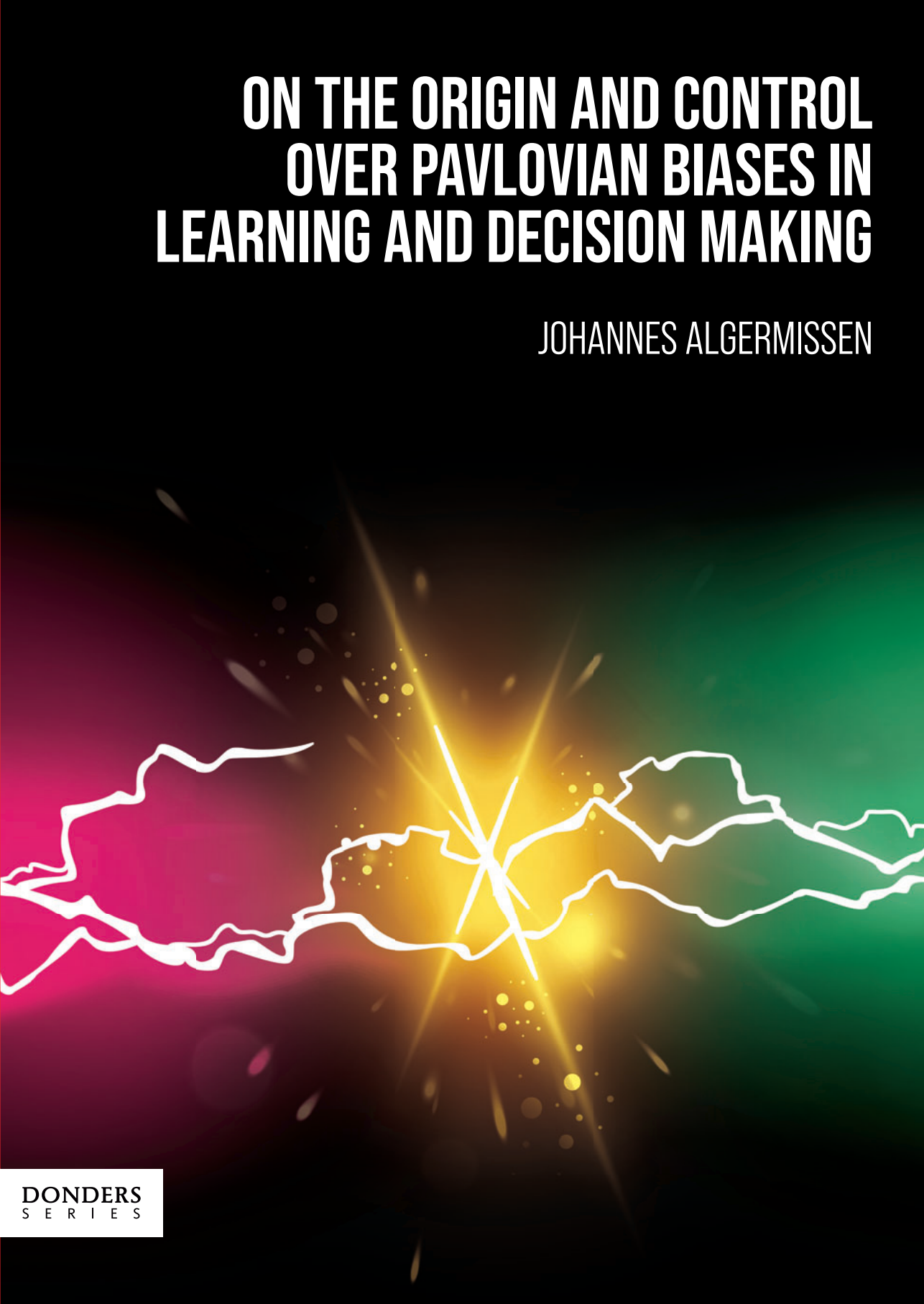
ON THE ORIGIN AND CONTROL OVER PAVLOVIAN BIASES IN LEARNING AND DECISION MAKING

JOHANNES ALGERMISSEN



DONDERS
SERIES

On the Origin and Control Over Pavlovian Biases in Learning and Decision Making

Johannes Algermissen

Colofon

The work described in this thesis was carried out at the Donders Institute for Brain, Cognition, and Behaviour, Radboud University.

Author:	Johannes Algermissen
Layout:	Johannes Algermissen
Cover layout:	Sandra Tukker | Studio Ridderprint
Printing:	Ridderprint | www.ridderprint.nl
ISBN:	978-94-6284-369-1

On the Origin and Control Over Pavlovian Biases in Learning and Decision Making

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op

maandag 3 april 2023
om 14:30 uur precies

door

Johannes Algermissen
geboren op 30 oktober 1991
te Hildesheim, Duitsland

Promotor:
Prof. dr. W.P. Medendorp

Copromotor:
Dr. H.E.M. den Ouden

Manuscriptcommissie:
Prof. dr. I. Toni
Prof. dr. M.J. Frank (Brown University, Verenigde Staten)
Dr. J.X. O'Reilly (University of Oxford, Verenigd Koninkrijk)

On the Origin and Control Over Pavlovian Biases in Learning and Decision Making

Dissertation

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.H.J.M. van Krieken,
according to the decision of the Doctorate Board
to be defended in public on

Monday, April 3, 2023
at 2:30 pm

by

Johannes Algermissen
born on October 30, 1991
in Hildesheim, Germany

PhD supervisor:
Prof. dr. W.P. Medendorp

PhD co-supervisor:
Dr. H.E.M. den Ouden

Manuscript Committee:
Prof. dr. I. Toni
Prof. dr. M.J. Frank (Brown University, United States of America)
Dr. J.X. O'Reilly (University of Oxford, United Kingdom)

ὃ γὰρ κατεργάζομαι οὐ γινώσκω· οὐ γὰρ ὃ θέλω τοῦτο πράσσω, ἀλλ᾽ ὃ μισῶ τοῦτο ποιῶ.
εἰ δὲ ὃ οὐ θέλω τοῦτο ποιῶ, σύμφημι τῷ νόμῳ ὅτι καλός.

Επιστολή προς Ρωμαίους, 7:15–16 (Novum Testamentum Graece NA 28)

Quod enim operor, non intelligo: non enim quod volo bonum, hoc ago: sed quod odi, illud facio.
Si autem quod nolo, illud facio: consentio legi, quoniam bona est.

Epistula beati Pauli apostoli ad Romanos, 7:15-16 (Biblia Sacra Vulgata)

For I do not understand my own actions. For I do not do what I want, but I do the very thing I hate.
Now if I do what I do not want, I agree with the law, that it is good.

Epistle of St. Paul to the Romans, 7, 15–16 (English standard version)

Wat ik doe, doorzie ik niet, want ik doe niet wat ik wil, ik doe juist wat ik haat.
Maar wanneer mijn daden in strijd zijn met mijn wil, dan erken ik dat de wet goed is.

Brief van Paulus aan de Romeinen, 7, 15–16 (Nieuwe Bijbelvertaling)

Denn ich weiß nicht, was ich tue. Denn ich tue nicht, was ich will; sondern was ich hasse, das tue ich.
Wenn ich aber das tue, was ich nicht will, stimme ich dem Gesetz zu, dass es gut ist.

Brief des Paulus an die Römer, 7, 15-16 (Lutherbibel 2017)

CONTENTS

APPENDICES

Chapter 1

General Introduction

1 General Introduction

1.1 The multiplicity of behavioral control systems

Our behavior does not always align with our best intentions. We eat that extra slice of cake, forgo our gym class, or drift off while working on an important assignment. We do not speak up when our friends are bullied, we do not save for retirement, and we do not take appropriate measures in face of climate change. This discrepancy between intentions and actual behavior, termed "self-control failure", "akrasia", or "**weakness of will**", has troubled humankind for millennia. Several philosophical and psychological accounts have tried to explain these phenomena by postulating the human mind to not be uniform, but the battleground of different sub-systems competing for control over behavior. This thesis addresses the conflict between two particular sub-systems, termed the "Pavlovian" and "instrumental" (or "goal-directed") systems, presenting four empirical studies that yield new insights in their neural origin as well as the arbitration between them.

Various philosophical and psychological accounts have postulated **different systems** that compete for control over behavior, explaining why agents might make choices—triggered by one system—that appear "wrong" from the perspective of another system. Important philosophical examples are, among others, the chariot drawn by a white "noble" and a black "troublesome" horse in Plato's Phaidros, St. Paul's and Augustine's strong distinction between mind and body ("flesh"), and, most prominently maybe, Sigmund Freud's distinction between Id, Ego, and Super-Ego. In psychology, in the 1990s and early 2000s, several *dual-process theories* have postulated a competition between two systems, namely a fast, automatic, impulsive "hot" system and a slower, controlled, deliberate "cold" system (Schneider and Shiffrin 1977; Shiffrin and Schneider 1977; Metcalfe and Mischel 1999; Loewenstein and O'Donoghue 2004; McClure et al. 2004; Strack and Deutsch 2004; Gawronski and Bodenhausen 2006). Probably the most famous of these dual-process theories is Daniel Kahneman's distinction between "System 1" and "System 2" (Kahneman 2011).

Dual-process theories of behavior propose the existence of multiple control systems because these systems are supposedly adaptive in different contexts, complementing each other (Milli et al. 2021). System 1 is fast and requires little deliberation, similar to a mental "auto-pilot", which leaves mental capacity to pursue other, putatively more important thoughts. Such a system is useful when behavior can be based on simple cues in a familiar and well-trained manner, e.g., stop at a red traffic light and move on when it turns green. In contrast, System 2 is slow and requires "active deliberation". Investing mental capacity into deliberation requires commitment and keeps an agent from pursuing other goals simultaneously. However, in the end, this system is more likely to yield a response in line with the agent's long-term goals. In such a dual-systems architecture, phenomena of "weakness of will" will arise when the fast System 1 determines behavior in a context in which the slow System 2 should have been relied on. For example, when considering one's lunch options, certain cues (e.g., the sight of a piece of cake) can activate the fast System 1 and elicit an impulsive action (buy and eat the cake) unless the slow System 2 takes over and instead picks a healthier food option in line with the agent's long-term goals. Such a "mis-reliance" on the fast System 1 might happen when an agent is distracted or mentally occupied with other thoughts that keep them from focusing on the decision at hand (Kahneman 2011).

The above-mentioned dual-process theories leave the actual information on which these two systems base their decisions rather opaque. The fast System 1 seems to be triggered by certain "biological primitives" such as appetitive foods that are postulated, but not summarized under an overarching principle. Such primitives—supposedly shared by all agents—cannot explain why different agents make different choices or why an agent behaves differently in two instances of the same context. A crucial driver that can explain both the variability in choice as well as how choices change over time is **(reinforcement) learning**, i.e., agents learning the "values" of different choice options from previous encounters with them. Agents can do so by comparing the outcomes of their actions with their expectations (i.e., their current estimate of the **value** of the action): If an action returns a better outcome than expected (a positive "prediction error"), they increase their estimate of an option's value, while when the action outcome is worse than expected (a negative "prediction error"), they decrease their value estimate. Selecting actions based on their past outcomes is termed **"instrumental" control**, a process by which agents actively manipulate their environment to maximize the chances for rewards (Dickinson and Balleine 1994). When choosing between different courses of action, agents could consider the values of all possible outcomes of a given action, combine each outcome with the respective probability of obtaining it, and then compute an integrative value of the action. When choices are rooted in (learned) values, agents might differ in their choices—because they have learned different values for different actions given a different reinforcement history—and choices might change over time due to updates in value.

When supplementing a dual-systems architecture with reinforcement learning that provides values as inputs to both systems, one assumes that both systems assign the same value to the same outcome. For example, both systems similarly appreciate the hedonic pleasure from a piece of cake based on past experiences. Still, different decisions can instead arise from both systems using **different styles of information processing** to integrate values of possible outcomes (Milli et al. 2021). The fast System 1 might have a narrow focus, solely relying on the presence of cues of particularly positive or negative value. A particularly positive cue—such as the sight of a piece of cake, but also a signpost announcing the availability of a piece of cake—would trigger System 1 to immediately elicit an automatic "approach" action. In contrast, the slow System 2 would have a wider focus and consider other potential outcomes—weight gain, disapproval by others—before reaching a decision. The opposite pattern would arise for cues of particularly negative value, e.g., the sight of a spider on a cupboard: the fast System 1 would automatically inhibit any action, while the slow System 2 could consider other items on the cupboard—such as the box with coffee powder needed to make a cup of coffee—and eventually decide to approach the cupboard. Under this perspective, System 1 relies on the same inputs as System 2, namely "values", but responds rather inflexibly to positive or negative value cues with the automatic invigoration or inhibition of actions. This inflexible way of action selection is termed **"Pavlovian" control** (Dayan et al. 2006; Boureau and Dayan 2011; Guitart-Masip, Duzel, et al. 2014).

Pavlovian control constitutes a candidate explanation of how values (learned from past experiences) automatically trigger action invigoration or inhibition and in this way could lead to phenomena of "weakness of will". Instead of postulating biological primitives, the concept of Pavlovian control gives concrete instructions how to artificially design (high value/ low value) cues that trigger the fast System 1. Hence, Pavlovian control can be a fruitful framework to study involuntarily triggered actions in a controlled (lab) environment. Initial studies suggest that

Pavlovian control assessed under lab environments is linked to "weakness of will" phenomena in real life: Alterations in Pavlovian control have been observed in patients suffering from psychiatric disorders such as depression (Huys et al. 2016), anxiety disorder (Mkrtchian, Aylward, et al. 2017), or alcohol addiction (Garbusow et al. 2016; Sekutowicz et al. 2019; Sommer et al. 2020; Chen et al. 2023), which tentatively suggests that altered Pavlovian control might contribute to the etiology and maintenance of such disorders. Understanding the neural mechanisms underlying Pavlovian control could thus eventually lead to the development of interventions to change it and in this way facilitate the treatment of psychiatric disorders.

In this thesis, I investigate **how the Pavlovian control system is balanced against the instrumental system**, both during action selection and during value updating after feedback has been received. In particular, I focus on ways by which the Pavlovian system can be **up- or down-regulated** in situations in which it is adaptive or maladaptive. I present four studies using functional magnetic resonance imaging (fMRI; Chapters 2 and 3), electroencephalography (EEG; Chapters 2 and 3), eye-tracking (Chapter 4), and magnetoencephalography (MEG; Chapter 5) that yield insights into the neural processes by which the Pavlovian system shapes behavior.

1.2 PAVLOVIAN BIASES IN ANIMAL AND HUMAN BEHAVIOR

The **Pavlovian system** comprises a class of automatic response tendencies typically called **"Pavlovian" or "motivational" biases**. These biases describe the fact that cues signaling the chance for positive or negative outcomes trigger fast, seemingly "automatic" responses: Cues signaling the opportunity for gaining a reward invigorate behavior (termed "Go" in the following), while cues signaling the risk of a punishment (a "threat") lead to behavioral inhibition (termed "NoGo"). Such response tendencies are amply visible in everyday life: tasty pieces of food such as cake or crisps, but also alcohol or other drugs, can trigger automatic approach and consumption that makes them "hard to resist". On the other hand, apparent threats such as the sight of a spider or snake can make us "freeze" on the spot (Fig. 1.1).

Figure 1.1. Coupling between action and valence.
Behavioral control can be mapped onto two seemingly independent axes: on the one hand, action running from action invigoration ("Go") to action inhibition ("NoGo"), and on the other hand, valence running from reward to punishment. For an instrumental control system, both axes are independent of each other. It should thus perform equally well at invigorating actions to obtain rewards ("Go2Win"), invigorating actions to avoid punishments ("Go2Avoid"), inhibiting actions to win rewards ("NoGo2Win") and inhibiting actions to avoid punishments ("NoGo2Avoid"). However, for the Pavlovian control system, chances to win rewards are intrinsically linked to action ("Go"), while risks of punishments are intrinsically linked to inaction ("NoGo"). Figure freely adapted after (Guitart-Masip, Duzel, et al. 2014).

Approach reward-predictive cues and freezing in face of punishment-predictive cues are likely **adaptive** in a majority of cases (Dayan et al. 2006; Guitart-Masip, Duzel, et al. 2014; O'Doherty et al. 2017), especially in those situations typical of our evolutionary past: Most rewards require active work to obtain them, and many threats (e.g., predators) can be evaded by staying still and waiting for them to disappear. In competition with other foragers and predators, response speed is key, and showing appropriate responses without much deliberation can yield an evolutionary advantage: In environments in which food resources are scarce (unlike our Western "obesogenic" environment), it is not worth deliberating whether to pick up a piece of food, but rather, one should immediately grab and consume it (lest it is consumed by competitors) (Hunt et al. 2016). Similarly, when becoming aware of nearby threats such as a predator, it might be adaptive to (initially) inhibit any ongoing behavior lest the predator becomes aware of oneself (Roelofs 2017). Hence, on evolutionary time scales, having a rigid "prior" (Nesse 2001; Haselton and Nettle 2006; Haselton et al. 2009; Fawcett et al. 2014; Parpart et al. 2018) that quickly triggers an action that is appropriate most of the times will be advantageous.

While contexts in which Pavlovian biases are adaptive might be ample, there are **exceptions**: At times, we have to wait for obtaining a reward, e.g., for a fruit to become ripe or for batter to turn into a cake. A famous instance in which humans' ability to wait for a (bigger) reward is measured is the "Marshmallow experiment" in which children have the opportunity to forego a single piece of marshmallow (or other preferred food) in order to obtain two pieces after a certain delay (Mischel and Ebbesen 1970; Mischel and Moore 1973; Mischel and Baker 1975). Here,

immediate approach to the single marshmallow interferes with the goal of obtaining the most marshmallows possible. Immediate gratification needs to be suppressed in service of long-term goal pursuit, a challenge common to human life, but particularly prominent in addictive behaviors (Garbusow et al. 2016; Sommer et al. 2020; Chen et al. 2023). Similarly, while freezing might be adaptive as an initial response in presence of a threat, other, more "active" behaviors such as fight or flight require effort mobilization to chase off or escape from the threat. Showing extensive freezing in face of potential threatening (e.g., novel) environments will hinder exploration and thus prevent the detection of novel food grounds, allies, or mates, behaviors common in social anxiety or depression (Huys et al. 2016; Mkrtchian, Aylward, et al. 2017). Pavlovian biases might thus constitute somewhat "global" stimulus-response associations, akin to "priors" on action selection (Moutoussis, Bullmore, et al. 2018) that need to become refined by learning "local", contextually more appropriate responses from experience. In each new situation, humans have to trade-off their "global" Pavlovian priors with their "local" knowledge about action-outcome associations.

While examples of Pavlovian biases are ample in human everyday life, they have first been systematically studied in **in other animal species**. The first descriptions of animals showing automatic (and quite vigorous) approach behavior to cues endowed with "value" are given in Breland and Breland's "The Misbehavior of Organisms" (Breland and Breland 1961), likely intended as a rebuttal to B. F. Skinner's "The Behavior of Organisms" (Skinner 1938). While Skinner claimed that learning makes animal behavior increasingly refined and "rational" in maximizing rewards, Breland and Breland presented case studies suggesting the contrary: Diverse animals such as chickens, pigs, and racoons initially learned to submit a token to the experimenter to obtain a food reward. However, over time, these tokens appeared to acquire a value on their own, making the animals more and more reluctant to let them go. These tokens—merely signaling the chance for a food reward—apparently acquired the ability to trigger automatic approach behaviors on their own, which interfered with the instrumental goal of food acquisition. Such behaviors became stronger and stronger over time, with animals seemingly unable to suppress them. Another illustration of animals' inability to overcome such biases is the famous "looking glass experiment" (Hershberger 1986). In this experiment, a food-carrying car was locked to chicken, mimicking its movements, but at twice the speed. Chickens continued to approach the car, which withdrew at twice the speed. However, they never learned to move away from the car to make it return. While these anecdotes demonstrated the pervasive and seemingly inescapable nature of Pavlovian control in the animal realm, further knowledge on the exact conditions under which such Pavlovian behavior occurs was only gained in systematic lab experiments.

Automatic approach behavior to cues endowed with positive value has first been systematically described in controlled lab experiments with pigeons (Morse and Skinner 1957; Brown and Jenkins 1968; LoLordo et al. 1974). A typical example is a classical conditioning experiment in which a light cue on one side of the pigeon's cage predicts food delivery on the other side of the cage. Typically, pigeons learn that the light predicts food and approach the food delivery side already in anticipation of food delivery. Such behavior is termed **goal-tracking** because it is oriented towards the animal's goal, i.e., the food reward itself. However, often, a subset of animals develops the persistent strategy of approaching the predictive light on the other side of the cage, foregoing the food. Such behavior is termed **sign-tracking** because it is oriented towards the reward-predictive cue rather than the reward itself. Sign-tracking appears to be a stable

animal trait (Flagel et al. 2007, 2009, 2010) and predicts an animal's response to addictive drugs (Yager and Robinson 2013; Robinson et al. 2014; Flagel and Robinson 2017).

Sign-tracking and goal-tracking only become visible when they elicit mutually exclusive behaviors, e.g., when rewards and reward-predictive cues are placed at opposite ends of an animal's cage. In contrast, when cue and reward are aligned—for example, by placing a light cue next to a lever that the animal has to press to obtain rewards—the cue can in fact help invigorate a behavior required to achieve rewards, facilitating reward acquisition (Estes 1943, 1948; Rescorla and Soloman 1967; LoLordo et al. 1974; Schwartz 1976; Lovibond 1983). This faciliatory nature of reward-predictive cues is systematically studied in **Pavlovian-to-Instrumental Transfer (PIT)** paradigms (Dickinson and Balleine 1994; Corbit and Balleine 2005; Corbit and Janak 2007; Holmes et al. 2010) consisting of three phases: In Phase 1 ("Pavlovian training"), an animal learns that a cue—such as a light or a tone—predicts the delivery of a reward. In Phase 2 ("Instrumental training"), the animal learns that a certain action—such as pressing a certain lever—returns rewards. In Phase 3 ("Transfer phase"), the animal receives again the opportunity to press a lever, but this time without rewards (i.e., in extinction). Typically, animals continue to press the lever, and do so more vigorously when the cue from Phase 1 is presented, suggesting that the reward-predictive cue can facilitate an action that used to return rewards. The PIT paradigm has also been translated to human research (Bray et al. 2008; Talmi et al. 2008) in which case reward-predictive cues presented in the background of a task can invigorate button presses to a foreground task. Most recent research has even started to quantify the effect of such background cues on performance to isolate a "sign-tracking" phenotype in humans (Garofalo and di Pellegrino 2015; Colaizzi et al. 2020; Schad et al. 2020). Furthermore, human research allows for the inclusion of cue predicting "punishments" (i.e., monetary losses), which exhibit effects opposite to reward-predictive cues: while reward-predictive cues invigorate button presses, punishment-predictive cues suppress it (Huys et al. 2011; Geurts et al. 2013a, 2013b), demonstrating that the opposite valence of these cues also manifests in opposite implications for behavior.

A shortcoming of PIT paradigms in humans is the fact that reward- and punishment-predictive cues are typically presented as "irrelevant" to the instrumental task. This lack of constraint leaves open the possibility that participants ignore those cues or come up with their own naïve theories about their putative role (Mahlberg et al. 2021). A possible remedy is to explicitly highlight the valence of cues, e.g., by linking them to the outcome a participant can get on a given trial. In the so-called **Motivational Go/ NoGo (MGNG) Learning Task** (Guitart-Masip, Fuentemilla, et al. 2011; Swart et al. 2017), certain cues signal the chance for a (monetary) reward (opposed to no outcome; called "Win" cues), while other cues signal the chance for a (monetary) loss (opposed to no outcome; "Avoid" cues; Fig. 1.2). Half of these cues require a Go response (button press) to obtain the reward or avoid the loss, while the other half requires a NoGo response (no button press). In this task, participants tend to perform more Go responses to Win cues, but more NoGo responses to Avoid cues, reflecting again that reward-predictive cues invigorate action, while punishment-predictive cues suppress it. These response tendencies are termed **"Pavlovian" or "motivational biases.** This task is currently the most straightforward way of assessing Pavlovian biases and will thus be used in Chapters 2 and 3 of this thesis.

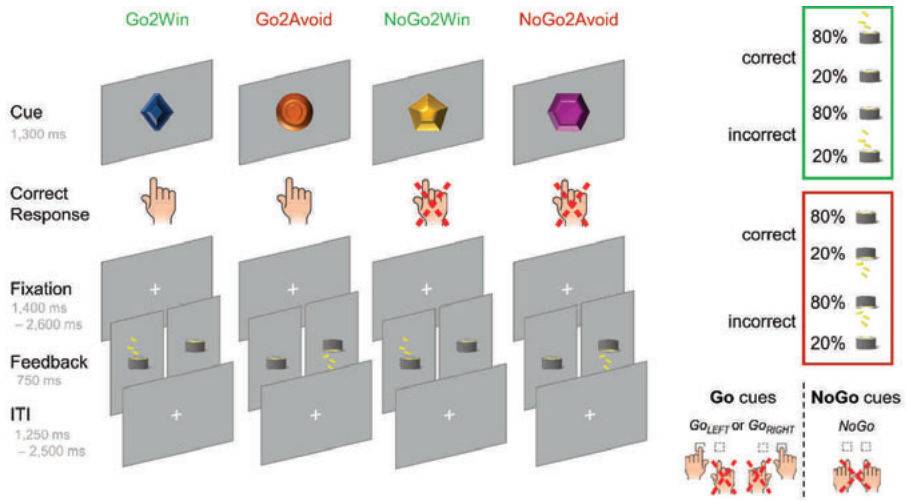


Figure 1.2. Motivational Go/ NoGo Learning Task as used in Chapters 2 and 3.
On each trial, a Win or an Avoid cue appears. Valence of the cue is not signaled, but needs to be learned from the task. Each cue features one correct response option, either GoLEFT, GoRIGHT, or NoGo, which participants have to learn from trial and error. Cue offset is also the response deadline. Response-dependent feedback follows after a jittered interval. For Win cues, actions can lead to rewards or neutral outcomes; for Avoid cues, actions can lead to neutral outcomes or punishment. Rewards and punishments are depicted by money falling into/ out of a can. Feedback is probabilistic: Correct actions to Win cues lead to rewards in 80% of cases, but neutral outcomes in 20% of cases. For Avoid cues, correct actions lead to neutral outcomes in 80% of cases, but punishments in 20% of cases. For incorrect actions, these probabilities are reversed. Figure freely adapted after (Swart et al. 2017) and (van Nuland et al. 2020).

Taken together, research in both humans and animals demonstrates how Pavlovian control can at times interfere with instrumentally controlled action selection, e.g., in paradigms eliciting sign-tracking, while at other times facilitate action selection, e.g., in PIT paradigms. However, up to now, two big questions about the interaction between Pavlovian and instrumental control remain: Firstly, it is **unclear at what neutral stage Pavlovian control interferes with action selection**. Notably, both Pavlovian and instrumental control influence the very same effector—the human's finger or the animal's paw—and must at some stage compete for control over a certain neural process that eventually triggers vs. inhibits a response. In the first part of the remainder of this Chapter, I will describe a particular neural architecture that could explain such a competition and then test its predictions in Chapters 2 and 3 using combined EEG-fMRI recordings. Secondly, in previous research, the experimental design determined whether Pavlovian and instrumental control were aligned or not, with the animal or human participant as a mere "subject" to environmental constraints. In the second part of the remainder of this Chapter, I will describe an **alternative possibility in which humans can actively contribute to the alignment of both systems**. I will first describe the role visual attention can play in controlling the input to the Pavlovian system and how this input could be selectively chosen based on action plans held by the instrumental system. I will then test the proposed role of visual attention in Chapters 4 and 5 using eye-tracking and MEG recordings.

1.3 Neural origin of Pavlovian response biases

Given that Pavlovian biases are shared across the animal realm (Breland and Breland 1961), they have been hypothesized to arise from evolutionary ancient, deep brain structures shared across many species. An architecture that implements instrumental control must receive projections from all areas containing relevant (sensory) information for action selection—i.e., almost all parts of cortex—and itself project towards relevant motor output regions such as motor cortex and deep brain nuclei (Fig. 1.3, 1.4). To explain how reward- and punishment-predictive cues can affect action selection, this architecture must also receive knowledge about (changes in) estimated values (i.e., reward prediction errors). A prime candidate for such an architecture are **cortico-basal-ganglia loops** (Haber et al. 2000; Haber 2003; Peters et al. 2021), i.e., recurrent connections from frontal cortical subregions to a range of subcortical basal ganglia nuclei which then project back to the same cortical origins. At the same time, these nuclei receive dopaminergic projections from midbrain nuclei which communicate reward predictions errors, informing these structures about the (positive/ negative) value of a given context.

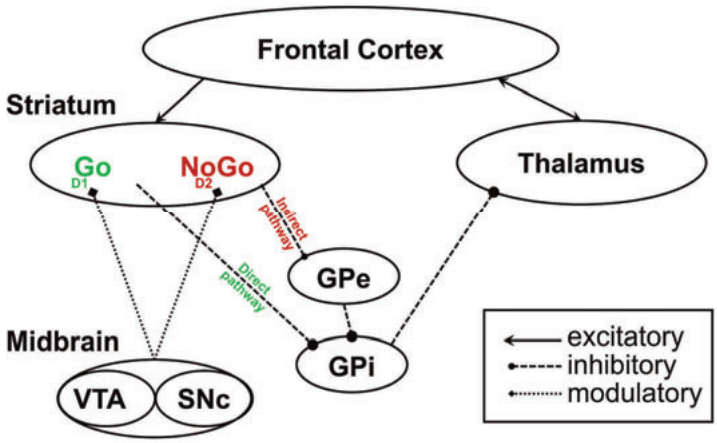


Figure 1.3. Cortico-striatal and thalamo-cortical loops including the direct and indirect pathway.
Frontal (i.e., prefrontal and motor) cortices send candidate action plans to the striatum, which can be divided into two pathways: Via the direct ("Go") pathway, the striatum projects directly onto the globus pallidus interior (GPi) and inhibits it. Globus pallidus inhibition disinhibits the thalamus, which makes it more likely that motor programs represented in motor cortex will be implemented and a Go action emitted. In contrast, via the indirect ("NoGo") pathway, the striatum projects indirectly to the globus pallidus interior via the globus pallidus exterior (GPe) and thus disinhibits it, which in turn inhibits the thalamus and suppresses any motor response. Dopaminergic neurons on the direct pathway predominantly express D1-type receptors, which sensitize its cells in high dopamine states (e.g., dopamine bursts elicited by positive prediction errors). In contrast, dopaminergic neurons on the indirect pathway predominantly express D2-type receptors, which suppress cell activity; hence, this pathway is most sensitive in low dopamine states (e.g., dopamine dips elicited by negative prediction errors). Ascending projections from the dopaminergic midbrain (ventral tegmental area, VTA; and substantia nigra pars compacta, SNc) signal dopaminergic reward and punishment prediction errors to the striatum, modulating the relative sensitivity of direct/ indirect pathways to cortical inputs. Under high dopamine states as induced by positive prediction errors (rewards), the direct pathway is more sensitive to cortical inputs and more likely gate a Go action, while in low dopamine states as induced by negative prediction errors (punishments), the indirect pathway is more sensitive to cortical inputs and more likely late a NoGo action. Figure freely adapted after (Frank, Rudy, et al. 2005) and (Frank 2006).

Cortico-basal-ganglia loops have been hypothesized as the target structure in which dopaminergic prediction error signals—conveying changes in values estimates—modulate ongoing action selection and give rise to Pavlovian biases in behavior. The crucial role of these structures in movement control was first appreciated in animals studies finding that lesioning these nuclei induced akinetic, Parkinson-like symptoms (Albin et al. 1989; DeLong 1990). The exact pathways from cortex throughout the basal ganglia nuclei and back were then systematically mapped in tracer studies (Alexander et al. 1986; Albin et al. 1989; Haber et al. 2000). Early basal ganglia models suggested that movement plans were generated by cortical regions and then forwarded to the basal ganglia, which acted as a filter-like selection mechanism boosting desired motor actions and inhibiting competing programs (Mink 1996; Hikosaka 1998). Further model refinements suggested this action selection process to be implemented via two opponent pathways (Frank 2005; Collins and Frank 2014): a **direct pathway** starting from the striatum and via the globus pallidus interior exciting the thalamus, and an **indirect pathway** starting from the striatum and via the globus pallidus exterior and interior inhibiting the thalamus (see Fig. 1.3). The starting point of both pathways, the striatum, is the major target of dopaminergic neurons from the dopaminergic midbrain (the ventral tegmental area, VTA, and substantia nigra pars compacta, SNc, Fig. 1.4). Neuron in the striatal part of both pathways show differential sensitivity to dopamine levels (Gerfen 1992; Aubert et al. 2000; Joel and Weiner 2000): Neurons in the direct pathway primarily feature D1-type dopamine receptors, which are excited by dopamine, making them more sensitive to cortical inputs under high dopamine levels (Hernández-López et al. 1997). States of high dopamine, such as after positive prediction errors induced by rewards (or reward-predictive cues), sensitize the direct pathway to frontal input and make it more likely to disinhibit the thalamus. A disinhibited thalamus is more likely to gate motor programs currently under consideration in motor cortex, and a motor action ("Go" response) is more likely to be emitted. In contrast, neurons in the indirect pathway primarily feature D2-type dopamine receptors, which are inhibited by dopamine, making them more sensitive under low dopamine levels (Hernandez-Lopez et al. 2000). States of low dopamine, such as after negative prediction errors induced by punishments (or punishment-predicting cues), excite the indirect pathway and make it more likely to inhibit the thalamus, meaning that any motor program is suppressed (leading to a "NoGo" response). This neural architecture links reward- and punishment-predictive cues to Go and NoGo responses and thus provides a mechanistic hypothesis about the neural origin of Pavlovian biases.

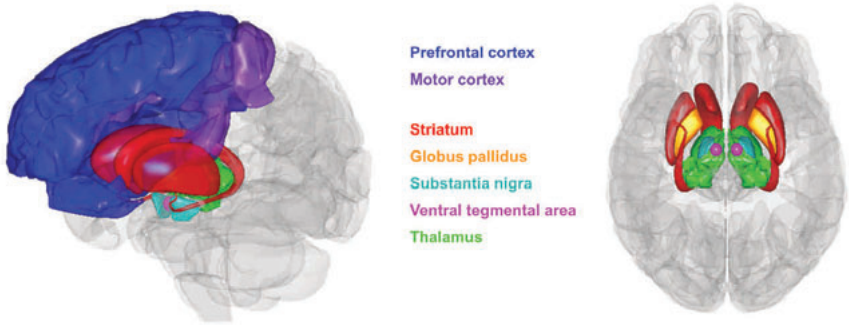


Figure 1.4. Anatomical locations of core regions involved in cortico-striatal and thalamo-cortical loops.
Recurrent connections between frontal (prefrontal and motor) cortex and the basal ganglia are involved in the modulation of candidate action plans. The sensitivity of direct and indirect pathways to cortical inputs is differentially modulated by dopaminergic projects from the midbrain (ventral tegmental area and substantia nigra). Left depiction from the front left, right depiction from the bottom of the brain. Images created using Anatomography, a website maintained by the Life Science Databases (LSDB), under CC-BY-SA-2.1-jp.

The idea that dopaminergic reward prediction errors from the midbrain affect the invigoration vs. inhibition of motor plans in cortico-basal-ganglia loops has received ample **empirical support**: Activity in the dopaminergic midbrain of animals (Schultz et al. 1997; Schultz 2019) as well as fMRI BOLD signal from the striatum of humans reflects reward prediction errors (O'Doherty et al. 2002; Tobler et al. 2007; Niv et al. 2012), which are modulated by changes in dopamine levels as induced by dopaminergic drugs (Pessiglione et al. 2006; Jocham et al. 2011). Behaviorally, elevated dopamine levels in the striatum have been found to lead to behavioral activation, enhancing the ability of reward-predictive cues to trigger responses (Taylor and Robbins 1984; Wyvell and Berridge 2000; Peciña and Berridge 2013; Halbout et al. 2019), while lesions of the striatum (Taylor and Robbins 1986; Hall et al. 2001; Corbit and Janak 2007), administration of dopamine antagonists (Dickinson et al. 2000; Corbit et al. 2007; Lex and Hauber 2008; Wassum et al. 2011; Ostlund and Maidment 2012) and dietary depletion of the dopamine precursor tyrosine (Hebart and Gläscher 2015) dampens responding to appetitive cues.

The direct/ indirect pathway model of cortico-basal ganglia loops makes straightforward predictions for how Win vs. Avoid cues in the MGNG Task should affect neural signals underlying motor actions: **Striatal BOLD signal** should be higher to Win than to Avoid cues when participants perform the MGNG Task, contributing to a higher propensity to emit Go vs. NoGo responses as visible in behavior. However, studies directly measuring striatal BOLD signal during the MGNG Task have unanimously found the striatum to reflect the executed action (Go/ NoGo) rather than the valence of the cue (Win/ Avoid) (Guitart-Masip, Fuentemilla, et al. 2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012; Moutoussis, Rutledge, et al. 2018). Even more, performance in the MGNG Task has been found to be altered under **dopaminergic drug administration** (Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Economides, et al. 2014; Swart et al. 2017; van Nuland et al. 2020), but the direction of effects varied substantially across studies. **Till today, it thus remains open whether Pavlovian biases truly arise through dopaminergic prediction errors biasing the striatal action selection process.** For higher statistical power, the above-mentioned studies exclusively focused on the striatum and midbrain using region-of-interest (ROI) analyses, but ignored other regions, leaving open whether other

areas might encode the cue valence. Cortical regions, such as ventromedial prefrontal cortex (vmPFC) or anterior cingulate cortex (ACC), have often been found to encode the valence or value associated with a certain cue, as well (Kable and Glimcher 2009; Bartra et al. 2013; Haber and Behrens 2014). Instead of dopaminergic prediction errors from the midbrain affecting striatal dopamine levels, value information provided by cortical regions might the crucial driver of Pavlovian biases. This idea makes the prediction that BOLD signal in cortical regions reflects cue valence during the MGNG Task.

Even when assuming cortical inputs instead of dopaminergic midbrain inputs as the source of valence information, it remains dubious why previous studies found striatal BOLD signal to not reflect cue valence in the MGNG Task. A potential explanation comes from recent studies finding that striatal activity does not reflect a pure reward prediction error. During action selection, agents should experience prediction errors as they make progress towards the goal. However, prediction error-like dopamine responses have only been observed in the context of "Go" actions, but not "NoGo" actions (Hamid et al. 2016; Syed et al. 2016). A novel hypothesis poses that during action selection, the striatum might not compute the value of the current context per se, but rather the "**value of work**", i.e., how much it is worth to invest effort into making the attainment of a positive outcome more likely (Berke 2018; Westbrook et al. 2020; Hamid et al. 2021). Under this hypothesis, dopamine tracks the "value" of each additional unit of effort, which increases the closer an agent gets to the eventual reward (i.e., the more certain reward attainment becomes). Support for this hypothesis has surged recently due to observations of dopamine "ramping" over several seconds before the attainment of a reward (Phillips et al. 2003; Roitman et al. 2004; Howe et al. 2013; Gershman 2014; Howe and Dombeck 2016; da Silva et al. 2018; Engelhard et al. 2019; Mohebi et al. 2019; Kim et al. 2020), but only if the animal can control outcome delivery (Hamid et al. 2021).

Observations that the striatum encodes the "value of work" during action selection suggest that striatal BOLD signal in the MGNG Task should primarily reflect whether an agent deems it worth (i.e., plans to) perform a Go response or not. Cue valence encoding might instead happen in other regions, such as vmPFC or ACC. I tested these two predictions in Chapter 2 of this thesis. Notably, valence and action coding should be visible at different time points during action selection: (cortical) cue valence signals should be present soon after a cue appears, while (striatal) signals reflecting the eventually executed response should be present much later, close to the actual response. While fMRI recordings provided access to deep brain structures such as the striatum, they lack the temporal resolution required to infer when exactly a signal appears—a resolution afforded by other techniques such as EEG and MEG. In Chapter 2 of this thesis, I present results from simultaneous EEG and fMRI recordings while participants performed an adapted version of the MGNG Task. Relating trial-by-trial estimates of the cortical and striatal BOLD response to millisecond-by-millisecond EEG data over the scalp allowed me to (a) test for EEG correspondents of cortical and subcortical BOLD signal, and (b) infer the relative timing of when these signals appeared (Fig. 4). I expect to see cortical signals reflecting the cue valence early after cue onset, but striatal signals reflecting the eventual response to occur closer to the eventual response.

1.4 Active suppression of Pavlovian biases

Unlike animals, humans are able to suppress Pavlovian biases in order to perform alternative behaviors in line with their long-term goals (Guitart-Masip, Fuentemilla, et al. 2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012; Cavanagh et al. 2013; Swart et al. 2017, 2018): They can wait for rewards—as famously exemplified in the Marshmallow Test in which infants will sometimes forego one piece of marshmallow in favor of two pieces at a later time point. Similarly, they take action to avoid an upcoming punishment, e.g., attack a potential predator (or just a small spider) to chase it away (Roelofs 2017). Still, even if participants are well aware which response they have to show in the MGNG Task, they tend to show "action slips" in response to Win cues and involuntarily omit actions in response to NoGo cues (Swart et al. 2017, 2018). Such "late errors" in action selection demonstrate that Pavlovian biases do not simply vanish once action-outcome contingencies are learned, but continue to shape behavior. In fact, research suggests that Pavlovian biases have to be actively suppressed to not interfere with instrumental action selection (Cavanagh et al. 2013; Swart et al. 2018).

The process of arbitrating between competing response tendencies is called **cognitive control**. The ability to suppress inappropriate response tendencies has previously been studied in other tasks, such as the Stroop, Simon, and Flanker tasks (Cohen et al. 1990; Botvinick et al. 2001). Studies have observed heightened activity in the ACC and adjacent regions, such as a presupplementary motor area (pre-SMA), when participants successfully inhibited inappropriate responses in favor of appropriate responses (Carter et al. 1998; Botvinick et al. 1999). Similarly, studies using EEG have observed increased midfrontal theta power (i.e., increases in rhythmic activity in the range of 4–8 Hz) in situations that required cognitive control (Cohen and Cavanagh 2011; Cohen and Ridderinkhof 2013), assumed to constitute the EEG equivalent to ACC activity. Computational models have conceptualized the ACC as a central hub monitoring for any conflict between competing response tendencies (Botvinick et al. 2001, 2004). Recent extensions such as the "expected value of control" (EVC) model suggest that ACC does not simply boost appropriate over inappropriate response tendencies, but evaluates the particular costs and benefits of recruiting additional resources to boost one response tendency over the other (Shenhav et al. 2013, 2016). In the context of the MGNG Task, the ACC is thus a prime candidate for detecting conflicts between Pavlovian and instrumental response tendencies and launching further steps to suppress Pavlovian control when it suggests an inappropriate action.

Two previous studies have observed increases in theta power when participants successfully inhibited Pavlovian biases (Cavanagh et al. 2013; Swart et al. 2018). However, it is unclear how, mechanistically, increases in theta power affect the expression of Pavlovian biases in subcortical action selection circuits. One explanation is provided by computational models of the basal ganglia suggesting the frontal cortex to trigger the subthalamic nucleus (STN) via the so-called "hyper-direct" pathway, which then elevates the response thresholds in basal ganglia loops (Frank 2006; Wiecki and Frank 2013). With higher thresholds, more evidence has to be accumulated until a response is elicited, slowing responses, but allowing appropriate (slower) response tendencies to take over the lead from inappropriate (faster) response tendencies. While this model has received considerable empirical support (Cavanagh et al. 2011; Zavala et al. 2014; Frank et al. 2015; Herz et al. 2016), it seems to only apply to competitions between two or more active ("Go") responses. It is unclear how elevated response thresholds should help agents to overcome a NoGo tendency

induced by Avoid cues and instead invigorate a Go response. **It is thus till today unclear how midfrontal control processes—as visible in midfrontal theta power increases—contribute to the suppression of Pavlovian biases, which likely originate from subcortical processes.** An alternative explanation is that ACC and striatum interact directly via cortico-striatal loops (Haber et al. 2000; Haber 2003) (Fig. 1.3) with recurrent inputs from ACC attenuating striatal signals. In Chapter 2, I combine simultaneous EEG and fMRI measures to investigate the downstream effects of midfrontal theta power increases on subcortical BOLD signal.

1.5 Neural origin of Pavlovian learning biases

Apart from action selection, also **learning** from rewards and punishments appears to be subject to Pavlovian biases (Swart et al. 2017, 2018; de Boer et al. 2019): Humans find it relatively easier to learn to perform Go responses to obtain rewards, but more difficult to unlearn NoGo responses to avoid punishments (Fig. 1.5). In the same way as action selection, learning can be a compromise between "global" priors about typical action-outcome associations—apparent as **Pavlovian learning biases** in behavior—and "locally" acquired knowledge about contextually appropriate action-outcome relationships. The global priors of crediting actions for rewards, but not blaming inactions for punishments might again reflect environmental statistics: Most rewards need to be actively collected; hence, responses appropriate for their collection should be quickly acquired. Vice versa, most threats (e.g., predators) might be avoided by staying still, a tendency that should be robust to occasional situations in which freezing does not prevent detection by the predator (but instead leads to a "punishment"). Taken together, on the one hand, Pavlovian learning biases might speed up learning and make it more robust against environmental noise. On the other hand, these biases might give rise to rare phenomena of "animal superstition" like negative auto-maintenance, i.e., when pigeons consistently perform a response they believe to be necessary to obtain a reward while, in fact, this response is unrelated to (or even delays) reward delivery (Brown and Jenkins 1968; Williams and Williams 1969; Dayan et al. 2006). Even humans have a propensity to perceive agency over factually random events (Wegner et al. 2004; Aarts et al. 2005) and perform effortful responses that are factually in vain. Conversely, in the domain of punishments, humans have apparent problems to attribute harmful outcomes to someone's inactions, making them more lenient in judging others' action omissions (Ritov and Baron 1990, 1995; Baron and Ritov 1994; Zeelenberg et al. 2000). In the MGNG Task, biased behavior appears to (at least partially) arise from learning biases in addition to response biases (Swart et al. 2017, 2018; de Boer et al. 2019). In sum, Pavlovian learning biases might direct agents to infer likely action-outcome relationships that are correct in many situations, while at the same time making them prone to occasional misattributions.

Pavlovian learning biases are predicted by the asymmetric nature of direct and indirect pathways in the basal ganglia, as well (Frank 2005; Collins and Frank 2014) (Fig. 1.3): Positive prediction errors as elicited by rewards lead to dopamine bursts and long-term potentiation (LTP) in the direct pathway, facilitating the acquisition of Go responses following rewards, while negative predict errors as elicited by punishments lead to dopamine dips and LTP in the indirect pathways, hindering the unlearning of NoGo responses following punishments. In fact, also Pavlovian learning biases seem to be under dopaminergic control (Swart et al. 2017; de Boer et al. 2019). **However, the putative striatal origin of Pavlovian learning biases has remained untested.**

It is possible that, alternatively, learning biases arise through cortical regions biasing striatal learning mechanisms. Learning guided by instructions has been found to reflect prefrontal cortical influences (Doll et al. 2009, 2011; Atlas et al. 2016). Similarly, the BOLD signal in ACC has been found to reflect environmental volatility, which impacts the learning rate and thus the speed with which participants incorporate new information in their value estimates (Behrens et al. 2007; O'Reilly et al. 2013; Meder et al. 2017). Hence, it is well possible that cortical signals regulate how much striatal learning mechanisms learn from positive and negative outcomes.

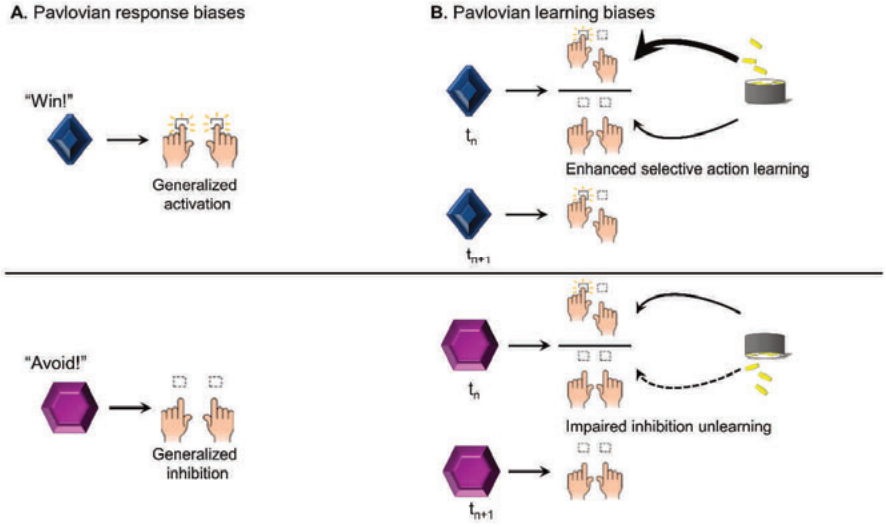


Figure 1.5. Illustration of Pavlovian response biases and learning biases.
A. Pavlovian response biases arise when humans see the chance to win rewards or the risk of receiving punishments: Reward prospect generally invigorates any motor action, while punishment prospect generally inhibits any motor action. **B.** Pavlovian learning biases arise when agents preferentially update the values of actions/ inactions after reward/ punishment outcomes. When a Go action leads to a reward, agents show an increased learning rate, preferentially attributing rewards to their own actions. Notably, this attribution only affects the value of the action previously shown, not the values of other actions. Conversely, when a NoGo action leads to a punishment, agents exhibit a reduced learning rate, showing an impairment in attributing punishment to them having held back. The introduction of multiple Go actions (Go_{LEFT} vs. Go_{RIGHT}) allows for disentangling the influences of response and learning biases using computational reinforcement learning models. Figure freely adopted after (Swart et al. 2017).

In Chapter 3 of this thesis, I directly test whether trial-by-trial striatal BOLD signals were better explained by "standard" reward prediction errors without any bias or in fact by biased prediction errors with heightened credit for rewarded actions and diminished credit for punished inactions. Similarly, I test whether cortical activity was better explained by biased prediction errors. Finally, by combining simultaneously acquired EEG and fMRI recordings, I test whether EEG correspondents of cortical prediction error signals preceded EEG correspondents of striatal prediction error signals—suggesting that, indeed, cortical regions instruct striatal learning in a way that gives rise to learning biases in behavior.

1.6 THE PUTATIVELY ADAPTIVE ROLE OF BIASES

So far, I have suggested that Pavlovian biases can be adaptive in constituting global "priors" on action selection, suggesting responses that are likely conducive to obtaining rewards or avoiding punishments. Having such biases might be adaptive on "large", evolutionary time scales (Dayan et al. 2006; O'Doherty et al. 2017) rather than locally in each modern context. However, we know little about the environments in which these biases were putatively shaped, and any claim about such environments appears difficult to empirically verify or falsify (Haselton and Nettle 2006; Haselton et al. 2009). Still, it remains dubious why human participants show frequent action slips or omissions in the MGNG Task even if they have learned the required action of each cue (Cavanagh et al. 2013; Swart et al. 2018). Apparently, biases do not vanish with time and instrumental control taking over, but they continue to shape behavior. In Chapter 2, I focus on how these biases are actively suppressed by cortical regions involved in cognitive control. In contrast, **in Chapters 4 and 5, I test whether humans are able to up- or down-regulate their Pavlovian biases using selective visual attention in a way that is line with their ongoing action plans.** Instead of actively suppressing Pavlovian biases, these could in fact be actively recruited to pursue the same goals as instrumental control.

Previous theoretical perspectives have casted Pavlovian and instrumental control as two separate systems that operate largely in parallel and compete for control over behavior at the motor output level (Dayan et al. 2006). Arbitration between such systems is typically assumed on be based on the reliability of each system, i.e., a system is given more consideration if it predicts action outcomes accurately, but downweighed if its predictions do not match obtained outcomes (Daw et al. 2005; Keramati et al. 2011; O'Doherty et al. 2017; Dorfman and Gershman 2019). Pavlovian control only considers cues irrespective of actions to predict future outcomes, leading to frequent errors in outcome predictions, while instrumental control considers knowledge about action-outcome contingencies. Under such an architecture, the influence of Pavlovian biases should vanish overtime and eventually become zero, which does not match behavioral data demonstrating the pervasive nature of these biases (Cavanagh et al. 2013; Swart et al. 2018). In contrast, in the new perspective that I advocate in Chapters 4 and 5, the **influence of the Pavlovian system stays high because it is regularly recruited by the instrumental system in service of its goals**. The Pavlovian system acts ballistically: the sight of a reward-predictive cue makes it trigger action, while the sight of a punishment-predictive cue makes it suppress action. By being sensitive to a single stimulus feature, i.e., predicted value, it is more robust to interference and distraction than an instrumental system that has to monitor incoming sensory information to potentially update its inferences about the current context and the values of possible actions. Beyond more robust action selection, the instrumental system may also become faster in eliciting actions and collecting outcomes when it is supported by the Pavlovian system. Having both systems aligned could thus gain a crucial temporal advantage over competitors.

The tool by which the instrumental system could recruit the Pavlovian system is **selective visual attention.** Agents are not just passively exposed to cues, but can actively choose which cues to attend to ("active sensing") (Gottlieb 2018; Gottlieb and Oudeyer 2018). Given how ballistically the Pavlovian system responds to reward or punishment inputs, attending to the "correct" cues is key. By actively steering visual attention to reward or punishments cues, the instrumental system could strategically recruit the Pavlovian system to invigorate or suppress

actions. If the instrumental system has the plan to invigorate an action, it could direct attention to a reward cue in the environment, which activates the Pavlovian system and automatically triggers a Go response. Vice versa, if the instrumental system plans to hold in, it could direct attention to a punishment cue in the environment that activates the Pavlovian system and automatically suppresses any action (Fig. 1.6). Under this perspective, the instrumental system does not need to overrule the Pavlovian system at the output level, but can align it with its own goals by controlling its inputs. By focusing on rewards/ punishments, the instrumental system can set the Pavlovian system on an almost "ballistic" Go/ NoGo track, ensuring execution of the intended action even in face of distractors or other interferences (Fig. 1.6). Such instances of one decision system recruiting or training another have been proposed and observed before, e.g., between model-based and model-free systems in retrospective reward revaluation (Robinson and Berridge 2013; Gershman et al. 2014), credit assignment (Moran et al. 2019), and memory replay (Mattar and Daw 2018). In Chapters 4 and 5, I test whether participants strategically attended to reward- and punishment-predictive cues in a way that would make the Pavlovian system trigger the response required by task demands.

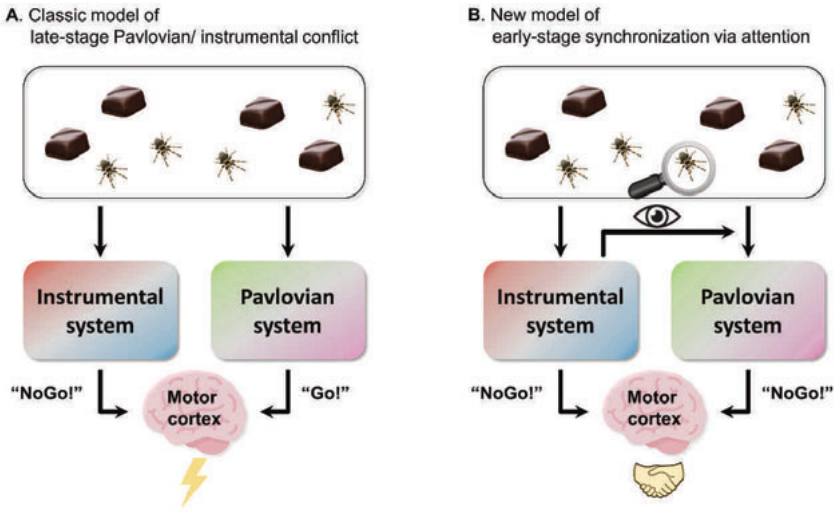


Figure 1.6. Illustration of different models of resolving Pavlovian-instrumental conflicts.
A. Previous research has assumed that instrumental and Pavlovian control systems work largely independently, potentially triggering incompatible action plans at the motor output stage. At this late stage, conflict is detected and mechanisms that suppress the influence of the Pavlovian system need to be recruited. **B**. According to the new framework I suggest, both systems do not work independently, but the instrumental system can use selective visual attention to control the visual input to the Pavlovian system. Under a Go action plan, the instrumental system can steer attention towards reward cues, which will activate the Pavlovian system and lead to a Go action in line with the original action plan. Vice versa, under a NoGo action plan, the instrumental system can steer attention towards punishment cues, which will activate the Pavlovian system and lead to a NoGo action in line with the original action plan. This early-stage synchronization prevents late-stage conflicts and makes both system work in concert to pursue the same goal.

As an illustrative example for how instrumental recruitment of Pavlovian control might have been adaptive in a past "environment of evolutionary adaptedness", imagine a hunter that waits for the right moment to throw a spear at a prey animal. During this delay, their plans might be

disrupted by other events that capture their attention: They might hear a nearby bird chirp, spot another prey further away, or mentally drift off to their plans for the evening. In such cases, they might forget about the action they initially intended—or alternatively perform it too late such that the prey escapes or is caught by a competitor. Hence, focusing on a cue that both reminds them of the action they intend to perform and that ensures that they do so quickly enough can potentially increase their hunting success. Hence, the hunter could focus his attention on a cue that signals the chance for reward (e.g., the prey itself) and in this way ensure that their plan is eventually smoothly executed. Notably, action invigoration might be particularly strong for a fat (compared to a scrawny) prey, such that the hunter's action (and effort investment) might still be tuned by the factual outcome prospects in a particular situation. Conversely, in other situations, the hunter might have to remain still to not be caught by a nearby lion. Again, they might be distracted by a bird or an unrelated thought, causing them to step on a branch or start talking to their companions. Here, again, focusing on a cue that reminds them to not perform any response could be adaptive. The hunter might fixate a cue signaling threat (e.g., the predator itself) to constantly suppress any action. The level of suppression might still depend on the level of threat associated with the cue (e.g., whether it is a lion or merely a spider), tuning the hunter to be more careful and risk-averse in presence of more dangerous animals. In sum, a hunting situation in face of predators might very well profit from selectively attending to reward- or punishment-predictive cues to flexibly invigorate or suppress action at the right moment.

In sum, beyond suggesting actions in accordance with global environmental statistics, Pavlovian biases could also be useful in supporting the execution of action intentions. Such a mechanism can be particularly advantageous when actions unfold over time and are prone to interference. Apart from hunting prey in previous millennia, also the modern world can require humans to act quickly once a certain opportunity opens—in summer/ winter sales or online concert ticket purchases, in claiming spots in a lecture hall, or in searching team mates for a group work. Having a clear action goal in mind and focusing on potential positive outcomes could give a competitive advantage over someone who has the same goal, but is occupied by negative thoughts that might even inhibit action. Similarly, some activities require us to keep still and wait for the right moment—not just Marshmallow Tests, cake recipes, or Dutch (tulip) auctions, but even talking to a certain person or delivering a certain piece of (good) news (e.g., promotion) in an appropriate context. To not impulsively deliver this information at the wrong moment, focusing on negative cues or thoughts could be adaptive. In such a role, Pavlovian biases do not restrict the range of possible behaviors, but instead constitute a mechanism that gives agents additional flexibility and freedom to pursue their goals. In Chapters 4 and 5, I test whether human participants make active use of this opportunity, measuring their attentional focus via eye-tracking (Chapter 4) and MEG (Chapter 5).

1.7 Flexible recruitment of Pavlovian biases via selective visual attention

In Chapter 4, I develop an adapted version of the MGNG Task to test whether action plans shape how humans attend to reward or punishment cues. This research was motivated by two existing lines of research: one line showing how action plans influence which visual features or

locations humans attend to, and another line showing that the amount of attention humans direct to their choice options has a strong effect on their eventual action.

Ample research has shown that action plans direct attention: humans are more easily distractable by cues that share features with the task goal (Folk et al. 1992; Eimer and Kiss 2008; Van der Stigchel and Hollingworth 2018). Theoretical perspectives of ***active sensing*** describe how attention and eye-gaze are not merely determined by bottom-up saliency, but shaped by top-down goals that recruit attention to actively interrogate the environment (Yang et al. 2016; Gottlieb and Oudeyer 2018). The ***premotor-theory of attention*** goes as far as claiming that the primary role of attention is to facilitate the preparation of a (manual) action towards an intended target (Rizzolatti et al. 1987; Sheliga et al. 1997; Olivers and Roelfsema 2020). Indeed, action intentions sharpen visual sensitivity to action-relevant features, such as object location for reaching movements or object size and orientation for grasping movements (Craighero et al. 1999; Bekkering and Neggers 2002; Fagioli et al. 2007; Welsh and Pratt 2008; Wykowska et al. 2009), with similar effects recently reported for working memory performance (Heuer and Schubö 2017; Heuer et al. 2017).

While there is little research in the domain of value-based decision-making, some evidence suggests that the plan to choose an option directs attention towards this option—even if attention is costly and the option has already been rendered superior to other available options (Hunt et al. 2016, 2018; Kaanders et al. 2021). Such findings that humans seek out "positive evidence" before selection an option suggest that also in value-based choice, attention is guided by action plans. Such evidence gives reason to assume that attention to reward- and punishment-predictive cues is not static, but might be flexibly adjusted based on (Go/ NoGo) task demands.

Vice versa, previous research suggests not only an effect of action plans on attention, but also a strong link between what people attend to and what they eventually choose. In choices among generally positive objects such as food snacks, looking longer at an item makes it more likely to be eventually chosen. The opposite link holds for generally negative objects, e.g., rotten food items (Armel et al. 2008), effortful tasks (Westbrook et al. 2020), or risky gambles (Pachur et al. 2018), in which case longer attention to the negative features of an item predicts its eventual rejection. Still, it is unclear by which mechanism attention could possibly make an appetitive object more appetitive or an aversive object more aversive. Theoretical perspectives have suggested that attention facilitates the retrieval of positive/ negative features of an object from memory (Shadlen and Shohamy 2016), but direct (neural) evidence for such an attention-facilitated retrieval mechanism has not been provided yet. Such theoretical positions propose that considerations of the features of an object—which contribute to the object's value and are thus relevant to the correct choice—drive attentional effects. However, alternatively, it may be that attention to any positive or negative information—even if this information does not contribute to the object's value and is in fact complete choice-irrelevant—has similar effects. Possibly, any positive or negative information has the potential to trigger Pavlovian biases and in this way lead to the selection or rejection of an object. In Chapter 4, I indeed test the hypothesis that attention to task-irrelevant positive or negative information affects an eventual Go/ NoGo response—supporting the view that attentional effects on choice can be explained by Pavlovian biases.

In sum, past research suggests influences of action plans on attention (to action-relevant objects) as well as influences of attention (on reward- or punishment-related information) on eventual action (selection or rejection). In Chapter 4, I combine and extend these two literatures in the domain of value-based decision-making to test whether attention to reward- and punishment-predictive cues might be selectively used to trigger Go/ NoGo actions in line agents' action plans.

1.8 Neural influences of action plans on visual processing in the recruitment of Pavlovian biases

Recruiting Pavlovian biases via visual attention implies that information about action plans—likely from motor cortex—must become available to other brain regions regulating visual attention—such as visual cortex, but also frontal, parietal or subcortical regions (Corbetta and Shulman 2002; Baluch and Itti 2011; Squire et al. 2013). Obtaining "online" neural measures of action planning and visual processing is crucial because these processes are dynamic and can change over time without consequences for overt behavior. For example, participants might be tempted by a reward cue to make a Go action, but inhibit themselves at the last moment—which leads to the same behavioral response as intending a NoGo action all along. Similarly, participants might fixate a certain cue for a long time period while their thoughts are wandering off, but be fully alert during the short time period that they fixate another cue. The dissociation between gaze and "attentional focus" renders the time participants fixate a given cue an imperfect measure of the "amount of processing" it receives. These considerations highlight the shortcomings of overt behavioral measures and the added benefit that can be gained from neural measures that provide online estimates of evolving action planning and attention. For achieving an adequate temporal resolution for tracking changes in the processes, techniques such as EEG or MEG are more suited than fMRI.

An established measure of both action selection and execution is a desynchronization of **beta oscillations (13–33 Hz) over sensorimotor cortex** (Salmelin and Hari 1994; Neuper et al. 2006; van Ede, Chekroud, Stokes, et al. 2019; Boettcher et al. 2021). Power in this range decreases when humans select an action (Boettcher et al. 2021) and again when they execute it, with stronger decreases contralateral to the executing hand (Donner et al. 2009; O'Connell et al. 2012). When ongoing action preparation is disrupted or voluntarily inhibited, beta power re-synchronizes (Walsh et al. 2010; Gluth et al. 2013). Assessing the state of beta power of sensorimotor cortex could thus give insight into the current state of action preparation and how it changes when subjects are confronted with—or voluntarily attend to—reward- or punishment-predictive cues.

An established measure of covert spatial attention to the left/ right are **asymmetries in alpha oscillations (8–13 Hz) over parietal and occipital cortex**. Alpha power decreases at stimulus onset and does so more strongly contralaterally to an attended stimulus (Worden et al. 2000; Thut et al. 2006; Rihs et al. 2007). Alpha power decreases likely reflect selective disinhibition of higher visual cortices to allow further processing of incoming visual information (Klimesch et al. 2007; Jensen and Mazaheri 2010). Tracking the state of ongoing alpha oscillations could thus give insight into the relative amount of attention that reward and punishment-predictive cues presented on the left/ right side of the screen receive. Previous research has indeed shown that reward- or

punishment-associated cues presented in a lateralized manner attract attention as reflected in posterior alpha power lateralization (Marshall et al. 2018).

In real life, reward and punishment cues suitable for invigorating action plans might not always be in the immediate field of view. Instead, humans have to actively sample those cues, or even stock their environment with potential cues. In the eye-tracking paradigm in chapter 4, I required participants to actively direct their fixation at certain cues locations in order to render those cues visible. Crucially, such sampling requires an up-front plan of where to attend first, second, and so on. If this plan is ill-chosen, participants' attention might be caught by a distractor, which potentially disrupts their ongoing action plan. Hence, this paradigm incentivized a strategy that uses ***anticipatory, proactive attention*** to selectively attend to helpful and ignore unhelpful cues.

However, while the eye-tracking design might come close to active sampling in real life, it does not mimic situations in which cues are visible in the periphery and participants can covertly screen them before deciding which cues warrant more attention. Such a ***reactive attentional strategy*** (Geng 2014) might be more efficient in many situations given that attentional shifts driven by bottom-up information are faster than attentional shifts driven by top-down control (Wolfe et al. 2000). In situations in which speed is not prioritized and the occurrence of distractors is rather unlikely, a reactive strategy might be advantageous over a proactive strategy (Aron 2011; Braver 2012). Alpha power lateralization might be a marker of such reactive attention that first scans which stimuli appear on the screen and where they are located. In particular, it has been argued that (unlike target selection processes) distractor suppression processes cannot rely on pre-knowledge of where on the screen distractors will occur (Noonan et al. 2018; van Moorselaar and Slagter 2020). Instead, such processes need to either capitalize on statistical regularities describing where distractors tend to occur or have to spontaneously react to the sudden occurrence of unexpected distractors. Such a slow, reactive strategy was not possible in the gaze-contingent paradigm employed in Chapter 4. However, the version of this paradigm employed in Chapter 5 allows participants to freely view both stakes, while instructing them to keep their focus at the center of the screen. Hence, in Chapter 5, alpha power lateralization measured via MEG recordings could give insights into whether also a merely reactive attentional strategy incorporated Go/ NoGo action plans.

So far, I have considered decreased/ increased alpha power as the prime candidate of how the processing can be enhanced/ suppressed. However, it is also possible that the bottom-up processing of cues can be modulated independently of spatial attention as visible in alpha power lateralization. Instead, parietal saliency maps (Itti and Koch 2000; Gottlieb and Oudeyer 2018) could detect at a very early stage with cues deserve spatial consideration and thus upregulate the processing of these cues. Such facilitated bottom-up processing of a given cue could be visible in stronger event-related fields (ERFs) as well as increased posterior gamma power (33–100 Hz) contralateral to that cue. For example, associations of a cue with rewards or punishments have been found to enhance early evoked responses towards it already in the first 100–200 ms after cue onset, resulting in larger N1/ P1 component (Hickey et al. 2010; Luque et al. 2017). Similarly, research has found enhanced posterior gamma power contralateral to reward- or punishment-associated stimuli compared to neutral stimuli (Marshall et al. 2018), indicating their preferential bottom-up processing (van Kerkoerle et al. 2014; Bastos et al. 2015). In sum, influences of action plans on reward and punishment processing might not (only) be apparent in the deployment of

spatial attention as indexed by alpha lateralization, but also in enhanced bottom-up processing of cues that match action plans as indexed in stronger ERFs and contralateral gamma power.

Taken together, in Chapter 5, I test complementary neural mechanisms by which action plans could inform visual processing using MEG recordings. I expect that attention to reward- and punishment-predictive cues—as indexed by posterior alpha power lateralization—would be aligned to Go/ NoGo action plans—as indexed by the state of sensorimotor beta power desynchronization. In addition, I test two complementary mechanisms for such an alignment: One possibility entails the *proactive* biasing of spatial attention already before cues appear, while another possibility entails the *reactive* biasing of alpha after cues have appeared. Furthermore, the latter strategy might be supplied by enhanced visual bottom-up processing of cues that match action plans. This chapter provides new insights into how visual attention could be recruited in a goal-directed manner to elicit Pavlovian biases in line with instrumental action plans.

1.9 Aims and outline of this thesis

This thesis presents new insights into the origin of and control over Pavlovian biases in both response selection and learning. Both issues are addressed as follows:

In **Chapter 2**, I use simultaneous EEG-fMRI recordings to investigate the **origin of** Pavlovian **response biases** as well as **how biases can be suppressed by midfrontal circuits**. The direct/ indirect pathway model of the basal ganglia predicts that value prediction signals from midbrain nuclei should affect action selection in the basal ganglia, leading to higher striatal BOLD signal for reward-predictive compared to punishment-predictive cues. Data from previous studies have not matched this prediction (Guitart-Masip, Fuentemilla, et al. 2011; Guitart-Masip, Huys, et al. 2012). In this chapter, firstly, I test for the possibility that value signals arise in alternative areas, notably prefrontal regions such as vmPFC and ACC. Also, I test whether the striatum selectively codes for the "value of work" (i.e., the expected value for making a Go action) rather than value per se. Combining trial-by-trial fMRI BOLD amplitudes with simultaneously recorded EEG measures allows me to test whether prefrontal and striatal signals reflected processes early after cue onset (i.e., reflecting cue encoding) or later processes close to the eventual response (i.e., reflecting response selection). Secondly, previous studies have observed increased midfrontal theta power when human participants successfully suppressed Pavlovian biases (Cavanagh et al. 2013; Swart et al. 2018). However, it still remains unclear how midfrontal influences could possibly facilitate the suppression of inappropriate responses in the context of a Go/ NoGo decision—especially how an inappropriate NoGo tendency in light of a punishment-predictive cue could potentially be overcome (and a Go tendency invigorated instead). Combining EEG and fMRI recordings allows me to investigate the downstream consequences of midfrontal theta power increases on striatal BOLD signals, particularly whether theta power increases lead to an attenuation of striatal valence signals. This chapter yields novel insights into the origin of Pavlovian response biases and how they can be suppressed by cortical circuits.

In **Chapter 3**, I use simultaneous EEG-fMRI recordings to investigate the **neural origin of Pavlovian learning biases**. The direct/ indirect pathway model of the basal ganglia predicts not only biased action selection, but also biased learning: Positive prediction errors facilitate plasticity in the direct ("Go") pathway and thus crediting Go actions for positive outcomes, while negative

prediction errors promote plasticity in the indirect ("NoGo") pathway and impair the extinction of NoGo actions after negative outcomes. However, alternatively, biased learning might not be an emerging feature of the basal ganglia architecture itself, but instructed by cortical regions, most notably vmPFC and ACC. In this chapter, firstly, I test whether BOLD signal in striatum and prefrontal regions is better described by assuming biased prediction errors (in line with Pavlovian learning biases) rather than "standard" prediction errors. Secondly, by combining EEG and fMRI recordings, I test whether EEG correlates of prefrontal learning signals arise earlier than EEG correlates of striatal learning signals, suggestive of a prefrontal influence on striatal learning that might give rise to learning biases. This chapter yields insights in the origin of Pavlovian learning biases and whether those arise purely from the basal ganglia architecture or reflect prefrontal influences.

In **Chapter 4**, I use eye-tracking recordings to test whether humans **use overt visual attention to seek out reward- and punishment-related information in a way that triggers Pavlovian biases matching their action plans**. Previous models have considered instrumental and Pavlovian systems as operating in isolation and frequently triggering incompatible actions, leading to conflict that needs to be arbitrated. Instead, in this chapter, I investigate whether the instrumental system can selectively seek out reward- or punishment-predictive cues to align the Pavlovian system with its action plans. More specifically, I test whether participants' action plans influence their attention allocation to reward- and punishment-predictive cues, and whether, vice versa, their attention to these cues predicts their eventual response. This chapter yields novel insights into whether and how humans can selectively recruit Pavlovian biases that match their action plans using selective visual attention.

In **Chapter 5**, I use MEG recordings to test complementary neural mechanisms by which action plans could influence visual processing of reward- and punishment-predictive cues. I use the state of beta power desynchronization as an index of ongoing action preparation as well as posterior alpha power lateralization as an index of the current focus of attention. First, I test whether the attentional focus on reward and punishment-predictive cues was aligned with participants' action plans. Second, I test two complementary mechanisms by which action plans could shape visual processing: As one, more "proactive" alternative, I test whether humans align their attentional focus with their action plans in a proactive, anticipatory manner even before cues appear, or, alternatively, whether they do so in a reactive manner only after cues have appeared. Furthermore, as a complementary strategy to bias cue processing that goes beyond spatial attention as indexed in alpha power lateralization, I test whether bottom-up visual processing—reflected in ERFs and posterior gamma power—was enhanced for cues that matched action plans. This chapter yields novel insights into how action plans shape visual processing in the recruitment of Pavlovian biases.

These four empirical Chapters shed new light on the origin of and control over Pavlovian biases. They provide insight into which brain structures make us automatically approach a slice of cake and shrink back at the sight of a spider. They show how phenomena of "weakness of will" might be rooted in evolutionary old neural circuits shared across the animal realm. Furthermore, they suggest ways in which humans could use these automatic stimulus-response links to deliberately invigorate or suppress other actions. Exhibiting (and adaptively using) Pavlovian biases might be integral to normal human functioning—they help us approach positive and avoid

negative stimuli in a fast and easy manner. Altered Pavlovian biases might in fact contribute to the etiology and maintenance of psychiatric disorders such as depression (Huys et al. 2016), social anxiety (Mkrtchian, Aylward, et al. 2017), or (alcohol) addiction (Garbusow et al. 2016; Sekutowicz et al. 2019; Sommer et al. 2020; Chen et al. 2023). Together, these chapters can help to reframe seemingly "impulsive", cue-triggered actions as part of normal human life and as a tool rather than a burden.

Chapter 2

Striatal BOLD and midfrontal theta power express motivation for action

2 STRIATAL BOLD AND MIDFRONTAL THETA POWER EXPRESS MOTIVATION FOR ACTION

2.1 ABSTRACT

Action selection is biased by the valence of anticipated outcomes. To assess mechanisms by which these motivational biases are expressed and controlled, we measured simultaneous EEG-fMRI during a motivational Go/NoGo learning task (N=36), leveraging the temporal resolution of EEG and subcortical access of fMRI. VmPFC BOLD encoded cue valence, importantly predicting trial-by-trial valence-driven response speed differences and EEG theta power around cue onset. In contrast, striatal BOLD encoded selection of active Go responses and correlated with theta power around response time. Within trials, theta power ramped in the fashion of an evidence accumulation signal for the value of making a 'Go' response, capturing the faster responding to reward cues. Our findings reveal a dual nature of midfrontal theta power, with early components reflecting the vmPFC contribution to motivational biases, and late components reflecting their striatal translation into behavior, in line with influential recent "value of work" theories of striatal processing.

2.2 Introduction

Learning from rewards and punishments allows us to adapt action selection to our environment. At the same time, our responses are also shaped by seemingly automatic action tendencies that appear to be innate or acquired very early in development. A prime example are motivational action biases (also called "Pavlovian" biases), referring to the tendency to invigorate actions when there is a prospect of reward, but to hold back when there is a threat of punishments (Dayan et al. 2006; Guitart-Masip, Duzel, et al. 2014). Such an action 'prior' allows for fast responding, which is often adaptive given that reward pursuit typically requires active responses and threat avoidance typically requires action suppression. However, in environments in which these relationships do not hold, the action triggered by the motivational bias can interfere with the normative optimal action (i.e., the action that maximizes rewards and minimizes punishments) and needs to be suppressed. Keeping a balance between hardwired action tendencies and action values flexibly learned from experience can be challenging, and deficits have been linked to psychiatric disorders such as addiction, depression, trauma symptoms, and social anxiety (Garbusow et al. 2016, 2019; Huys et al. 2016; Mkrtchian, Aylward, et al. 2017; Ousdal et al. 2018). Thus, it is important to understand the mechanism by which humans selectively rely on these biases when they are helpful and suppress them when they are not. In this study, we investigate the neural interactions that accompany (un)successful suppression of motivational biases when needed.

Motivational biases have been hypothesized to arise from dopaminergic effects in the basal ganglia (Frank 2005; Collins and Frank 2014). In line with the putative role of the striatum in driving motivational biases, striatal fMRI BOLD signal has been found modulated by cues signaling the prospect of reward (O'Doherty et al. 2003; Tobler et al. 2007; Niv et al. 2012) and dopaminergic medication has been found to modulate these motivational biases (Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Economides, et al. 2014; Swart et al. 2017; van Nuland et al. 2020).

When actions triggered by motivational biases conflict with the action required to obtain a desired outcome, individuals need to detect this motivational conflict and mobilize control mechanisms to suppress biases. This role has traditionally been attributed to interactions between the medial and lateral prefrontal cortex, particularly the anterior cingulate cortex (ACC). More recent approaches regard the medial frontal cortex, in particular the ACC, as a central decision hub evaluating whether to recruit cognitive control or not (Shenhav et al. 2013, 2016). Recruiting cognitive control is perceived as costly, but potentially worth the effort to overcome biases and select the optimal action.

Bursts of oscillatory synchronization in the theta range (4-8 Hz) over midfrontal cortex have been proposed as an electrophysiological signature of conflict processing. Recently, we and others have observed this signal also when participants successfully overcame motivational conflict between task-appropriate and task-inappropriate, bias-triggered actions (Cavanagh et al. 2013; Swart et al. 2018; Csifcsák et al. 2020). However, the downstream mechanisms by which midfrontal signals prevent the behavioral expression of motivational biases, putatively driven by the striatum, remain elusive. Previous research has focused on tasks with two active responses (e.g., the Stroop task), where midfrontal theta has been suggested to activate the subthalamic nucleus, which raises the threshold of striatal input needed to elicit an action and thus prevents impulsive actions (Zavala et al. 2014; Frank et al. 2015; Aron et al. 2016; Herz et al. 2016). In contrast, in our task, not only

reward-triggered action, but also punishment-triggered inhibition needs to be overcome, which might be achieved by a direct attenuation of the cortical inputs into the striatum implemented via recurrent fronto-striatal loops (Alexander et al. 1986; Mink 1996; Gurney et al. 2001). In the current study, we tested the hypothesis that midfrontal theta power is associated with an attenuation of subcortical signals that encode motivational biases when those need to be suppressed.

To test this hypothesis, participants performed a Motivational Go/NoGo learning task known to elicit motivational biases in humans (Swart et al. 2017, 2018) while simultaneously recording fMRI and scalp EEG. The temporal precision of EEG allowed us to separate cue-induced conflict signals from later, response-locked signals. We tested the specific hypotheses that i) striatal BOLD encodes cue valence, ii) this valence signaling is attenuated when participants successfully overcome motivational biases, and iii) valence signal attenuation correlates with midfrontal theta power. Specifically, using BOLD signal to predict EEG power at different time points allowed us to separate regions involved in (early) valence cue processing vs. (later) response biases.

2.3 Materials and Methods

2.3.1 Participants

Thirty-six participants (M_{age} = 23.6, age range 19–32; 25 women, all right-handed) performed the motivational Go/NoGo learning task while simultaneous EEG and fMRI were recorded. Sample size was based on previous EEG studies (Cavanagh et al. 2013; Swart et al. 2018) accounting for potential dropout. The study was approved by the local ethics committee (CMO2014/288; Commissie Mensgebonden Onderzoek Arnhem-Nijmegen). All participants provided written informed consent. Exclusion criteria comprised claustrophobia, allergy to gels used for EEG electrode application, hearing aids, impaired vision, colorblindness, history of neurological or psychiatric diseases (including heavy concussions and brain surgery), epilepsy, metal parts in the body, or heart problems.

Participants attended a single three-hour recording session and were compensated for participation (€30). Additionally, they received a performance-dependent bonus (range €0–5, M_{bonus} = €1.28, SD_{bonus} = 1.54). The reported behavioral and EEG results are based on all 36 participants. For two participants, fMRI co-registration failed due to excessive orbitofrontal distortion in the T1 image; thus, fMRI results are based on 34 participants (M_{age} = 23.5, age range 19–32; 25 women). fMRI-informed EEG results are based on 29 participants (Mage = 23.4, 21 women): Apart from the two participants for whom co-registration failed, we excluded four further participants who exhibited strong head motion (i.e., at least 5 volumes with relative displacement larger than the voxel size of 2 mm). These four participants also exhibited stronger overall head motion (i.e., mean relative displacement across all volumes), M = 0.213, SD = 0.084, compared to the other participants, M = 0.088, SD = 0.040. Head motion is a particular problem in the EEG-fMRI combined analysis, as it will lead to high and spatially uniform correlations between fMRI and EEG data. Indeed, for these participants, regression weights for all regressors were an order of magnitude larger than for the other participants and largely uniform across time and frequency.

We repeated behavioral, EEG and fMRI results for the subgroup of these 29 participants in S2.1; conclusions were identical unless mentioned otherwise in the main text.

2.3.2 Motivational Go/NoGo learning task

Participants performed the Motivational Go/NoGo learning task as detailed in (Swart et al. 2018) with trial timings adjusted to the fMRI acquisition (Fig. 2.1). The task was programmed in MATLAB 2014b (The MathWorks, Natick, MA, United States) / Psychtoolbox-3.0.13. Each trial started with the presentation of one of eight cues (a colored geometric shape) in the center of the display (Fig. 2.1A). This cue determined the outcome valence (Win reward/ Avoid punishment) and which action (Left Go/ Right Go/ NoGo) was required for the desired outcome (win reward/ avoid punishment). Participants had to learn both the valence of the cue and the required action from trial-and-error (Fig. 2.1B). For Win cues, participants should aim to win a reward and avoid neutral outcomes, while for Avoid cues, they should aim to achieve neutral outcomes and avoid punishments. Participants could respond by pressing a left button (Left Go), right button (Right Go), or choose not to press (NoGo) while the cue was on screen for 1300 ms. After a pseudorandomly jittered fixation period of 1400–2600 ms, the outcome was presented for 750 ms in the form of money falling into a can (reward), money falling out of a can (loss / punishment), or just the can (neutral outcome). Outcomes were probabilistic so that the optimal action led to the desired outcome in only 80% of trials, while suboptimal actions led to desired outcomes in 20% of trials (Fig. 2.1C). Each trial ended with a pseudorandomly jittered inter-trial interval of 1250–2000 ms, resulting in an overall trial length of 4700–6650 ms. Analyses of learning and outcome-based activity in EEG and fMRI will be reported in a separate publication.

Participants received two button boxes in the scanner, one for each hand, and were instructed to use only one of the four keys on each button box. When participants accidentally pressed one of the three other buttons, the text "invalid response" appeared instead of an outcome. For analyses purposes, such invalid button presses were recoded into the valid button press of the respective hand, assuming that participants aimed for the correct key of the respective button box.

Before the actual task, participants underwent a practice session in which they were familiarized first with each condition separately (using practice stimuli) and then practiced all conditions together, using different cues from the actual experiment. They were informed about the probabilistic nature of feedback and that each cue features one optimal action. The actual task comprised 320 trials, split into three blocks of approximately ten minutes with short breaks between blocks. Participants performed the task twice, with a different set of cues, yielding 640 trials in total. Introducing a new set of cues allowed us to prevent ceiling effects in performance and investigate continuous learning throughout the task.

2.3.3 Behavioral data analysis

For behavioral analyses, all trials with RTs before 1.3 seconds (i.e., cue offset) were treated as Go responses (in line with fMRI and EEG analyses). Button presses were treated as Go responses irrespective of whether the correct button or an incorrect button was pressed. We recoded 43 trials with invalid button presses (i.e., pressing a button that was not instructed as response button) into the correct response button of the respective hand (0.19% of trials; max. 14/640 per participant). For analyses of RTs, 12 trials with RTs smaller than 200 ms (0.08% of trials, max. 5/640 per participant) were excluded as such responses were unlikely to follow from cue-based action

selection. Furthermore, 980 trials with RTs larger than 1300 ms (i.e., after cue offset, which was the instructed response time limit; 6.5% of trials, max. 139/640 per participant) were excluded from RT analyses as it was unclear whether such button presses were still intended as responses to the cue or were mere "action slips". Results did not qualitatively change when including these trials.

We used mixed effects logistic regression (lme4 package in R) to analyze participants' behavioral responses (Go vs. NoGo). We assessed a main effect of required action (Go/ NoGo), reflecting whether participants learned the task (i.e., showed more Go response to Go than NoGo cues), a main effect of cue valence (Win/ Avoid), reflecting the motivational bias (i.e., more Go response to Win than Avoid cues), and the interaction between valence and required action. The model written in Wilkinson notation was:

$$response \sim cueValence * requiredAction + (cueValence * requiredAction | participant)$$

All variables were treated as factors, with sum-to-zero coding. To achieve a maximal random effects structure (Barr et al. 2013), we added random intercepts and random slopes of all three predictors for each participant and further allowed for random correlations between all predictors. The same predictors and random effects structure were used to analyze RTs, for which linear regression was used. *P*-values were computed using likelihood ratio tests (package afex in R). *P*-values smaller than $\alpha = 0.05$ were considered statistically significant. In all plots, whiskers indicate standard errors, which were computed across participants based on the per-condition-per-participant means using the Cousineau-Morey method (Morey 2008).

2.3.4 fMRI data acquisition

MRI data were acquired on a 3T Siemens Magnetom Prisma fit MRI scanner. In the scanner, participants' heads were stabilized with foam pillows, and a strip of adhesive tape was applied to participants' forehead to provide motion feedback and minimize head movement (Krause et al. 2019). After two localizer scans to position slices, functional images were collected using a whole-brain T2*-weighted sequence (68 axial-oblique slices, TR = 1400 ms, TE = 32 ms, voxel size 2.0 mm isotropic, interslice gap 0 mm, interleaved multiband slice acquisition with acceleration factor 4, FOV 210 mm, flip angle 75°, A/P phase encoding direction) and a 64-channel head coil. This sequence yielded a short TR at high spatial resolution, which allowed us to disentangle BOLD signal related to cue and outcome presentation. The sequence parameters were piloted to find a sequence with minimal signal loss in the striatum. The first seven volumes of each run were automatically discarded.

After task completion, when the EEG cap was removed, an anatomical image was collected using a T1-weighted MP-RAGE sequence (192 sagittal slices per slab, GRAPPA acceleration factor = 2, TI = 1100 ms, TR = 2300 ms, TE = 3.03 ms, FOV 256 mm, voxel size 1.0 mm isotropic, flip angle 8°) for registration and a gradient fieldmap (GRE; TR = 614 ms, TE1 = 4.92 ms, voxel size 2.4 mm isotropic, flip angle 60°) for distortion correction. For one participant, no fieldmap was collected due to time constraints. At the end of each session, an additional diffusion tensor imaging (DTI) data collection took place; results will be reported elsewhere.

2.3.5 fMRI preprocessing

fMRI data for each of the six blocks per participants were preprocessed using FSL 6.0.0 (Smith et al. 2004). Functional images were cleaned for non-brain tissue (Smith 2002), segmented, motion-corrected (Jenkinson et al. 2002), and smoothed (FWHM 3 mm). Field maps were used for B0 unwarping and distortion correction in orbitofrontal areas. We used ICA-AROMA (Pruim et al. 2015) to automatically detect and reject independent components in the data that were associated with head motion (non-aggressive denoising option).

To prevent empty regressors on a block level, we concatenated blocks and performed a single first-level GLM per participant. For this purpose, we registered the volumes of all blocks to the middle image (the default registration option in FSL) of the first block of each participant (using MCFLIRT) and then merged files. The first and last 20 seconds of each block did not contain any trial events, such that, when modelling trial events, no carry-over effects from one block to another could occur.

After concatenation and co-registration of the EPI, we performed high-pass filtering with a cutoff of 100 s, and pre-whitening. We then computed the co-registration matrices of EPI images to high-resolution anatomical images (linearly with FLIRT using Boundary-Based Registration) and to MNI152 2mm isotropic standard space (Andersson et al. 2007).

2.3.6 ROI selection

We used masks of selected ROIs for three purposes: fMRI GLMs with small volume correction, BOLD-RT correlations, and fMRI-informed EEG analyses. For GLMs with small-volume correction, we used anatomical masks of striatum and ACC/pre-SMA based on our a-priori hypothesis and previous literature. Anatomical masks were based on the Harvard-Oxford atlas, thresholded at 10%. For BOLD-RT correlations and fMRI-informed EEG analyses, we used conjunctions of anatomical masks and relevant functional contrasts. These are follow-up analyses to results observed in the fMRI GLMs and test independent hypotheses. The well-established rationale for using these conjunctive constraints (O'Reilly et al. 2012) is that the anatomical constraints ensure that all voxels are from the same anatomical region, while the functional constraints ensure that voxels reflect the signal of interest. These masks were obtained by thresholding the z-map of the relevant functional contrast at $z > 3.1$ (cluster-forming threshold) and then combining it with the anatomical mask using the logical AND operation.

We obtained the following anatomical masks by combining sub-masks of the Harvard-Oxford atlas: striatum (bilateral caudate, putamen, and nucleus accumbens), midfrontal cortex (anterior division of the cingulate gyrus and juxtapositional lobule cortex—formerly known as supplementary motor cortex), ACC (anterior division of the cingulate gyrus), left and right motor cortex (precentral and postcentral gyrus), vmPFC (frontal pole, frontal medial cortex, and paracingulate gyrus).

Masks were either used in the group-level GLM for small-volume correction or back-transformed to participants' native space to extract either parameter estimates from the GLM or the raw BOLD signal time series. All masks are displayed in S2.2.

2.3.7 fMRI analysis

We fitted a first-level GLM to the data of each participant using the fixed-effects model in FSL FEAT. The four task regressors of interest were the four conditions resulting from crossing cue valence (Win/Avoid) and performed action (Go/NoGo irrespective of Left vs. Right Go), all modeled at cue onset. We also included five regressors of no interest: two at cue onset, namely response side (Go left = +1, Go right = -1, NoGo = 0) and errors (i.e., participants chose the incorrect action), and three at outcome onset to control for outcome-related activity, namely outcome onset (intercept of 1 for every outcome), outcome valence (reward = +1, punishment = -1, neutral = 0), and invalid trials (invalid buttons pressed and thus not feedback given). We further added the following nuisance regressors (separate regressors for each block): intercept, six realignment parameters from motion correction, mean cerebrospinal fluid (CSF) signal, mean out-of-brain (OOB) signal, and separate spike regressors to model out each volume on which relative scan-to-scan displacement was more than 2 mm (occurred in 10 participants; in those participants: M = 7.40, range 1 – 29). Task regressors were convolved with a double-gamma haemodynamic response function (HRF) and high-pass filtered at 100 s. The full model is also displayed in S2.3.

We hypothesized that a main effect of valence in the striatum would be attenuated under motivational conflict. Thus, the positive BOLD response to a Win cue would be reduced when having to make a NoGo response for Win, and the negative BOLD response to an Avoid cue would be less negative for a Go response for Avoid (Fig. 2.3A). This hypothesized conflict-mediated attenuation would thus manifest itself as a (weaker) main effect of action in the presence of a (stronger) main effect of valence. Taken together, we predicted that striatal BOLD signal would exhibit both a main effect of valence and a main effect of action. Indeed, previous literature using a simpler version of this task (Guitart-Masip et al. 2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012) has reported a main effect of action in the striatum. In addition to testing the main effects of valence and action, we also computed a congruency contrast, reflecting motivational conflict. Finally, left vs. right Go responses were contrasted to identify lateralized motor cortex activation as a quality control check. Regressors were specified across both correct and incorrect trials; we report results for regressors on the correct trials only (in line with our EEG analysis approach) in S2.3.

First-level participant-specific contrasts were fitted in native space. Co-registration and re-slicing was then applied on each participant's contrast maps, which were combined at the group-level (using FSL FLAME for mixed effects models; (Woolrich et al. 2004, 2009) with a cluster-forming threshold of $z > 3.1$ and cluster-level error control at $\alpha < .05$ (i.e., two one-sided tests with $\alpha < .025$).

Given our specific hypotheses about the striatum and midfrontal cortex, we additionally tested contrasts using a small-volume correction: For the Valence and Action contrasts, we used an anatomical mask of the striatum, and for the Congruency contrast (i.e., NoGo2Win and Go2Avoid minus Go2Win and NoGo2Avoid), we used a mask of the entire midfrontal cortex (ACC and pre-SMA). While theoretical models predict differences in BOLD signal in ACC, empirical fMRI studies have actually found correlates in more dorsal regions, specifically pre-SMA (Botvinick et al. 2004). Indeed, source reconstruction of midfrontal theta suggested a rather superficial source in pre-SMA (Cohen and Ridderinkhof 2013) and a simultaneous EEG-fMRI study found choice

conflict related to pre-SMA BOLD (Frank et al. 2015). Thus, the anatomical mask comprised the entire midfrontal cortex.

2.3.8 BOLD – behavior correlations

To assess whether regions that encoded cue valence, i.e., information driving motivational biases, predicted reaction times, we performed regression analyses of the trial-by-trial BOLD signal in relevant regions on RTs (for trials where participants made a Go response). The main regions we found to encode cue valence were vmPFC (positively) and ACC (negatively). We computed conjunctions of anatomical vmPFC and ACC masks with the cue valence contrast and back-transformed the resultant masks to each participant's native space. We then extracted the first eigenvariate of the signal in both ROIs, returning one summary measure of BOLD signal in that ROI per volume. The volume-by-volume signal in each ROI was then highpass-filtered at 128 s and nuisance regressors (6 realignment parameters, CSF, OOB, single volumes with strong motion, same as in the fMRI GLM) were regressed out. Afterwards, the signal was upsampled by factor 10, epoched into trials of 8 s duration (Hauser et al. 2015), and a separate HRF was fitted for each trial (i.e., 57 upsampled datapoints).

We then tested whether BOLD signal in those ROIs correlated with reaction times on a trial-by-trial basis. In order to account for overall differences between trials with Win cues and trials with Avoid cues, we standardized reaction times and BOLD signal separately for Win trials and Avoid trials within each participant such that differences between cue valence conditions were removed. We computed correlations between trial-by-trial reaction times and BOLD signal HRF amplitude for each participant, applied Fisher-z transformations to correlations to make them normally distributed, and then tested with a one-sample t-test whether correlations were significantly different from zero at a group level.

2.3.9 EEG data acquisition and preprocessing

EEG data were acquired using 64 channels (BrainCap-MR-3-0 64Ch-Standard; Easycap GmbH; Herrsching, Germany; international 10-20 layout, reference electrode at FCz) at a sampling rate of a 1,000 Hz using MRI-compatible EEG amplifiers (BrainAmp MR plus; Brain Products GmbH, Gilching, Germany). Additional channels for electrocardiogram, heart rate, and respiration were used recorded for MR artifact correction. Recordings were performed with Brain Vision Recorder Software (Brain Products). EEG amplifiers were placed behind the scanner, and cables were attached to the cap once participants were positioned in the scanner. Cables were fixated with sand-filled pillows to reduce artifacts induced through cable movement in the magnetic field. During functional scans, the MR scanner helium pump was switched off to reduce EEG artifacts. A Polhemus FASTRAK device was used to record the exact location of each EEG electrode on the participant's head relative to three fiducial points. For four participants, no Polhemus data were recorded due to time constraints and technical errors; for these participants, the average channel positions of the remaining 32 participants were used.

EEG data were cleaned from MR scanner and cardioballistic artifacts using BrainVisionAnalyzer (Allen et al. 2000). Pre-processing was performed in Fieldtrip (Oostenveld et al. 2011) in MATLAB 2017b by rejecting channels with high residual MR noise (mean 4.8 channels per participant, range 0–13), epoching trials (-1750–2800 ms relative to cue onset, total duration of 4550 ms), re-referencing the channels to their grand average and recovering the reference as

channel FCz, band-pass filtering the data in the 0.5–15 Hz range using a two-pass 4th order Butterworth IIR filter (Fieldtrip default), and finally linear baseline correction based on the 200 ms prior to cue onset. The low-pass filter cut-off of 15 Hz allowed us to dissociate theta from its adjacent bands while filtering out residual high-frequency MR noise. We used ICAs to visually identify and reject independent components related to eye blinks, saccades, head motion, and residual MR artifacts (mean 12.94 components per participant, range 8–19), and afterwards manually rejected trials that were still contaminated by noise (mean 29.6 trials per participant, range 2–112). Finally, we computed a Laplacian filter with the spherical spline method to remove global noise (using the exact electrode positions obtained with the Polhemus FASTRAK), which we also used to interpolate previously rejected channels. This filter attenuates more global signals (deep sources) and noise (heart-beat and muscle artifacts) while accentuating more local effects (superficial sources).

For response-locked analyses, we re-epoched Go actions trials, time-locked to the time of response (RT). For NoGo response trials, we re-epoched the data time-locked to the average RTs (for each participant) of Go actions for that cue valence, as a proxy for 'latent RTs' on these trials.

2.3.10 EEG TF decomposition

Time-frequency decomposition was performed using Hanning tapers between 1–15 Hz in steps of 1 Hz, every 25 ms with 400 ms time windows. We first zero-padded trials to a length of 8 sec. and then performed time-frequency decomposition in steps of 1 Hz by multiplying the Fourier transform of the trial with the Fourier transform of a Hanning taper of 400 ms width, centered around the time point of interest. This procedure results in an effective resolution of 2.5 Hz (Rayleigh frequency), interpolated in 1 Hz steps, which is more robust to the exact choice of frequency bins. Given that all pre-processing was performed on data epoched into trials, we aimed to exclude the possibility of slow drifts in power over the time course of the experiment. We thus performed baseline correction by fitting a linear model across trials for each channel/frequency combination. This model included trial number as a regressor and the average power in the last 50 ms before cue onset as outcome. The power predicted by this model was then removed from single-trial data. Note that in absence of any drift, this approach amounts to correcting all trials by the grand-mean across trials per frequency in the selected baseline window per participant. Next, we averaged over trials within each condition spanned by of valence (Win/ Avoid) and action (Go/ NoGo; correct trials only). Finally, power was converted to decibel for all analyses to ensure that data across frequencies, time points, electrodes, and participants were on the same scale.

In line with (Swart et al. 2018), we restricted our analyses to correct trials, i.e., trials where required and performed action matched. We assumed that on correct incongruent (Go2Avoid and NoGo2Win) trials, participants successfully detected and resolved conflict, while no such processes were required on congruent (Go2Win and NoGo2Avoid) trials. Although the same processes might be initiated on incorrect trials, though unsuccessfully, these trials are potentially confounded by error-related synchronization in the theta range (Cavanagh, Zambrano-Vazquez, et al. 2012), which makes the interpretation of any effects in the theta range less straightforward. To be consistent with fMRI results, we report results across both correct and incorrect trials in S2.6 and S2.10.

2.3.11 EEG data analysis

All analyses were performed on the average signal of the a-priori selected channels Fz, FCz, and Cz based on previous findings (Swart et al. 2018). We performed non-parametric cluster-based permutation tests (Maris and Oostenveld 2007) as implemented in Fieldtrip for the selected electrodes in the theta range (4–8 Hz) during cue presentation (0–1300 ms). This procedure is suited to reject the null hypothesis of exchangeability of two experimental conditions, but not suited to exactly determine when or where differences occur (Sassenhagen and Draschkow 2019). Our interpretations of when and where conditions differed in power are thus based on visual inspection of the signal time courses.

Given our a-priori hypothesis of midfrontal theta power reflecting conflict, we performed the test contrasting bias-incongruent than bias-congruent actions to the theta range (4–8Hz). Furthermore, since visual inspection of the condition-specific time courses of theta power suggested major differences in theta power between Go and NoGo responses, we additionally performed an exploratory test contrasting Go and NoGo responses. Given its exploratory nature, this test was performed on broadband power (1–15 Hz).

2.3.12 fMRI-informed EEG analysis

The sluggish nature of the BOLD signal makes it difficult to determine when exactly different brain regions become active. In contrast, EEG provides much higher temporal resolution. Identifying distinct EEG correlates of the BOLD signal in different regions could thus reveal when these regions become active (Hauser et al. 2015). Furthermore, using the BOLD signal from different regions in a multiple linear regression allows to control for variance that is shared among regions (e.g., changes in global signal; variance due to task regressors) and test which region is the best unique predictor of a certain EEG signal. In such an analysis, any correlation between EEG and BOLD signal from a certain region reflects an association above and beyond those induced by task conditions.

To link BOLD signal from distinct regions to time-frequency power, we applied the same approach as for BOLD-RT correlation analyses and fitted a trial-by-trial hemodynamic response function (HRF) to the BOLD signal in selected ROIs. We then used the trial-by-trial HRF amplitudes to predict TF power at each time-frequency-channel bin (Hauser et al. 2015)(building on existing code from https://github.com/tuhauser/TAfT). For BOLD signal extraction, we specified six ROIs using a combination of functional and anatomical constraints based on our fMRI GLM results: vmPFC (valence contrast), ACC (valence and action contrast), left and right motor cortex (response side contrast, which captured lateralized motor activity better than the action contrast), and striatum (action contrast). For ACC, we obtained two separate masks (valence and action contrast), which strongly overlapped. Analyses included only one of those masks at a time; conclusions were identical with either mask. The resultant masks were back-transformed to each participant's native space and the first eigenvariate of the signal in each ROI was extracted, returning one summary measure of BOLD signal in that ROI per volume. We applied the same high-pass filter, nuisance regression, upsampling procedure, and trial-by-trial HRF estimation as in BOLD-RT correlation analyses. The trial-wise HRF amplitude estimates for each ROI were used as regressors in a multiple linear regression to predict the TF power for each 3D time-frequency-channel bin across trials, resulting in a 3D map of regression weights (*b*-map) for each ROI. In these regressions, we also added behavioral regressors for the main effects of required

action, valence, and their interaction as covariates of no interest to account for task-related variance in EEG power. In all analyses, predictors and outcomes were demeaned so that any intercept was zero. Finally, participants' *b*-maps were Fisher-z transformed (which makes the sampling distribution of correlation coefficients approximately normal and allows to combine them across participants).

Finally, to test whether BOLD signal in certain ROIs uniquely predicted variance in TF power, we performed cluster-based one-sample permutation *t*-tests across participants (Hunt et al. 2013). We performed these tests on the mean regression weights of the channels that exhibited condition differences in the EEG-only analyses (FCz and Cz; Fz was dropped because it did not show significant power differences between conditions) in the range of 0–1300 ms (i.e., duration of cue presentation), 1–15 Hz. We first obtained a null distribution of maximal cluster mass statistics from 10000 permutations. In each permutation, the sign of the *b*-map of a random subset of participants was flipped. Then, a separate *t*-test for each time-frequency bin (bins of 25 ms, 1 Hz) across participants was computed. The resulting *t*-map was thresholded at $|t| > 2$, from which we computed the maximal cluster mass statistic (i.e., sum of all *t*-values) of any cluster (i.e., adjacent voxels above threshold). We next computed a *t*-map for the real data, from which we identified the cluster with largest cluster mass statistic. The corresponding *p*-value was computed as the number of permutations with larger maximal cluster mass than the maximal cluster mass of the real data, and considered significant for a p-value of < .05.

2.4 RESULTS

Thirty-six healthy participants performed a motivational Go/NoGo learning task. In this task, they needed to learn by trial and error which response (left Go/ right Go/NoGo) to make in order to gain rewards ("Win" cues) or avoid losses ("Avoid" cues); see Fig 1. We simultaneously measured EEG and fMRI while participants performed this task.

2.4.1 Task performance

Participants successfully learned the task, as they performed significantly more Go actions to Go cues than NoGo cues (Required action: $\chi^2(1) = 32.01$, $p < .001$; Fig. 2.1D-E). Furthermore, participants showed a motivational bias, as they performed more Go actions for Win cues than Avoid cues (Valence: $\chi^2(1) = 23.70$, $p < .001$). The interaction of Required action x Valence was not significant ($\chi^2(1) = 0.20$, $p = .658$), suggesting that motivational biases occurred similarly for Go and NoGo cues.

When making a (Go) action, participants responded faster to Go cues (averaged over correct and incorrect responses) than to NoGo cues (where responses were by definition incorrect) (Required action: $\chi^2(1) = 25.64$, $p < .001$). Furthermore, a motivational bias was present also in RTs, with significantly faster actions to Win than Avoid cues (Valence ($\chi^2(1) = 44.58$, $p < .001$). Again, the interaction was not significant ($\chi^2(1) = 1.51$, $p = .219$; see Fig. 2.1F).

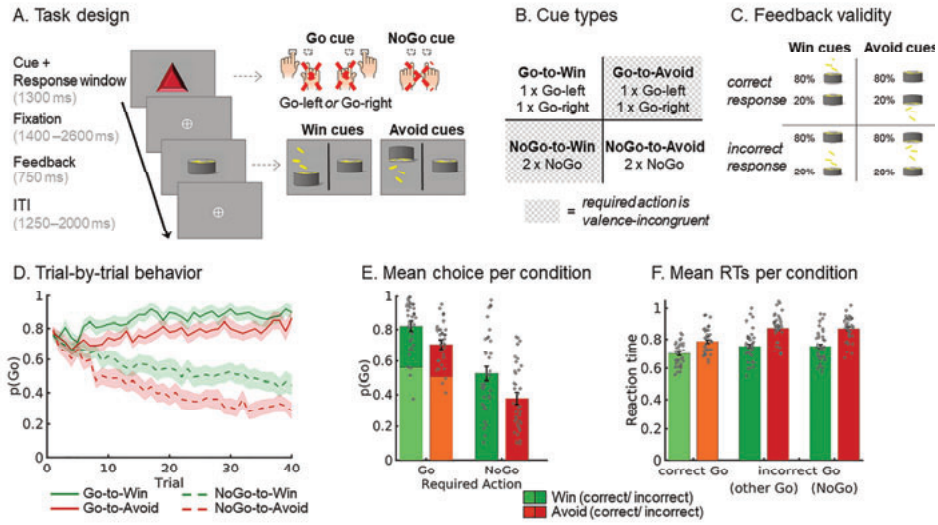


Figure 2.1. Motivational Go/NoGo learning task and performance.
Motivational Go/NoGo learning task design and performance. **A.** On each trial, a Win or Avoid cue appears; valence of the cue is not signaled but should be learned. Participants should respond during cue presentation. Response-dependent feedback follows after a jittered interval. Compared to a previous study with the same task (Swart et al. 2018), fixations between cue and feedback were jittered and 700–1900 ms longer and ITIs were 250 ms longer to allow us to disentangle cue- and outcome-related activity in the fMRI signal. Each cue has only one correct action (Go-left, Go-right, or NoGo), which is followed by the desired outcome 80% of the time. For Win cues, actions can lead to rewards or neutral outcomes; for Avoid cues, actions can lead to neutral outcomes or punishments. **B.** There are eight different cues, orthogonalizing cue valence (Win versus Avoid) and required action (Go versus NoGo). The motivationally incongruent cues, for which the motivational action tendencies are incongruent with the instrumental requirements, are highlighted in gray. **C.** Feedback is probabilistic: Correct actions to Win cues lead to rewards in 80% of cases, but neutral outcomes in 20% of cases. For Avoid cues, correct actions lead to neutral outcomes in 80% of cases, but punishments in 20% of cases. For incorrect actions, these probabilities are reversed. Rewards and punishments are depicted by money falling into/ out of a can. **D.** Trial-by-trial proportion of Go actions (±SEM) for Go cues (solid lines) and NoGo cues (dashed lines). D) Trial-by-trial proportion of Go actions (±SEM) for Go cues (solid lines) and NoGo cues (dashed lines). Shadows indicate standard errors for per-condition-per-participant means across participants using the Cousineau-Morey method (Morey 2008). The motivational bias is defined as the tendency to make more Go actions to Win than Avoid cues (i.e., green lines are above red lines). Additionally, participants clearly learn whether to make a Go actions or not (solid lines go up, dashes lines go down). Orange asterisks below indicate trial-by-trial significance of motivational bias, blue asterisks indicate performance accuracy above chance (i.e., correct Go_{Left}, Go_{Right} or NoGo response) (light color: $p < .05$ uncorrected; dark color: $p < 0.0013$; Bonferroni corrected for number of trials). **E.** Mean (±SEM) proportion Go actions per cue condition (points are individual participants' means). Proportion Go actions is higher for Go than NoGo cues, indicative of task learning, and higher for Win than Avoid cues, reflecting the influence of motivational biases on behavior. **F.** Mean (±SEM) reaction times for correct and incorrect Go actions, the latter split up in whether the other Go response or the NoGo response would have been correct (points are individual participants' means). Participants respond faster on correct than on incorrect Go actions and faster to Win than Avoid cues, reflecting the influence of motivational biases on behavior.

2.4.2 fMRI

Valence. We hypothesized higher BOLD signal in the striatum for Win compared with Avoid cues (Fig. 2.2A). There were no significant clusters in the striatum in a whole-brain corrected analysis. When restricting our analyses to an anatomical mask of the striatum, there were two significant clusters (for a complete list of all significant clusters and *p*-values, see S2.4): BOLD signal in left posterior putamen was significantly higher for Win than Avoid cues (Fig. 2.2E; no longer significant when excluding the five participants that were excluded from the final EEG-

fMRI analysis; S2.1). In contrast, BOLD signal in the bilateral medial caudate nucleus was, surprisingly, higher for Avoid than Win cues (Fig. 2.2F). While the effect in left putamen appeared robust over the time course of the task, the effect in medial caudate was strongest at the beginning of the task, but then disappeared towards the end of the task (see S2.5). At a whole-brain level cluster correction, the largest cluster of higher BOLD signal for Win than Avoid cues was observed in the ventromedial prefrontal cortex (vmPFC). Positive valence coding was also observed in bilateral dorsolateral prefrontal cortex (dlPFC), bilateral ventrolateral prefrontal cortex (vlPFC), posterior cingulate cortex (PCC), bilateral amygdala, and bilateral hippocampus (Fig. 2.2A). Conversely, BOLD was higher for Avoid cues in dorsal ACC and bilateral insula. Repeating analyses on correct trials only yielded identical results in the whole-brain analyses, while effects posterior putamen and medial caudate with small-volume correction were not significant anymore (see S2.6).

RTs. We reasoned that any region translating cue valence into motivational biases in behavior should also predict the speed-up of RTs for Win compared with Avoid cues. Our finding that vmPFC and ACC BOLD reflected cue valence is in line with a wealth of previous literature (Haber and Knutson 2010; Bartra et al. 2013), raising the question whether signals from these two regions impact action selection in a way that gives rise to motivational biases. We reasoned that if such signals have a causal effect on behavior, they should predict RT differences, i.e., both overall RT differences between Win and Avoid cues, but also RT differences within each valence condition. To assess whether BOLD signal in vmPFC and ACC indeed related to the speed of selected actions, we computed correlations of reaction times with the trial-by-trial deconvolved BOLD signal in the identified vmPFC and ACC clusters. To account for overall differences between Win and Avoid cues, we standardized RTs and BOLD signal separately for Win and Avoid cues, removing the overall difference between both conditions. There was a strong negative association for vmPFC, $t(33) = -4.11$, $p < 0.001$, $d = -0.71$, with higher vmPFC BOLD predicting faster reaction times, and a strong positive association for ACC, $t(33) = 7.83$, $p < .001$, $d = 1.34$, with higher ACC BOLD predicting slower reaction times (Fig. 2.2H). These results are consistent with vmPFC and ACC, both signaling cue valence, influencing the speed of downstream actions and thus contributing to motivational biases in reaction times. However, we cannot infer a causal role of these regions from the observed correlation.

Action. We further hypothesized that the valence signal in the striatum would be modulated by the congruency between valence and action, such that increased striatal BOLD signal for Win cues would be dampened when (bias-incongruent) NoGo responses were required, while decreased striatal activity for Avoid cues should be elevated when (bias-incongruent) Go actions were required (Fig. 2.2A). This interaction effect between valence and congruency is equivalent to a main effect of action.

Striatal BOLD (bilateral caudate nucleus, putamen, and nucleus accumbens) was indeed significantly higher for Go than NoGo responses (Fig. 2.2C). This effect was absent on the first block, but strongly emerged over time (see S2.5). However, in absence of a clear effect of valence on striatal BOLD, this main effect of action might not reflect an attenuation of valence signaling. Rather, the striatum appears to predominantly encode action itself—even in the left posterior putamen, which significantly encoded valence (Fig. 2.2E). This finding replicates previous studies using a different version of the Motivational Go/NoGo task (Guitart-Masip, Fuentemilla, et al.

2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012). Other regions that responded more strongly to Go versus NoGo responses included the ACC, thalamus and bilateral cerebellum. For a complete list of significant clusters, see S2.4. Repeating analyses on correct trials only yielded identical results (see S2.6).

Valence x Action interaction (Congruency). Based on prior work, we expected increased BOLD for bias-incongruent compared to bias-congruent actions in midfrontal cortex, the putative cortical source of midfrontal theta oscillations. At a whole-brain cluster level significance correction, there were no clusters in which BOLD signal differed between bias-congruent and -incongruent actions. When restricting the analysis to a mask comprising ACC and pre-SMA, there was a cluster in pre-SMA (Fig. 2.2D, G; not significant when excluding the five more participants excluded in the final EEG-fMRI analysis; see S2.1; not significant in the GLM on correct trials only, see S2.6). This finding is in line with source reconstruction studies of midfrontal theta (Cohen and Ridderinkhof 2013) and EEG-fMRI findings observing choice conflict-related activity in pre-SMA (Frank et al. 2015). The effect of conflict in pre-SMA was robust over time (see S2.5).

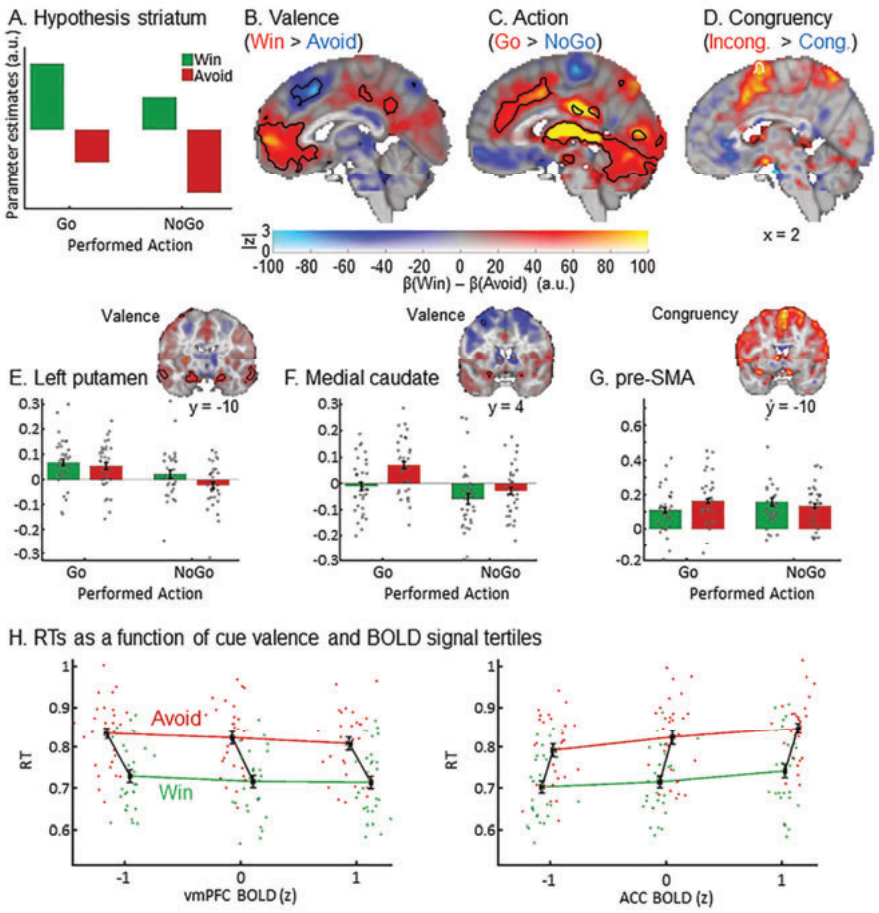


Figure 2.2. fMRI results.
BOLD signal as a function of cue valence, performed action, and congruency. **A.** We hypothesized striatal BOLD to encode cue valence (main effect of valence), with an attenuation of this valence signal when actions incongruent to the bias-triggered actions were performed (main effect of action). **B.** BOLD signal was significantly higher for *Win* compared to *Avoid* cues in ventromedial prefrontal cortex (vmPFC; whole brain corrected) and left putamen (small-volume corrected), but higher for *Avoid* compared to *Win* cues in ACC and medial caudate (small-volume corrected). **C.** BOLD signal was significantly higher for *Go* compared with *NoGo* responses in the entire striatum as well as ACC, thalamus, and cerebellum (all whole-brain corrected). **D.** BOLD signal was significantly higher for bias-incongruent actions than bias-congruent actions in pre-SMA (small-volume corrected). **B-D.** BOLD effects displayed using a dual-coding visualization with color indicating the parameter estimates and opacity the associated *z*-statistics. Contours indicate statistically significant clusters ($p < .05$), either small-volume corrected (striatal and SMA contours explicitly linked to a bar plot) or whole-brain corrected (all other contours). **E-G.** Mean beta weights per task condition (x-axis) per participant (individual grey dots) in significant clusters in left putamen, medial caudate and pre-SMA (significant in small-volume correction). **E.** Left posterior putamen encoded valence positively (higher BOLD for Win than Avoid cues), but was dominated by an encoding of the performed action (higher BOLD for Go than NoGo responses). **F.** Medial caudate encoded valence negatively (higher BOLD for Avoid than Win cues), again predominantly showing a main effect of action. **G.** BOLD signal in pre-SMA was higher for bias-incongruent than bias-congruent actions (small-volume corrected). **H.** Reaction times (RTs) as a function of cue valence and BOLD signal tertiles (z-standardized) per participant (individual dots; x-location relative to all other participants). RTs were significantly predicted by BOLD signal in vmPFC (positively) as well as by BOLD signal in ACC and striatum (negatively). BOLD-RT correlations were independent of cue valence. Lines connect the means of RT tertiles. Error bars (±SEM) for both BOLD (vertically) and RTs (horizontally) are very narrow.

2.4.3 EEG

Action x Valence interaction (Congruency). We used a non-parametric, cluster-based permutation test to test whether time-frequency (TF) power in the theta band (4–8 Hz) was significantly higher on motivational incongruent (Go2Avoid, NoGo2Win) than congruent (Go2Win, NoGo2Avoid) trials over midfrontal channels (Fz, FCz, Cz). We indeed found that theta power was higher on incongruent than congruent trials (p = .023), most strongly around 175–325 ms after cue onset. However, this difference occurred markedly earlier than in our previous study (450–650 ms) (Swart et al. 2018) and visual inspection of the time-frequency plot showed that the peak of this cluster was located rather in the alpha band (8–12 Hz), leaking into the upper theta range (Fig. 2.3E-F), and restricted to an early, transient increase in alpha power over midfrontal channels. This congruency effect was not present in evoked activity (ERPs; see S2.7) and indeed remained unaltered after the subtraction of evoked activity (see S2.8). Furthermore, this change in alpha power occurred selectively on incongruent trials on which participants made a correct response, rather than nonspecifically on all incongruent or on all correct trials (see S2.9). In other words, power increased selectively when biases were successfully suppressed (Swart et al. 2018). In sum, we found an electrophysiological correlate of conflict, which was however different in time and frequency range from our previous finding (Swart et al. 2018).

Action and Valence main effects. We observed a modulation of time-frequency power by valence-action congruency in the alpha range shortly after cue onset (around 175–325 ms). We next performed permutation tests to explore whether time-frequency power was further modulated at any later time point by the individual task factors rather than their interaction (i.e., action or valence). Broadband power (1–15 Hz) was significantly higher on trials with Go actions than NoGo responses (cue-locked: p = .002; response-locked: p = .002): This difference between Go and NoGo responses occurred as a broadband-signal from 1–15 Hz, but peaked in the beta band (cue-locked) and theta band (response-locked; Fig. 2.3B and D). The topographies exhibited a bimodal distribution with peaks both at frontopolar (FPz) and central (FCz, Cz, CPz) electrodes (Fig. 2.3B and 3D). As visual inspection of Fig. 2.3C shows, theta power increased in all conditions until 500 ms post cue onset and then bifurcated depending on the action: For NoGo responses, power decreased, while for Go actions, power kept rising and peaked at the time of the response. This resulted in higher broadband power for Go versus NoGo responses for about 550–1300 ms after cue onset (see Fig. 2.3C-D; around -200–400 ms when response-locked, see Fig. 2.3A-B; same held in analyses across both correct and incorrect trials, see S2.10). When looking at the cue-locked signal, the signal peaked earlier and higher for Go actions to Win than to Avoid cues, in line with faster reaction times on Go2Win than Go2Avoid trials. When testing for differences in broadband power between Win cues and Avoid cues, broadband power was indeed higher for Avoid than Win cues around 825–1300 ms cue-locked (p = .002). When comparing power time courses for Go responses to Win and to Avoid cues selectively in the theta range, theta power was higher for Win than Avoid cues around 550–700 ms, p = 0.028, but then higher for Avoid than Win cues around 925–1300 ms, p = .002. This difference in latency and peak height of the ramping signal was not present in the response-locked signal, and the respective test of Win vs. Avoid cues not significant (p = .110; Fig. 2.3A and S2.11).

The occurrence of this signal close to response execution across a broadband frequency range raised the question whether it reflects (a) a decision process (incorporating decision parameters like cue valence and action values) or rather (b) a generic motor signal occurring for any manual response, or even (c) a signal artifact of head motion in the scanner. Regarding the latter, a number of control analyses indicated that the theta increase was likely not reducible to a motor artifact: first, results remained unchanged when accounting for fMRI-realignment parameters (reflecting head motion) using a linear regression approach (following (Fellner et al. 2016); see S2.12), and second, relative to the pre-trial baseline, increases were clearly focused on the theta band (see S2.13), started already 300 ms after cue onset, and even occurred on NoGo trials (i.e., where no overt response was executed). Furthermore, the theta signal was unlikely to reflect a generic motor response, as it was modulated by task demands: Theta (but not broadband) power was higher for left, non-dominant hand compared to the right, dominant hand (in line with reaction time findings; see S2.14), and higher for correct than incorrect responses (see S2.11). Taken together, we cannot exclude the possibility that motor execution processes or signal artifacts contributed to the observed differences in theta power; however, differences in theta are likely not reducible to such processes, and do at least in part reflect pre-response, decision-related processes.

Next, if midfrontal theta power reflects a decision process, we asked whether this signal bore resemble to evidence accumulation processes described in perceptual decision-making before (Gold and Shadlen 2007; O'Connell et al. 2012). In such processes, a response is elicited once an accumulation signal reaches a certain threshold. This observation was particularly the case for the theta band (Fig. 2.3C) rather than any other band (see alpha band in Fig. 2.3E). Three further tests corroborated this interpretation: First, the peak of the ramping theta signal in the cue-locked data predicted reaction times within participants (see S2.11), while differences in peak height and latency were absent in the response-locked data. Thus, faster accumulation of evidence in favor of a 'Go' response for 'Win' cues could be the driving mechanism for a motivational bias. This change in accumulation rate would then be putatively driven by a neural region encoding cue valence, such as vmPFC or ACC. Second, the "threshold" that theta needed to reach to elicit a response was higher for responses of the left (non-dominant) than the right (dominant) hand (in line with reaction times findings; S2.14), putatively reflecting response competition in which the dominant hand needs to be overruled by raising response thresholds. Third, peak theta was lower for incorrect than correct responses (S2.11), consistent with the idea that incorrect responses might (sometimes) reflect premature responding due to random fluctuations in response thresholds (O'Connell et al. 2012). These three observations are consistent with an interpretation of the ramping theta signal as reflecting an evidence accumulation process for Go-related evidence.

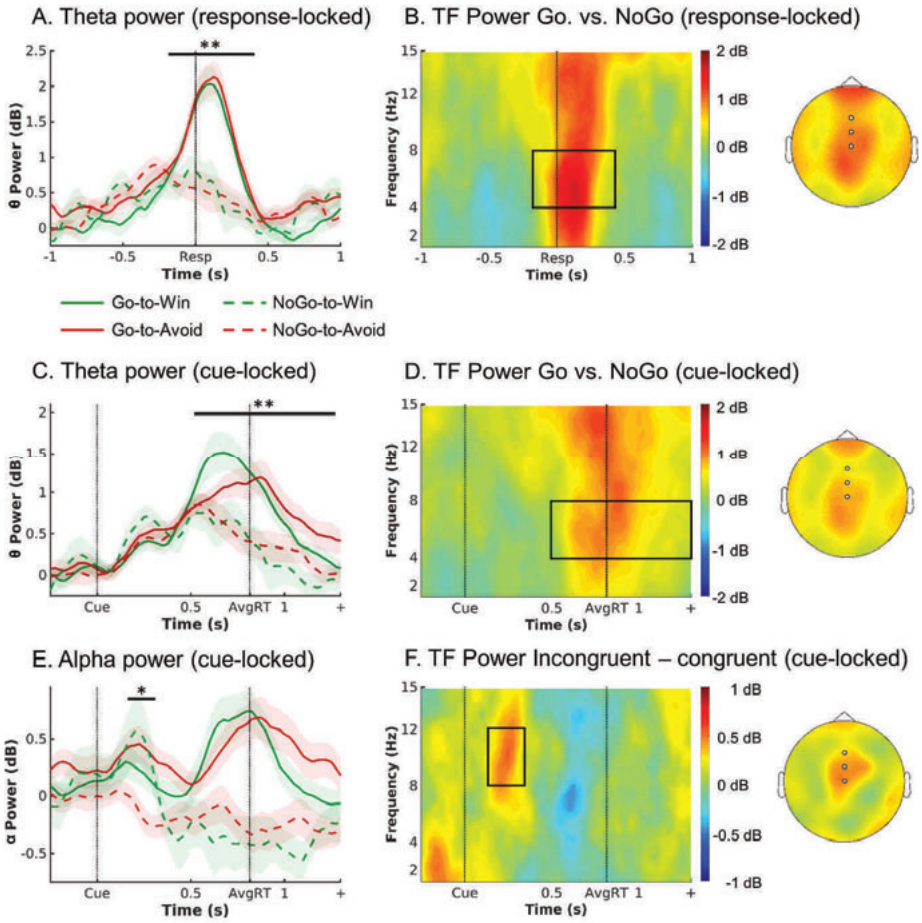


Figure 2.3. EEG results.
EEG time-frequency power as a function of cue valence and action. **A.** Response-locked within trial time course of average theta power (4–8 Hz) over midfrontal electrodes (Fz/ FCz/ Cz) per cue condition (correct-trials only). Theta increased in all conditions relative to pre-cue levels, but to a higher level for Go than NoGo trials. There were no differences in theta peak height or latency between Go2Win and Go2Avoid trials. **B.** Left: Response-locked time-frequency power over midfrontal electrodes for Go minus NoGo trials. Go trials featured higher broadband TF power than NoGo trials. The broadband power increase for Go compared to NoGo trials is strongest in the theta range. Right: Topoplot for Go minus NoGo trials. The difference is strongest at FZ and FCz electrodes. **C-D.** Cue-locked within trial time course and time-frequency power. Theta increased in all conditions relative to pre-cue levels, but to a higher level for Go than NoGo trials, with earlier peaks for Go2Win than Go2Avoid trials. **E.** Trial time course of average alpha power (8–13 Hz) over midfrontal electrodes per cue condition (correct trials only; cue-locked). Alpha power transiently increases for both incongruent conditions in an early time window (around 175–325 ms). **F.** Left: Time-frequency plot displaying that the transient power increase was focused on the alpha band (8 – 13 Hz), leaking into the upper theta band. Right: Topoplot of alpha power displaying that this incongruency effect was restricted to midfrontal electrodes (highlighted by white disks). * $p <$ 0.05. ** $p < 0.01$. Shaded errorbars indicate (±SEM). Box in TF plots indicates the time frequency window where *t*-values > 2.

2.4.4 fMRI-informed EEG analysis

Given that we observed action encoding in both BOLD (ACC, striatum) and midfrontal EEG power (theta), we next tested when differences in activity in these regions occurred by correlating the BOLD signal from those regions with midfrontal EEG time-frequency power. We extracted trial-by-trial BOLD signal from clusters encoding valence (vmPFC, ACC), action (striatum, ACC) and response hand (left and right motor cortex) and used these signals in a multiple linear regression to predict midfrontal time-frequency power.

First, trial-by-trial analyses revealed that striatal BOLD correlated positively with theta power around the time of response, over and above task and condition effects (p = .037 cluster-corrected; Fig. 2.4A). Conversely, and perhaps surprisingly, there was no such correlation between midfrontal theta power and ACC BOLD (mask based on action contrast: p = .268 cluster-corrected; mask based on valence contrast: p = .592). Supplementary analyses yielded correlations between motor cortex BOLD and midfrontal beta power (Jurkiewicz et al. 2006; Ritter et al. 2009), corroborating the overall ability of our approach in detecting well established BOLD-EEG associations (S2.15). Note that because the design matrix included all ROI timeseries as well as task regressors, these correlations only reflect variance uniquely explained by a specific ROI, over and above task effects in these regressors.

Next, to investigate whether vmPFC BOLD, which encoded cue valence rather than action, contributes to action selection in fronto-striatal circuits, we also assessed an association between midfrontal EEG power and BOLD in the vmPFC. Trial-by-trial deconvolved BOLD signal from the valence cluster in vmPFC correlated negatively with broadband power (p = .043, cluster-corrected) in an early cluster (around 2–15 Hz, peak in the upper theta range around 6–7 Hz, 0–0.4 sec., Fig. 2.4C). Regression of the time-domain EEG voltage on vmPFC BOLD yielded a negative association of vmPFC BOLD with a left frontal P2 component, which however showed a different topography and was unlikely to explain the negative vmPFC-theta correlations over midfrontal electrodes (see S2.16).

Finally, we performed complementary EEG-informed fMRI analyses using the trial-by-trial midfrontal alpha and theta signals identified in EEG-only analyses as regressors on top of the task regressors. While fMRI-informed EEG analyses allow to test which region is the best predictor of a certain EEG signal (competing with BOLD signal from other regions), this EEG-informed fMRI analysis allows to assess whether networks of several regions might reflect trial-by-trial changes in power (although none of them might do so uniquely). Alpha power correlated negatively with dlPFC and SMG BOLD, while theta power correlated positively with BOLD signal in bilateral pre-SMA, ACC, motor cortices, operculum, putamen, and cerebellum, corroborating the association between theta and motor regions, including the striatum (see S2.17). See S2.17 for a methodological discussion on these seemingly contrasting findings.

In sum, the fMRI-informed EEG analyses show that the amplitude of the ramping theta signal at the time of response correlates positively with striatal BOLD signal. In contrast, theta power early after cue onset correlates negatively with vmPFC activity. Finally, ACC activity in the cluster encoding Go vs. NoGo responses was not significantly linked to theta power, while other parts of ACC and pre-SMA (and further motor regions) were in fact related to trial-by-trial theta power. Taken together, we observed an early negative correlation of theta with vmPFC BOLD, which

encoded cue valence, and a late positive correlation of theta (ramping up to the response) with striatal BOLD, which encoded the selected action.

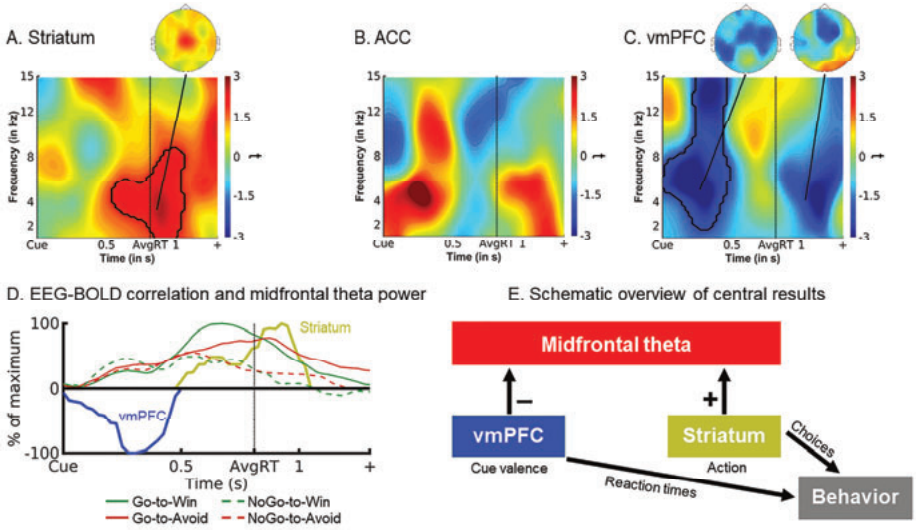


Figure 2.4. fMRI-informed EEG results.
Uniquely explained variance in EEG time-frequency power over midfrontal electrodes (FCz/Cz) by BOLD signal from (**A**) whole striatum, (**B**) ACC, and (**C**) vmPFC to the average Group-level *t*-maps display the modulation of the EEG time-frequency power by trial-by-trial BOLD signal in the selected ROIs. Striatal (but not ACC) BOLD correlates most strongly with theta/delta power around the time of response. vmPFC BOLD correlates with broadband (peak: theta) power soon after cue onset. Areas surrounded by a black edge indicate clusters of $|t| > 2$ with $p < .05$ (cluster-corrected). Topoplots indicate the topography of the respective cluster. Note that there are no significant clusters in ACC, but given our a-priori hypothesis regarding the relation between ACC BOLD and midfrontal theta, for completeness, we include this visualization. **D.** Time course of vmPFC and striatal BOLD correlations with theta power (*t*-values from clusters above threshold extracted and summed over frequencies), normalized to the peak of the time course of each region, overlaid with theta power for each valence x action condition. vmPFC-theta correlations emerge when theta is still similar for each condition, while striatum-theta correlations emerge when theta rises more strongly for Go than NoGo responses. **E.** Schematic overview of our main EEG-fMRI results: Both vmPFC and striatum modulate midfrontal theta power (striatum likely indirectly via motor areas, see S2.17). BOLD signal in both regions predicts the amplitude of theta power—the vmPFC early and negatively, the striatum late and positively. We speculate that the vmPFC encodes cue valence and sends this information to the striatum, where valence information biases the motivation for active responses in recurrent fronto-striatal loops and thus gives rise for motivational biases in behavioral responses and reaction times.

2.4.5 Summary of the main results

In sum, we observed motivational biases in both choice and reaction time data. Cue valence, which drives these biases, led to differences in BOLD signal in vmPFC (higher for Win cues) and ACC (higher for Avoid cues). These vmPFC and ACC BOLD responses to cue valence also predicted reaction times on a trial-by-trial basis. Striatal BOLD did not respond to cue valence, but predominantly reflect the action participants selected (higher for Go than NoGo). Motivational conflict was associated with higher BOLD in midfrontal cortex, though only weakly, and with early transient midfrontal alpha power. In contrast, midfrontal theta power reflected the selected action (higher for Go than NoGo) around the time of responses. Finally, trial-by-trial vmPFC BOLD correlated negatively with theta power early after cue onset, while striatal BOLD correlated positively with theta power around the time of responses.

2.5 DISCUSSION

The main aim of this study was to investigate the interaction of striatal and midfrontal processes during the suppression of motivational biases, using combined EEG-fMRI. We and others have previously found elevated theta power when such biases were successfully suppressed (Cavanagh et al. 2013; Swart et al. 2018). Furthermore, computational models of basal ganglia loops posited the origin of this bias in the striatum (Frank 2005; Collins and Frank 2014). We thus hypothesized that theta power would correlate with attenuated striatal valence coding during bias-incongruent actions. However, we found that both striatal and theta signals were strongly dominated by action per se, rather than by a combination of valence and action, as was predicted for the striatum, or by valence-action congruency, as was predicted for the midfrontal cortex. Most importantly, we found that the time course of theta power exhibited several features of a process accumulating evidence whether to select an active Go response or not. Interestingly, valence-driven vmPFC BOLD signal uniquely predicted variability in mid-frontal theta immediately following cue presentation, while action-driven striatal BOLD responses uniquely predicted variability in midfrontal theta around the time of the response. Taken together, these results suggest that striatum and vmPFC may act in concert to evaluate the value of performing an active Go response, i.e., the "value of work".

2.5.1 Striatal BOLD reflects motivation for action

We found that striatal BOLD signal was strongly dominated by the action participants performed (Go vs. NoGo) rather than cue valence. This finding replicates previous studies using a very similar task ((Guitart-Masip, Fuentemilla, et al. 2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012) and highlights the role of the striatum in value-based behavioral activation and invigoration (Taylor and Robbins 1984, 1986; Robbins and Everitt 1992, 2007; Niv et al. 2007; Salamone and Correa 2012; Howe and Dombeck 2016; Syed et al. 2016; Coddington and Dudman 2018, 2019; da Silva et al. 2018), yet appears to be at odds with the role of the striatum in reward expectation (Doya 2000; Dayan and Daw 2008; Collins and Frank 2014). A recent theory aimed to reconcile these roles by proposing that striatal (dopamine) signals do not reflect the value of anticipated outcomes per se, but rather the value of performing an action to obtain this outcome, i.e., the "value of work" (Berke 2018). Recent empirical work in rodents has indeed shown selective ramping of striatal dopamine signals for active responses approaching a goal state (Hamid et al. 2016, 2021; Syed et al. 2016; Mohebi et al. 2019). In light of these findings, it seems plausible that the striatum evaluates whether there are sufficient incentives to overcome a NoGo default and to instead take action to achieve a valuable goal. Any signal that reflects such an evolving value of Go should ramp over the trial time course and peak when a response is elicited, like an evidence accumulation process. While the BOLD response is too sluggish for capturing such fast within-trial signals, EEG can provide insight.

2.5.2 Late midfrontal theta power reflects striatal activity

In this study, midfrontal theta power, like striatal BOLD, was modulated by whether participants made an active Go response or not. Theta power bifurcated for Go and NoGo responses, peaking around the time of the response. Striatal BOLD and midfrontal theta signals were strongly linked, such that trial-by-trial fluctuations in striatal (rather than ACC or motor cortex) BOLD was the best predictor of trial-by-trial fluctuations in theta power around the time of response.

The observed link between striatal BOLD and midfrontal theta may seem surprising given that previous EEG source localization studies of conflict-related midfrontal theta power modeled a source in ACC or pre-SMA (Hanslmayr et al. 2008; Cohen and Ridderinkhof 2013), and previous resting-state EEG-fMRI studies reported negative correlations of frontal theta with regions of the default-mode network (Scheeringa et al. 2008, 2009). Our study might fill a blind spot in these literatures: by recording EEG-fMRI during a decision task, we show that theta power increases commonly observed in such tasks might reflect subcortical action selection processes. Arguably, the striatum is far away from the scalp and thus unlikely to be the direct neural source of midfrontal theta oscillations. It is possible that striatal action selection processes modulate activity in parts of midfrontal (or motor) cortex, reflected in the amplitude of theta power over the scalp (see also our EEG-informed fMRI analyses in S2.17). This finding suggests that scalp EEG can give insights into evolving action selection in the striatum, which is not visible in resting-state recordings, but can only be studied using appropriate task designs.

In contrast to our study, previous findings have reported elevated theta power mostly in situations of cognitive conflict (Cavanagh and Frank 2014; Cohen 2014), including our own EEG study using the same task (Swart et al. 2018). Importantly, however, these studies usually observe strong theta rises for *any* action, with conflict-induced theta constituting a minor increase on top of this much larger rise (Cohen and Cavanagh 2011; Swart et al. 2018). While conflict-induced theta was absent in our data, action-induced theta was strongly present—which might be especially visible in Go/NoGo tasks such as in this study, but concealed in other paradigms that only feature active responses. We speculate that both phenomena are related and reflect the evolving value of making a Go action: theta power rises prior to Go actions, but even further in situations of cognitive conflict. If response thresholds in striatal pathways are elevated during conflict, theta may not reflect a cortical top-down "trigger" that drives threshold elevation, but rather the extra bits of accumulated evidence in the striatum that follow from such elevated response thresholds. This alternative account of midfrontal theta power modulation provides a putative unifying explanation for both action- and conflict-induced theta increases.

2.5.3 Early vmPFC valence signals shape action selection

So far, we have suggested that striatal BOLD and theta power signals reflect how evidence for action is accumulated, putatively reflecting the value of work (i.e., the physical effort of taking action). This interpretation leaves open what drives this evidence accumulation—and does so differently for reward and punishment prospects, leading to the observed expression of motivational biases in behavior. Any neural "source" of these biases should show differential activity in response to Win and Avoid cues. While valence coding was weak and spatially heterogeneous in the striatum, it clearly emerged in vmPFC (positively) and ACC (negatively). Particularly the vmPFC appears to be a likely candidate source of motivational biases given that a wealth of previous studies (Haber and Knutson 2010; Bartra et al. 2013) has shown vmPFC BOLD to encode the expected outcomes. In behavior, cue valence affected both the probability and the speed of making a Go response, as people responded more often and faster to Win cues than to Avoid cues. A follow-up trial-by-trial analysis showed that the vmPFC is likely involved in eliciting this motivational bias, as fluctuations in vmPFC signal predicted response times also within each valence condition. This finding is consistent with the idea that valence information in this region feeds into fronto-striatal loops and gives rise to motivational biases in behavior. Of note, the region

of ACC encoding cue valence did not significantly correlate with midfrontal theta power, even though trial-by-trial ACC BOLD did correlate with RTs. Taken together, in line with past theories of recurrent fronto-striatal loops (Alexander et al. 1986; Mink 1996; Middleton and Strick 2000; Gurney et al. 2001), our result suggest that vmPFC encodes cue valence at an early time point and then biases the motivation for active responses, i.e., the value of work (Hamid et al. 2016; Berke 2018), in the striatum.

vmPFC BOLD correlated negatively with midfrontal theta power very early after cue onset (Fig. 2.4C), consistent with previous EEG-fMRI findings (Scheeringa et al. 2008, 2009; Hauser et al. 2015) as well as electrophysiological data in humans (Harris et al. 2011; Hunt et al. 2012) and animals (Van Wingerden et al. 2010; Vinck et al. 2010; Seo et al. 2012; Knudsen and Wallis 2020). The vmPFC shows a more positive BOLD response to Win (compared to Avoid) cues. The observed negative vmPFC-theta correlations are in line with previous findings showing that both vmPFC and midfrontal theta power encode valence, though with opposite signs: vmPFC BOLD is typically higher for positive than negative events, while the opposite holds for midfrontal theta (and midfrontal BOLD signal) (Shackman et al. 2011; Cavanagh, Figueroa, et al. 2012; Cavanagh, Zambrano-Vazquez, et al. 2012; Braem et al. 2017). Our results indicate that vmPFC encodes cue valence very soon after this information becomes available, indexed in midfrontal theta power.

In contrast to the negative vmPFC-theta correlation immediately following cue onset, action-related theta and positive striatum-theta correlations occurred later, around the time of response (Cohen 2014). Although both vmPFC and striatal BOLD correlated with power in the same frequency band, correlations may well reflect different neural processes. Compared to vmPFC-theta correlations, striatum-theta correlations and in particular action-related theta exhibited a more centroparietal (rather than midfrontal) topography of rather short duration (for a discussion of the burst-like modulations of ongoing theta oscillations, see (Cohen 2014). Of note, timing and topography of action-related theta in the EEG-only analysis were similar to the "centroparietal positivity"/ P300, which has been suggested to reflect perceptual evidence accumulation (O'Connell et al. 2012; Kelly and O'Connell 2013; Philiastides et al. 2014; Twomey et al. 2015). However, this action-related theta signal was not visible in cue-locked ERP analyses and thus apparently not phase-locked (see S2.08). Taken together, early vmPFC-theta correlations likely reflect cue valence processing, while late striatum-theta correlations likely reflect the motivation for a final action.

2.5.4 Caveats and open questions

A priori, we expected elevated midfrontal theta power in situations of motivational incongruency between biases and required actions. Instead of theta power, we observed a transient increase in midfrontal alpha power, which specifically occurred on incongruent trials on which participants successfully overcame biases. Trial-by-trial midfrontal alpha power was negatively correlated with BOLD signal in dorsolateral prefrontal cortex and supramarginal gyrus. While we are not aware of previous literature reporting elevated frontal alpha in conflict situations, the observed associations with BOLD in regions of the fronto-parietal attention network might point at midfrontal alpha reflecting an unspecific mechanism of focused attention and increased task engagement (Brier et al. 2010; Harris et al. 2013; Helfrich et al. 2018) which might help to retrieve and focus on learned stimulus-response associations (Buschman et al. 2012). Of note, while heightened attention is typically associated with decreased (posterior) alpha, there have been

findings of increased alpha, as well (van der Meij et al. 2016). Nonetheless, future research is needed to understand the role midfrontal alpha might play in overcoming motivational conflict.

The absence of elevated theta power in situations of cognitive conflict could perhaps be due to relatively low performance in the current study compared to previous studies using the same task (Swart et al. 2017, 2018). This reduced performance is likely due to the fMRI environment and associated necessary task changes (longer and jittered response-outcome intervals), and may explain the inconsistency with previous findings (Swart et al. 2018). If participants learned the required action of a cue less well, they would be less aware of conflict between motivational bias and action requirements. Furthermore, reduced performance resulted in relatively fewer trials on which participants successfully overcame motivational biases, leaving less statistical power to test neural hypotheses on bias suppression. This might explain why we did not observe conflict-related midfrontal theta increases and why evidence for increased BOLD signal in pre-SMA during conflict was rather weak. Thus, our findings do not undermine the role of theta in motivational conflict, but rather highlight a putatively complementary role of theta in the motivation of actions.

Finally, the response-locked nature of the observed theta power difference raises the question whether this finding might simply reflect motor execution. Control analyses indicated that the signal was modulated by response hand and accuracy, which can be expected from a signal that reflects decision variables such as response conflict or threshold variability, but not from a signal that reflects simple motor execution or even a signal artifact. Moreover, the observation that striatal BOLD was the best predictor of trial-by-trial midfrontal theta power (rather than BOLD in motor cortices) again speaks for theta reflecting a pre-response, decision-related processes.

2.5.5 Conclusion

In sum, participants in this simultaneous EEG-fMRI study exhibited strong motivational biases and relatively poor instrumental learning in a motivational Go-NoGo learning task. This feature likely has rendered our set-up suboptimal for isolating the predicted fronto-striatal mechanisms involved in suppressing motivational biases. However, the presence of these strong (relatively uncontrolled) motivational biases enabled us to further dissect the mechanisms of bias expression. Specifically, the finding that, despite strong valence effects on behavior, striatal BOLD indexed the selected action (rather than cue valence) indicates that the striatum is unlikely to play a role in generating the motivational bias. Rather, striatal BOLD might selectively reflect the motivation to show an active Go response. The finding of strong cue valence signaling in the vmPFC, which also predicted reaction times on a trial-by-trial basis, suggests that the motivational bias might instead arise from the vmPFC. The negative association of this vmPFC signal with early midfrontal theta suggests that the vmPFC processes valence information very early after cue onset and may subsequently shape action selection. One putative mechanism through which the vmPFC could shape action selection is the modulation of the rate of evidence accumulation towards a Go response. Besides the negative correlation of vmPFC BOLD with midfrontal theta power early after cue onset, the positive correlation of striatal BOLD with late midfrontal theta power concurs with a complementary role for the striatum in the eventual decision to execute an active response. Together, these findings suggest a dual nature of midfrontal theta power, with early components reflecting valence processing in the vmPFC and late components reflecting motivation for action in the striatum. Taken together, our results are in line with "value of work" theories of the role of fronto-striatal loops in the evaluation of whether to perform an active response.

2.6 SUPPLEMENTARY MATERIALS FOR CHAPTER 2

2.6.1 S2.1: Behavioral, fMRI, and EEG analyses with only the 30 participants included in EEG-fMRI analyses

We repeated the behavioral, fMRI, and EEG analyses reported in the main text while excluding the six participants that were also not included in the fMRI-inspired EEG analyses reported in the main text: two participants due to fMRI co-registration failure (which were also not included in the fMRI-only analyses), and five further participants due to large outliers on the *b*-maps in the fMRI-inspired EEG analyses.

In this subgroup, similar to the entire sample, participants performed significantly more Go responses to Go cues than NoGo cues (Required action: $\chi^2(1) = 25.77$, $p < .001$; see Fig. S2.1A panels A-B). Furthermore, participants showed a motivational bias, as they performed more Go actions for Win cues than Avoid cues (Valence: $\chi^2(1) = 18.87$, $p < .001$). The interaction of Valence x Required Action was not significant ($\chi^2(1) = 0.13$, $p = .910$). Similarly, when making a (Go) response, participants responded faster when this was correct (Go cues; irrespective of whether a left or right hand response required) than when this was incorrect (NoGo cues) (Required action: $\chi^2(1) = 21.02$, $p < .001$). Furthermore, a motivational bias was present also in RTs, with significantly faster responses to Win than Avoid cues (Valence: $\chi^2(1) = 37.31$, $p < .001$). Again, the interaction was not significant ($\chi^2(1) = 2.25$, $p = .134$; see Fig. S2.1A panel C). In sum, all behavioral results also held in this subsample.

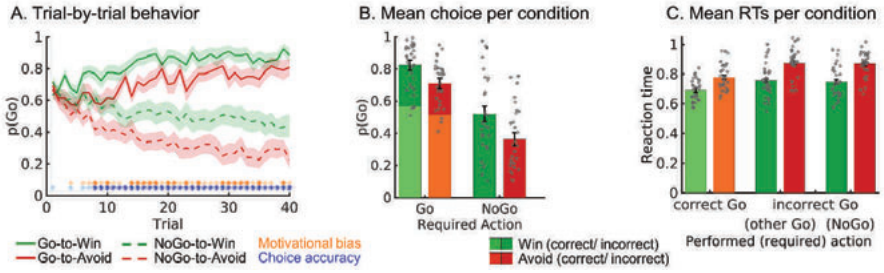


Figure 2.5. S2.1A. Motivational Go/NoGo learning task performance in the subgroup of 29 participants included in the fMRI-inspired EEG analyses.

A. Trial-by-trial proportion of Go responses (±SEM) for Go cues (solid lines) and NoGo cues (dashed lines). Shadows indicate standard errors for per-condition-per-participant means across participants using the Cousineau-Morey method (Morey 2008). The motivational bias is defined as the tendency to make more Go actions to Win than Avoid cues (i.e., green lines are above red lines). Additionally, participants clearly learn whether to make Go actions or not (solid lines go up, dashes lines go down). Orange asterisks below indicate trial-by-trial significance of motivational bias, blue asterisks indicate performance accuracy above chance (i.e., correct Go_{Left}, Go_{Right} or NoGo response) (light color: $p < .05$ uncorrected; dark color: $p < 0.0013$; Bonferroni corrected for number of trials). **B.** Mean (±SEM) proportion Go responses per cue condition (points are individual participants' means). Proportion Go responses is higher for Go than NoGo cues, indicative of task learning, and higher for Win than Avoid cues, reflecting the influence of motivational biases on behavior. **C.** Mean (±SEM) reaction times for correct and incorrect Go responses, the latter split up in whether the other Go response or the NoGo response would have been correct (points are individual participants' means). Participants respond faster on correct than on incorrect Go responses and faster to Win than Avoid cues, reflecting the influence of motivational biases on behavior.

In our fMRI analyses, when testing for differences in BOLD signal between Win and Avoid cues, there were again no significant clusters in the striatum in a whole-brain corrected analysis. When restricting our analyses to an anatomical mask of the striatum, BOLD signal in both left (z_{max} = 4.35, p = .00485, MNI peak coordinates: xyz = [-8 4 4]) and right medial caudate nucleus (z_{max} = 3.98, p = .0121, xyz = [12 6 6]) was higher for Avoid than Win cues as reported in the main text (Fig. S2.1B panel F), while differences in left posterior putamen reported in the main text were not significant any more (Fig. S2.1B panel E). Furthermore, at a whole-brain level cluster correction, BOLD signal was again higher for Win compared to Avoid cues in vmPFC (z_{max} = 5.09, p = 2.24e-16, xyz = [-6 44 2]), dlPFC (z_{max} = 5.01, p = .000142, xyz = [16 48 48]), PCC (z_{max} = 4.34, p = .000297, xyz = [6 -44 32]), and left amygdala/ hippocampus (z_{max} = 4.36, p = .0162, xyz = [-20 -2 -22]). There were additional clusters in left vlPFC (z_{max} = 4.30, p = .00702, xyz = [28 36 -10]), left middle temporal gyrus (z_{max} = 3.76, p = .014, xyz = [-62 -18 -12]),) right middle temporal gyrus (z_{max} = 3.92, p = .014, xyz = [62 -18 -6]), left angular gyrus (z_{max} = 4.99, p = 5.07e-6, xyz = [-44 -56 20]), and right amygdala/ hippocampus (z_{max} = 4.52, p = .0223, xyz = [20 -6 -20]) not featured in the results in the main text (Fig S2.1B panel B). Conversely, in line with results reported in the main text, BOLD was higher for Avoid stimuli in dorsal ACC (z_{max} = 4.12, p = 1.15e-05, xyz = [2 36 46]), left insula (z_{max} = 4.47, p = .0014, xyz = [-28 22 0]), left superior frontal gyrus (z_{max} = 4.12, p = .00398, xyz = [-22 -4 56]), right insula (z_{max} = 4.07, p = .00148, xyz = [32 26 0]), left vlPFC (z_{max} = 4.47, p = .00697, xyz = [-32 62 8]), right superior frontal gyrus (z_{max} = 3.99, p = .00588, xyz = [22 -4 52]), and right precuneous (z_{max} = 4.60, p = .00156, xyz = [8 -64 54]), in line with results reported in the main text (see Fig. 2.2B). In addition, we observed clusters in left frontal pole (z_{max} = 4.67, p = .00702, xyz = [-32 62 8]) and left angular gyrus (z_{max} = 3.98, p = .0162, xyz = [-38 -56 48]).

When correlating RTs and BOLD signal, there was again a significantly negative correlation between RTs and vmPFC BOLD, $t(29)$ = -3.89, $p < 0.001$, d = -0.71, and a significantly positive correlation between RTs and ACC BOLD, $t(29)$ = 7.41, $p < 0.001$, d = 1.35 (see Fig. S2.1B panel H).

When testing for differences in BOLD signal between responses, we observed again significantly higher BOLD signal for *Go* than *NoGo* action in the entire striatum (bilateral caudate nucleus, putamen, and nucleus accumbens), thalamus, and bilateral cerebellum (z_{max} = 7.05, p = 0, xyz = [-12 -24 10]), ACC (z_{max} = 6.87, p = 9.04e-05, xyz = [0 8 42]), left motor cortex (z_{max} = 4.59, p = .00915, xyz = [-54 -22 -24], Fig. S2.1B panel C), in line with results reported in the main text. Furthermore, there was an additional (separate) cluster in left frontal pole (z_{max} = 4.23, p = .0124, xyz = [-28 42 6]. Again, there were not clusters with higher BOLD signal for NoGo than Go responses (Fig S2.1B panel C).

When testing for differences in BOLD signal between bias-incongruent and bias-congruent actions, there were no clusters at a whole-brain cluster level significance correction, and—different from results report in the main text—also not when restricting the analysis to a mask comprising ACC and pre-SMA (Fig S2.1B panels D and G).

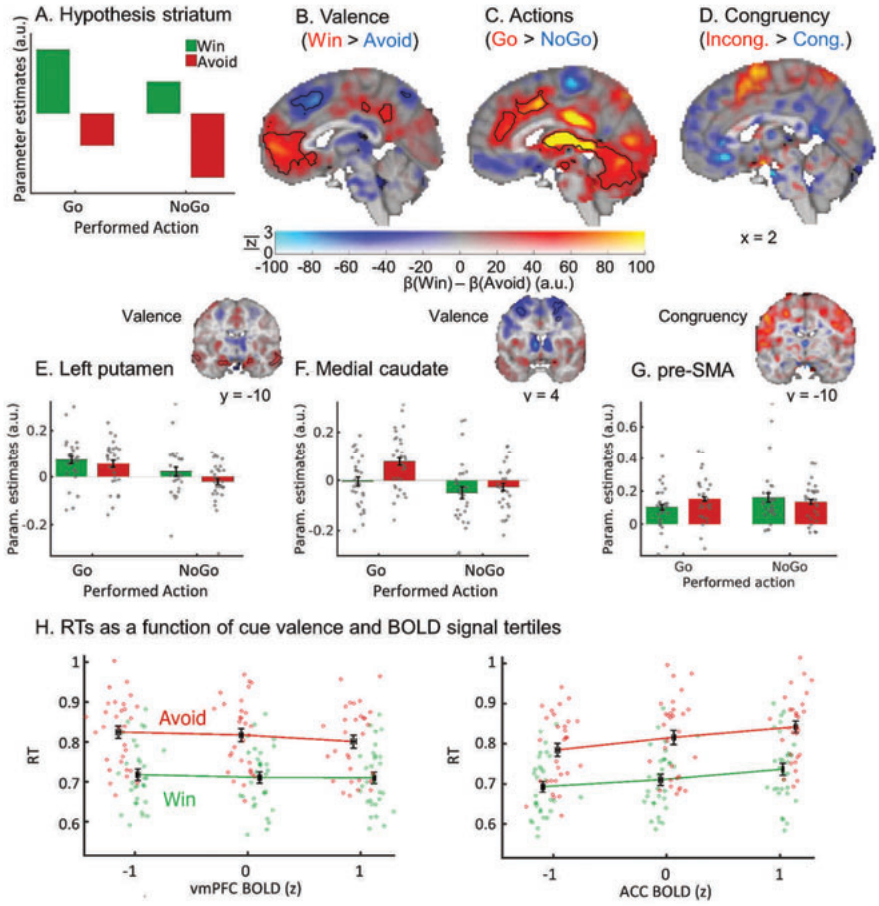


Figure 2.6. S2.1B. BOLD signal as a function of cue valence, performed action, and congruency in the subgroup of 29 participants included in the fMRI-inspired EEG analyses.
A. We hypothesized striatal BOLD to encode cue valence (strong main effect of valence), with an attenuation of this valence signal when actions incongruent with bias-triggered actions were performed (weak main effect of action). **B.** BOLD signal was significantly higher for *Win* compared to *Avoid* cues in ventromedial prefrontal cortex (vmPFC; whole brain corrected) and left putamen (small-volume corrected), but higher for *Avoid* compared to *Win* cues in ACC and medial caudate (small-volume corrected). **C.** BOLD signal was significantly higher for *Go* compared to *NoGo* actions in the entire striatum as well as ACC, thalamus, and cerebellum (all whole-brain corrected). **D.** Based on the plot, it appears that BOLD signal was higher for bias-incongruent actions than bias-congruent actions in pre-SMA, but contrary to the results reported in the main text, this was not significant. **B-D**. BOLD effects displayed using a dual-coding data visualization approach with color indicating the parameter estimates and opacity the associated *z*-statistics. Black contours indicate statistically significant clusters ($p < .05$, whole-brain corrected). **E-G.** Mean beta weights per task condition (x-axis) per participant (individual grey dots) in significant clusters in left putamen, medial caudate and pre-SMA (significant in small-volume correction). **E.** It appears that left posterior putamen encoded valence positively (higher BOLD for Win than Avoid cues), but contrary to results reported in the main text, this was not significant. **F.** Medial caudate encoded valence negatively (higher BOLD for Avoid than Win cues). **G.** Extracted BOLD signal from pre-SMA to illustrate (non-significant) congruency effects. **H.** Reaction times (RTs) as a function of cue valence and BOLD signal tertiles (z-standardized) per participant (individual dots; x-location relative to all other participants). RTs were significantly predicted by BOLD signal in vmPFC (positively) as well as by BOLD signal in ACC and striatum (negatively). BOLD-RT correlations were independent of cue valence.

In our EEG analyses, there was no significant difference between incongruent than congruent trials in the theta band (p = .236). In the alpha band, it was marginally significant (p = .052), most strongly around 200–325 ms after cue onset (Fig. S2.1C panels E-F). A permutation test on the broadband (1–15 Hz) TF power yielded a significant result (p = 0.046).

Furthermore, broadband power (1–15 Hz) was significantly higher on trials with Go actions than NoGo actions (cue-locked: p = .002, around 550–1300 ms after cue onset, see Fig. S2.1C panels A-B; response-locked: p = .002, around -150–425 ms relative to responses, see Fig. S2.1C panels C-D). Overall, all EEG results also held in this subsample.

Taken together, behavioral and EEG analyses yielded identical conclusions as results reported in the main text. In the fMRI analyses, differences between Win and Avoid cues in left posterior putamen and differences in pre-SMA between bias-incongruent and bias-congruent actions were not significant, while all other results were still significant and yielded identical conclusions as results reported in the main text.

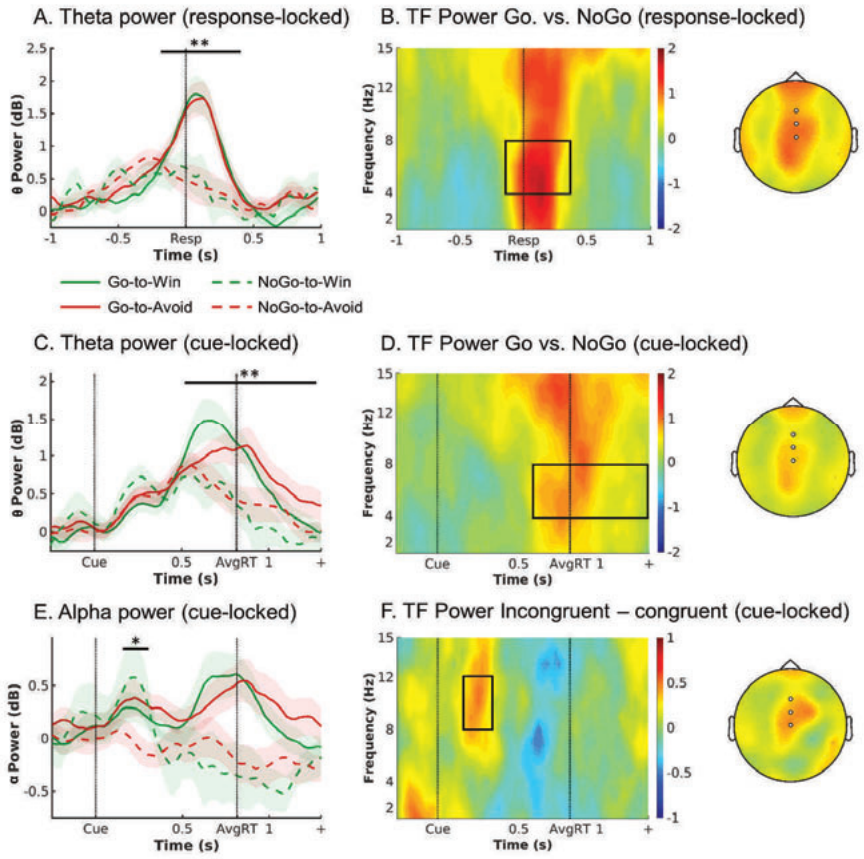


Figure 2.7. S2.1C. EEG time-frequency power as a function of cue valence and action.
A. Response-locked within trial time course of average theta power (4–8 Hz) over midfrontal electrodes (Fz/ FCz/ Cz) per cue condition (correct-trials only). Theta increased in all conditions relative to pre-cue levels, but to a higher level for Go than NoGo trials. There were no differences in theta peak height or latency between Go2Win and Go2Avoid trials. **B.** Left: Response-locked time-frequency power over midfrontal electrodes for Go minus NoGo trials. Go trials featured higher broadband TF power than NoGo trials. The broadband power increase for Go compared to NoGo trials is strongest in the theta range. Right: Topoplot for Go minus NoGo trials. The difference is strongest at Cz and FCz electrodes. **C-D.** Cue-locked within trial time course and time-frequency power. Theta increased in all conditions relative to pre-cue levels, but to a higher level for Go than NoGo trials, with earlier peaks for Go2Win than Go2Avoid trials. **E.** Trial time course of average alpha power (8–13 Hz) over midfrontal electrodes per cue condition (correct trials only; cue-locked). Alpha power transiently increases for both incongruent conditions in an early time window (around 175–325 ms). **F.** Left: Time-frequency plot displaying that the transient power increase was focused on the alpha band (8–13 Hz), leaking into the upper theta band. Right: Topoplot of alpha power displaying that this incongruency effect was restricted to midfrontal electrodes (highlighted by white disks). * $p < 0.05$. ** $p < 0.01$. Shaded error bars indicate (±SEM). Box in TF plots indicates the time frequency window where t-values > 2.

2.6.2 S2.2: Anatomical masks (for small-volume corrected analyses) and conjunctions of anatomical and functional masks (for fMRI-informed EEG analyses)

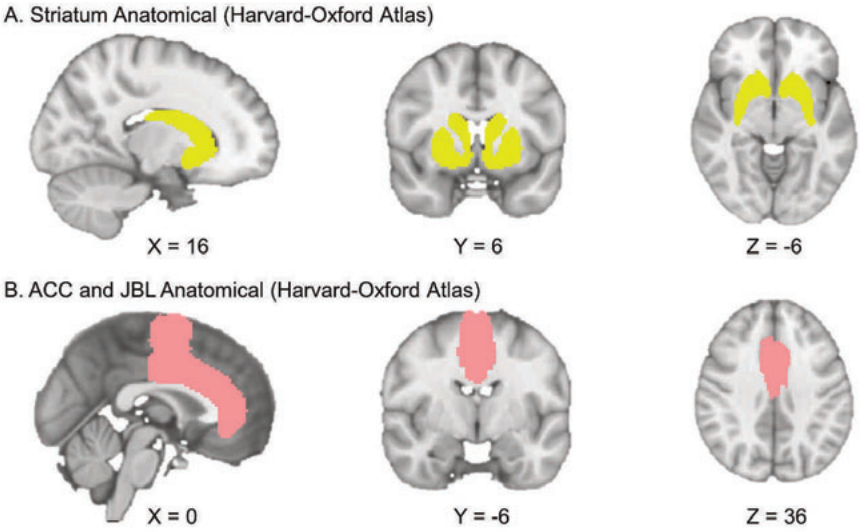


Figure 2.8. S2.2A. Anatomical masks used for small-volume corrected GLM analyses.
Anatomical masks of (**A**) striatum (yellow, conjunction of bilateral nucleus accumbens, caudate, and putamen) and (**B**) midfrontal cortex (pink, cingulate cortex anterior and juxtapositional lobule cortex) used for small-volume corrected GLM analyses. All masks were extracted from the probabilistic Harvard-Oxford Atlas, thresholded at 10%. Note that images are in radiological orientation (i.e., left brain hemisphere presented on the right and vice versa).

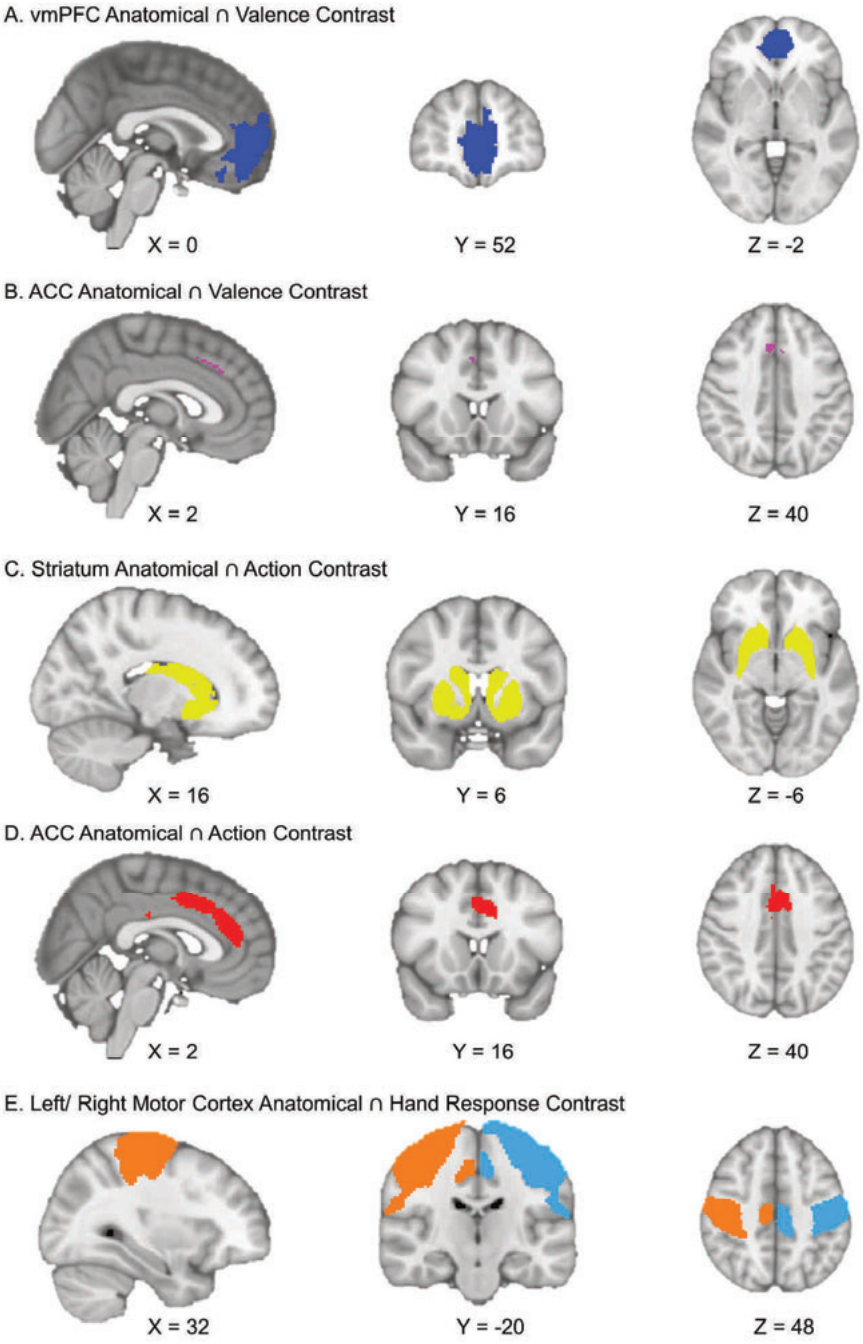


Figure 2.9. S2.2B. Conjunctions of anatomical masks (based on the Harvard-Oxford Atlas) and functional contrasts from fMRI GLM analyses (Valence, Action, and Hand Response contrasts) used for fMRI-informed EEG analyses.

A. vmPFC valence contrast (dark blue, conjunction of frontal pole, frontal medial cortex, and paracingulate gyrus. **B.** ACC valence contrast striatum (purple, cingulate cortex anterior). **C.** Striatum action contrast (yellow, conjunction of bilateral nucleus accumbens, caudate, and putamen). **D.** ACC action contrast (red, cingulate cortex anterior). **E.** Left (light blue) and right (orange) motor cortices hand response contrast (conjunction of precentral gyrus and postcentral gyrus) used for fMRI-informed EEG analyses. All anatomical masks were extracted from the probabilistic Harvard-Oxford Atlas, thresholded at 10%. Note that images are in radiological orientation (i.e., left brain hemisphere presented on the right and vice versa).

2.6.3 S2.3: Regressors and contrasts in fMRI analyses

Regressors:

- Win2GoOnset: for every trial with Win cue and Go action, at cue onset, duration 1, value +1
- Win2NoGoOnset: for every trial with Win cue and NoGo action, at cue onset, duration 1, value +1
- Avoid2GoOnset: for every trial with Avoid cue and Go action, at cue onset, duration 1, value +1
- Avoid2NoGoOnset: for every trial with Avoid cue and NoGo action, at cue onset, duration 1, value +1
- Handedness: for every trial, at cue onset, value +1 for left hand response, 0 for NoGo response, -1 for right hand response
- Error: for every trial, at cue onset, value +1 for incorrect response, 0 for correct response
- OutcomeOnset: for every trial, at outcome onset, duration 1, value +1 for every trial
- OutcomeValence: for every trial, duration 1, value +1 for positive outcome (reward, no punishment), -1 for negative outcome (no reward, punishment)
- InvalidOutcome: for trials where uninstructed button was pressed, at outcome onset, duration 1, value 1

Nuisance Regressors:

- 6 realignment parameters (obtained from co-registration) per volume
- Mean white-matter signal per volume
- Mean out-of-brain signal per volume
- Separate regressor for each volume where relative displacement > 2mm

Regressor	1	2	3	4	5	6	7	8	9
Contrast	Win2GoOnset	Win2NoGoOnset	Avoid2GoOnset	Avoid2NoGoOnset	Handedness	Error	OutcomeOnset	OutcomeValence	InvalidOutcome
1. Valence	1	1	-1	-1					
2. Executed Action	1	-1	1	-1					
3. Conflict	-1	1	1	-1					
4. Hand					1				

2.6.4 S2.4: Significant BOLD clusters in the valence, action, and congruency contrasts

Contrast Brain region	*Z*-value	Cluster size (voxels)	Corrected *p*	Peak coordinates x	y	z
Win > Avoid cues						
Ventromedial prefrontal cortex, caudal anterior cingulate gyrus	5.23	3533	1.06e-18	2	34	12
Left angular gyrus, left supramarginal gyrus	4.83	749	2.21e-06	-42	-54	18
Right dorsolateral prefrontal cortex	4.98	599	1.95e-05	16	46	48
Right ventrolateral prefrontal cortex	4.56	596	2.04e-05	30	34	-12
Right supramarginal gyrus, right middle temporal gyrus	4.14	510	7.68e-05	66	-42	10
Posterior cingulate gyrus	4.28	460	.000172	8	-32	36
Left hippocampus, left parahippocampal gyrus, left amygdala	4.72	366	.00085	-18	-6	-24
Left middle temporal gyrus	3.86	346	.00121	-60	-18	-12
Left precentral gyrus	4.12	303	.00268	-34	-12	70
Left ventrolateral prefrontal cortex	3.97	251	.00734	-40	36	-14
Right hippocampus, right parahippocampal gyrus, right amygdala	4.66	240	.00916	20	-6	-20
Left posterior middle temporal gyrus	3.88	204	.0194	-62	-46	-8
ROI in striatum:						
Left putamen	4.00	78	.00979	-28	-10	6
Avoid > Win cues						
Anterior cingulate cortex, superior frontal gyrus	4.38	690	5.13e-06	2	36	46
Left angular gyrus, left superior parietal lobule, left supramarginal gyrus	4.33	428	.000292	-38	-56	48
Left insula, left frontal operculum	3.98	303	.00268	34	24	0
Right insula, right frontal operculum	4.63	292	.0033	-28	24	0
Left ventrolateral prefrontal cortex	4.72	291	.00336	-32	62	8

Right middle frontal gyrus	4.11	213	.016	-20	-2	52
Left precuneous	4.81	207	.0182	8	-66	54
ROI in striatum:						
Left medial caudate	4.27	79	.00979	-6	4	2
Right medial caudate	3.9	56	.0194	12	8	0
Go > NoGo actions						
Cerebellum, bilateral thalamus, bilateral putamen, bilateral caudate, bilateral Nucleus Accumbens, posterior cingulate cortex, anterior cingulate cortex, paracingulate gyrus, bilateral ventrolateral frontal cortex	7.49	26731	0	26	-48	-28
Bilateral precuneous	5.29	595	.000141	-10	-62	38
Left postcentral gyrus, left central operculum	4.69	354	.00378	-54	-22	22
NoGo > Go actions						
No significant clusters						
Incongruent > Congruent actions						
No significant clusters						
ROI in ACC & pre-SMA:						
Pre-SMA	3.68	132	0.00431	4	4	66
Congruent > incongruent actions						
No significant clusters						
ROI in ACC & pre-SMA:						
No significant clusters						

Table S2.4. Significant clusters in the valence, action and congruency contrasts in the fMRI GLM.

2.6.5 S2.5: Changes in effects on fMRI BOLD signal over time

After identifying BOLD correlates of cue valence, performed action, and motivational conflict in the whole-brain and small-volume-corrected GLM analyses reported in the main text, we were interested in whether these effects change over the time course of the experiment. For this purpose, we extracted the first eigenvariate of the BOLD signal from the significant clusters above threshold (see Fig. 2.2 in the main text; for masks, see S2.2), fitted an HRF to each trial to obtain the trial-by-trial HRF amplitude (identical procedure to BOLD-RT correlations and fMRI-informed EEG analyses), and analyzed these amplitude as a function of the respective behavioral variable (cue valence, performed action, or motivational conflict), trial number, and their interaction, using mixed-effects linear regression.

Specifically, for vmPFC, ACC, left putamen and medial caudate signal, we fitted the following model (Wilkinson notation):

$$BOLD \sim cueValence * trialNumber + (cueValence * trialNumber|participant)$$

vmPFC signal was strongly modulated by cue valence, $\chi^2(1) = 31.313$, $p < .001$, with higher signal for Win than Avoid cues. In addition, the main effect of trial number was marginally significant, $\chi^2(1) = 3.351$, $p = .067$, with signal tending to increase over time. The interaction between valence and trial number was marginally significant as well, $\chi^2(1) = 2.959$, $p = .085$: The valence effect tended to decrease over time, driven by signal increasing for Avoid cues while staying at a constant high level for Win cues (Fig. S2.5 panel A).

ACC signal was also strongly modulated by valence, $\chi^2(1) = 15.213$, $p < .001$, with higher BOLD signal for Avoid than Win cues. There also was a significant main effect of trial number, $\chi^2(1) = 6.491$, $p = .011$, with signal decreasing over time. The interaction between valence and trial number was marginally significant, $\chi^2(1) = 2.935$, $p = .087$: The valence effect tended to decrease over time, driven by signal decreasing for Avoid cues while staying at a constant low level for Win cues (Fig. S2.5 panel B).

Signal in left putamen strongly encoded cue valence, $\chi^2(1) = 16.949$, $p < .001$, with higher signal for Win than Avoid cues. The main effect of trial number was not significant, $\chi^2(1) = 1.265$, $p = .261$, and neither was the interaction between valence and trial number, $\chi^2(1) = 1.544$, $p = .214$ (Fig. S2.5 panel C).

Signal in medial caudate strongly encoded cue valence, $\chi^2(1) = 17.330$, $p < .001$, with higher signal for Avoid than Win cues. The effect of trial number was just significant, $\chi^2(1) = 3.874$, $p = .049$, with signal decreasing over time. The interaction between valence and trial number was marginally significant, $\chi^2(1) = 3.769$, $p = .052$: The valence effect tended to decrease over time, driven by signal decreasing for Avoid cues while staying at a constant low level for Win cues (Fig. S2.5 panel D).

For striatal and ACC signal (different mask than for the valence signal reported above), we fitted the following model (Wilkinson notation):

$$BOLD \sim performedAction * trialNumber + (performedAction * trialNumber| participant)$$

For striatal signal, the main effect of action was not significant, $t(31.14) = 0.031$, $p = .975$, while the effect of trial number was strongly significant, $t(217.09) = -2.773$, $p = .006$, with signal

decreasing over time. The interaction was marginally significant, $t(54.61) = 1.736$, $p = .088$, driven by signal increasing for Go actions, but decreasing for NoGo actions, such that the action effect in striatum (higher signal for Go than NoGo actions) only emerged over time (because models using likelihood ratio tests failed to converge, *p*-values in this model are instead based on *t*-tests using Satterthwaite's method as implemented in the R package *lmerTest;* Fig. S2.5 panel E).

For ACC signal, the main effect of action was not significant, $\chi^2(1) = 0.270$, $p = .603$, while the main effect of trial number was significant, $\chi^2(1) = 5.342$, $p = .021$, reflecting overall decreasing signal over time. The interaction between action and trial number was not significant, $\chi^2(1) = 0.038$, $p = .845$. This inconsistency with our results in the whole-brain GLM analyses might reflect differential weighting of outliers and block-wise signal in FSL's FEAT vs. lme4's mixed effects models (Fig. S2.5 panel F).

For pre-SMA signal, we fitted the following model (Wilkinson notation):

$$BOLD \sim conflict * trialNumber + (conflict * trialNumber | participant)$$

For pre-SMA signal, there was a significant main effect of conflict, $\chi^2(1) = 5.064$, $p = .024$, with higher BOLD for bias-incongruent than -congruent action, and a significant negative effect of trial number, $\chi^2(1) = 10.530$, $p = .001$, with signal decreasing over time. The interaction between conflict and trial number was not significant, $\chi^2(1) = 0.142$, $p = .706$ (Fig. S2.5 panel G).

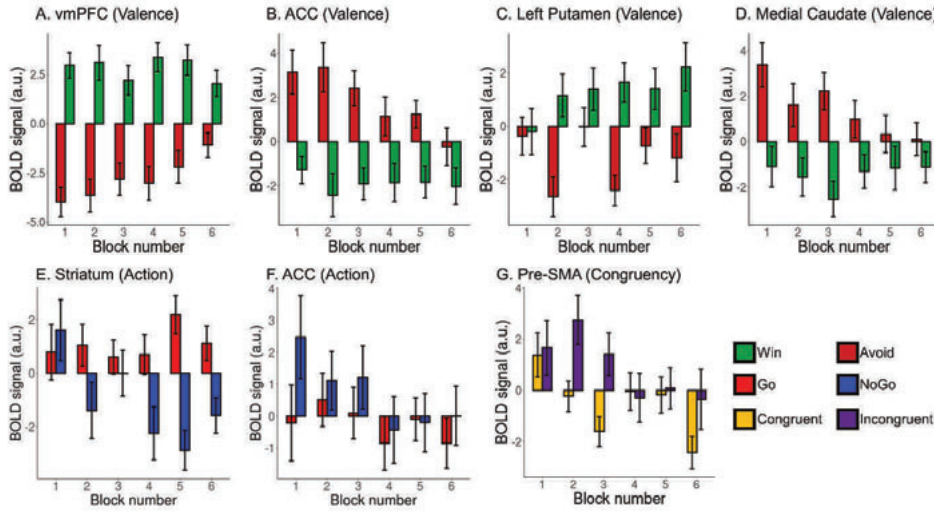


Figure 2.10. S2.5. Changes in BOLD signal differences over time.
BOLD signal in significant clusters identified with whole-brain and small-volume-corrected GLM analyses (see main text, Figure 2) as a function of behavioral variables (cue valence, executed action, motivational conflict) and block number. Bar represent means, whiskers present standard errors (per condition over participants, computed via the Cousineau-Morey method). **A.** vmPFC encoded cue valence (Win > Avoid), but this effect tended to decrease over time. **B.** ACC encoded cue valence (Avoid > Win), but this effect tended to decrease over time. **C.** Left putamen encoded cue valence (Win > Avoid). **D.** Medial caudate encoded cue valence (Avoid > Win), but this effect tended to decrease over time. **E.** Striatum encoded the performed action (Go > NoGo), but this effect only emerged over time. **F.** In contrast to whole-brain GLM analyses, ACC did not significantly encode the performed action. **G.** Pre-SMA encoded motivational conflict (congruency, incongruent > congruent).

2.6.6 S2.6: fMRI results for correct trials only

EEG and fMRI research have different analytical procedures of dealing with differences between correct and incorrect trials: While fMRI research typically uses multiple linear regression (GLMs), which allows to model error trials by a designated regressor, EEG research typically tests for differences between (categorical) conditions with a (mass-univariate) *t*-test approach. Because we used both approaches in the main text, here, for consistency, we also report fMRI results for regressors defined for correct trials only. Note that this analysis uses less trials than the one featured in the main text and thus has lower statistical power.

We fitted a GLM with eight task regressors, namely the four conditions resulting from crossing cue valence (Win/Avoid) and performed action (Go/NoGo irrespective of Left vs. Right Go) separately for correct and incorrect trials. We again added four regressors of no interest, namely response side (Go left = +1, Go right = -1, NoGo = 0), outcome onset (intercept of 1 for every outcome), outcome valence (reward = +1, punishment = -1, neutral = 0), and invalid trials (invalid buttons pressed and thus not feedback given). Note that compared to the GLM reported in the main text, we did not add an error regressor. This GLM failed to converge for one participant, leaving 33 participants in the group-level analysis.

When comparing BOLD signal between trials with Win cues and with Avoid cues, in the whole-brain corrected analysis, we again observed higher BOLD for Win cues in vmPFC (z_{max} = 5.20, p = 1.3e-9, xyz = [0 40 2]), as well as left superior lateral occipital cortex (z_{max} = 3.59, p = .00325, xyz = [-56 -64 30]), and left medial temporal gyrus (z_{max} = 3.63, p = .00474, xyz = [-70 -14 -14]; Fig. S2.6 panel A). Conversely, BOLD signal was higher for Avoid cues in left supramarginal gyrus (z_{max} = 3.91, p = .00235, xyz = [-36 -48 34]) and left ventrolateral prefrontal cortex (z_{max} = 2.34, p = .00453, xyz = [-26 58 4]) (Fig. S2.6 panel B). Note that higher activity in ACC for Avoid is clearly visible in Fig. S2.6 panel B (blue blob), but not statistically significant. Furthermore, analyses using small-volume correction on an anatomical mask of the striatum yielded no clusters of differential BOLD activity, also not in the regions reported in the main text, i.e., in left putamen (Fig. S2.6 panel E) nor medial caudate (Fig. S2.6 panel F). Overall, whole-brain results on correct trials only were similar to the results across both correct and incorrect trials reported in the main text, but weaker, suggesting that restricting analyses to correct trials only resulted in a considerable loss in statistical power.

When comparing trials with Go vs. NoGo actions, we observed higher BOLD signal for Go than NoGo actions in clusters in bilateral cerebellum, thalamus, striatum, and ACC (z_{max} = 7.01, p = 0, xyz = [-30 -50 -30]), right ventrolateral prefrontal cortex (z_{max} = 4.38, p = 1.79e-07, xyz = [34 48 6]), precuneous (z_{max} = 5.21, p = 1.97e-05, xyz = [-8 -64 38]), left operculum (z_{max} = 4.72, p = .000216, xyz = [-52 -22 18]), right supramarginal gyrus (z_{max} = 4.99, p = .000351, xyz = [-40 -50 36]), left precentral gyrus (z_{max} = 4.15, p = .00186, xyz = [-44 -18 62]), and right precentral gyrus (z_{max} = 3.98, p = .00758, xyz = [46 -18 66]; Fig. S2.6 panel C). This finding is in line with results across both correct and incorrect trials reported in the main text. Conversely, BOLD signal was higher for NoGo than Go trials in left inferior frontal gyrus (z_{max} = 4.43, p = .0128, xyz = [-58 26 20]), a finding not observed across both correct and incorrect trials reported in the main text.

Finally, when comparing both incongruent and congruent trials, there were again no significant clusters in a whole-brain corrected analysis (Fig. S2.6 panel D), and also not in an

analysis using small-volume correction on midfrontal cortex (Fig. S2. panel 6G). Again, this null result might be due to a considerate loss in power compared to the results across both correct and incorrect trials reported in the main text.

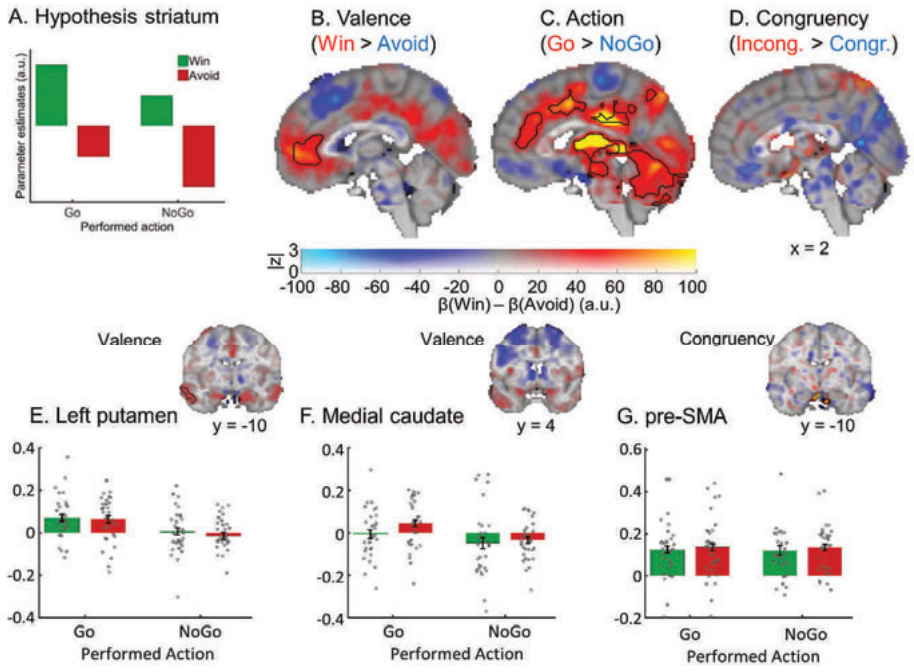


Figure 2.11. S2.6. BOLD signal as a function of cue valence, performed action, and congruency for correct trials only.
A. We hypothesized striatal BOLD to encode cue valence (main effect of valence), with an attenuation of this valence signal when actions incongruent to the bias-triggered actions were performed (main effect of action). **B.** BOLD signal was significantly higher for *Win* compared to *Avoid* cues in ventromedial prefrontal cortex (vmPFC; whole brain corrected), but in contrast to analyses across both correct and incorrect trials reported in the main text, BOLD was not significantly higher for *Avoid* compared to *Win* cues in ACC. **C.** BOLD signal was significantly higher for *Go* compared with *NoGo* actions in the entire striatum as well as ACC, thalamus, and cerebellum (all whole-brain corrected). **D.** BOLD signal was not significantly different between bias-incongruent actions (Go actions to Avoid cues and NoGo actions to Win cues) and bias-congruent actions (Go actions to Win cues and NoGo actions to Avoid cues), also not in the cluster in pre-SMA reported in the main text (small-volume corrected). **B-D.** BOLD effects displayed using a dual-coding data visualization approach with color indicating the parameter estimates and opacity the associated *z*-statistics. Contours indicate statistically significant clusters ($p < .05$), either small-volume corrected (striatal and SMA contours explicitly linked to a bar plot) or whole-brain corrected (all other contours). **E.** Numerically, left posterior putamen seemed to encode valence positively (higher BOLD for Win than Avoid cues), but in contrast to analyses across both correct and incorrect trials reported in the main text, this was not significant. (**F**) Numerically, medial caudate seemed to encode valence negatively (higher BOLD for Avoid than Win cues), but in contrast to analyses across both correct and incorrect trials reported in the main text, this was not significant. **G.** Extracted BOLD signal from pre-SMA to illustrate (the lack of) congruency effects.

2.6.7 S2.7: ERPs as function of action and valence

Given that the observed phasic alpha increase occurred soon after stimulus onset and much earlier than the theta effect in our previous study (Swart et al. 2018)—although more similar to the timing reported by (Cavanagh et al. 2013)—we investigated whether conditions differed in evoked rather and induced activity.

First, we again selected correct trials only, computed average ERPs for each condition per participant, and then tested for significant differences between the ERPs for incongruent and congruent trials using permutation tests on the average signal over midfrontal channels (Fz/ FCz/ Cz) in the time period of 0–700 ms post-cue (where evoked potentials occurred in the condition-averaged plot). We found no significant clusters in which the ERPs differed (no clusters above threshold; see Fig. S2.7A panels A and D). Visual inspection yielded an inconsistent picture such that, if anything, N1, N2 and P3 components tended to be slightly stronger on incongruent trials, while P2 components tended to be stronger on congruent trials. Numerically, when comparing all four conditions, the P2 seemed to be highest and the N2 lowest on Go2Avoid trials, which the opposite was the case for NoGo2Win trials (see Fig. S2.7B). Such opposite findings cannot explain why both conditions showed an increase in alpha power (see Fig. 2.3E in the main text), suggesting that the observed alpha power findings are not reducible to evoked activity.

Next, in line with the analyses in time-frequency space, we analyzed ERPs as a function of executed action and cue valence, contrasting trials with Go vs. NoGo actions and trials with Win vs. Avoid actions using permutation tests over the average signal of midfrontal electrodes (Fz/ FCz/ Cz). We found that ERPs differed significantly for Go vs. NoGo responses ($p = .008$) around 200–350 ms after cue onset, reflecting higher P2 (and lower N2) components for Go compared to NoGo responses (Fig. S2.7A panel B). The peak of the topography of this effect was over left and central frontal electrodes (Fig. S2.7A panel E).

When contrasting ERPs for Win vs. Avoid cues, we only obtained a marginally significant *p*-value of .068, which was driven by higher signal 420–470 ms after cue onset (Fig. S2.7A panels C and F). This difference occurred over midfrontal electrodes at the moment that the evoked signal rose towards the P3, but did not reflect differences in any of the component peaks.

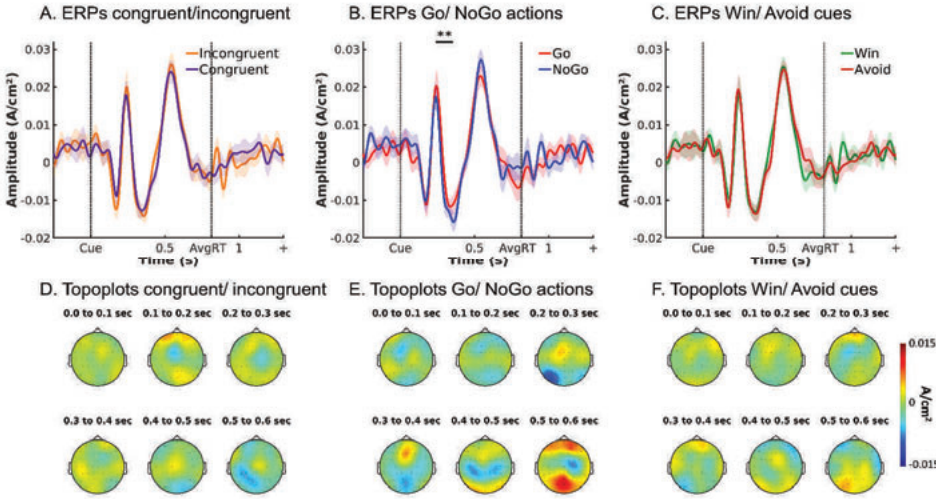


Figure 2.12. S2.7A. ERPs (±SEM) as a function of congruency, action, and cue valence over midfrontal electrodes (Fz/ FCz/ Cz; correct trials only).
A. There was no difference in midfrontal between congruent and incongruent trials, showing that the transient alpha effect observed on incongruent trials (main text Fig. 2.3E-F) was not reducible to evoked activity. **B.** The frontal P2 component was stronger for Go compared to NoGo actions (and N2 respectively weaker). ** $p < 0.01$. **C.** There was no difference between ERPs on Win and Avoid cues—apart from a small difference when the signal rises towards the P3 peak. **D-F.** Topoplots displaying differences in ERPs between (**D**) congruency, (**E**) action, and (**F**) valence conditions in steps of 100 ms from 0 to 600 ms.

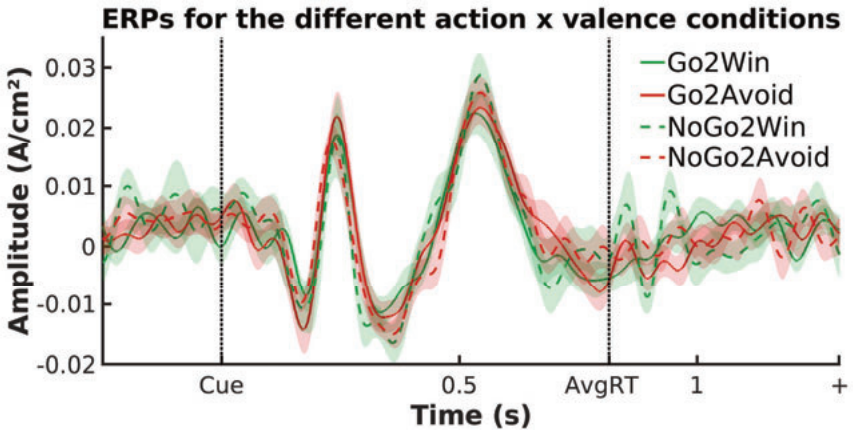


Figure 2.13. S2.7B. ERPs (±SEM) as a function of cue valence and action (correct trials only) over midfrontal electrodes (Fz/ FCz/ Cz).
There was no indication that trials with bias-incongruent (Go2Avoid and NoGo2Win) compared to bias-congruent (Go2Win and NoGo2Avoid) actions led to systematically higher/ lower component amplitudes.

2.6.8 S2.8: Conflict-related alpha power after ERPs are subtracted

To test whether the observed earlier phasic alpha increase for incongruent compared to congruent conditions was attributable to evoked rather than induced activity, we removed evoked components from our data (correct trials only) by computing the average ERP for each condition per participant and subtracting it from the trial-by-trial data before performing time-frequency decomposition (Cohen and Donner 2013). A permutation test on the alpha band yielded the same early phasic alpha increase for incongruent compared to congruent actions (p = .024; see Fig. S2.8) as reported in the main text (see Fig. 2.3E-F), suggesting that early alpha increase reflected induced rather than evoked activity.

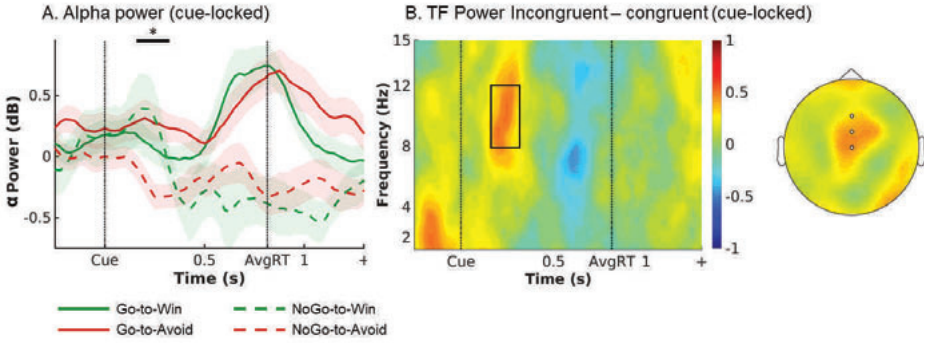


Figure 2.14. S2.8. EEG alpha power with stimulus-locked ERPs subtracted.
A. Trial time course of average (±SEM) alpha power (8–13 Hz) over midfrontal electrodes (Fz/FCz/Cz) per cue condition (correct-trials only; stimulus-locked). Alpha power transiently increases for both incongruent conditions in an early time window of around 175–325 ms. The time window where the tested data shows t-values > 2 is indicated by the box. * $p < 0.05$. **B.** Left: Time-frequency plot displaying that the transient power increase was focused on the alpha band, leaking into upper theta. Right: Topoplot of alpha power displaying that this incongruency effect was restricted to midfrontal electrodes (highlighted by white disks).

2.6.9 S2.9: Alpha signal as a function of cue valence, required action, and correctness

Given that we did not expect motivational conflict to be encoded in an early phasic signal in the alpha band, we conducted follow-up analyses. If this alpha signal reflected conflict detection that was causally involved in suppressing motivational biases, it should occur only when incongruent trials where met with the correct response, but be attenuated or even absent when those trials where met with an incorrect response, i.e., when participants failed to detect and/or overcome biases (Swart et al. 2018). Furthermore, the signal should occur only on incongruent trials, but not congruent trials, reflecting conflict detection mechanisms that are selectively recruited on incongruent trials rather than (possibly attentional) mechanisms improving accuracy more globally.

For this purpose, instead of global permutation tests across time and frequencies, we extracted average oscillatory power in a focal window of 175–325 ms after cue onset in the range of 8–13 Hz, averaged over midfrontal electrodes (Fz/ FCz/ Cz), for each participant, and performed repeated-measures ANOVAs with the independent variables valence (Win/ Avoid), required action (Go/ NoGo), and accuracy (Swart et al. 2018).

The RM-ANOVA yielded a significant main effect of valence, $F(1, 35) = 6.930$, $p = .013$, $\eta^2 = 0.007$, a significant two-way interaction between valence and action, $F(1, 35) = 8.368$, $p = .006$, $\eta^2 = 0.008$, but also a significant three-way interaction between valence, action, and accuracy, $F(1, 35) = 5.103$, $p = .03$, $\eta^2 = 0.005$ (see Fig. S2.9). The main effect of accuracy was not significant, $F(1, 35) = 2.02$, $p = .164$, $\eta^2 = 0.003$, suggesting that the observed alpha effect did not reflect an (attentional) process that was overall conducive to higher accuracy. For correct trials, we found the expected two-way interaction between valence and action, $F(1, 35) = 9.582$, $p = .004$, $\eta^2 = 0.023$, in absence of significant main effects, reflecting that alpha power was indeed higher on correct incongruent trials than correct congruent trials (simple effect: $t(35) = 3.096$, $p = .004$, $d = 0.397$). This reproduces the result of the permutation test from the main text. In contrast, for incorrect trials, we found only a significant main effect of valence, $F(1, 35) = 8.637$, $p = .006$, $\eta^2 = 0.011$, reflecting overall higher alpha for Win than Avoid cues (simple effect: $t(35) = 2.939$, $p = .006$, $d = 0.490$), but no significant interaction between valence and action, $F(1, 35) = 0.239$, $p = .628$, $\eta^2 < 0.001$. Incorrect incongruent trials did not lead to significantly higher alpha power than incorrect congruent trials, $t(35) = 0.489$, $p = .628$, $d = 0.081$.

These additional findings are in line with increased alpha power reflecting a conflict detection mechanism that is selectively recruited on incongruent trials on which biases are successfully overcome.

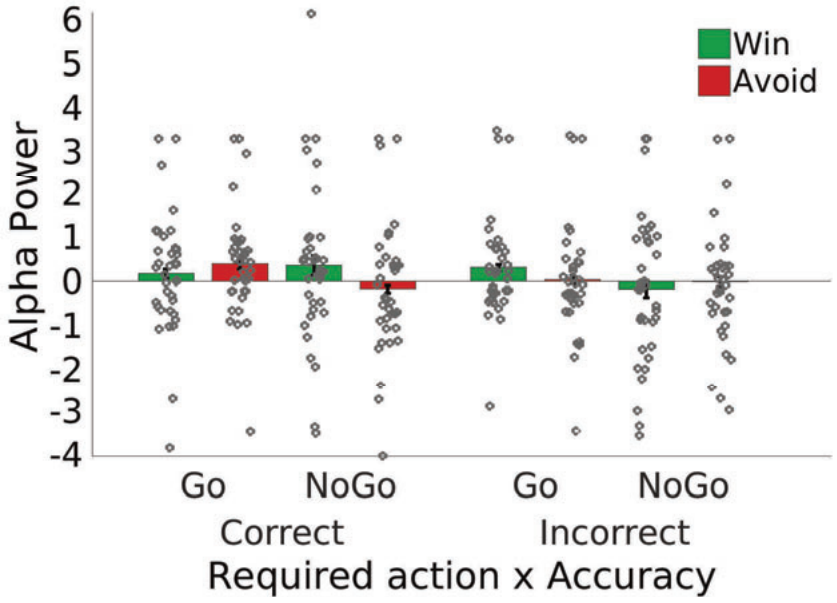


Figure 2.15. S2.9. Alpha power (±SEM) as a function of cue valence, performed action, and accuracy over midfrontal electrodes (Fz/ FCz/ Cz).
Alpha power selectively increased for correct bias-incongruent actions (correct Go2Avoid and NoGo2Win). Points are individual participant data points.

2.6.10 S2.10: EEG TF power as a function of action and valence across correct and incorrect trials

While EEG results reported in the main text only include correct trials in order to avoid contamination by error-related activity, fMRI results include all trials while explicitly modeling error trials with a designated regressor. To match this fMRI analysis approach, we report EEG analyses including both correct and incorrect trials, as well.

Results were highly similar to those of the correct trials only reported in the main text: Broadband power (1–15 Hz) was again significantly higher on trials with Go actions than NoGo actions (cue-locked: $p = .006$; response-locked: $p = .004$): This difference between Go and NoGo actions occurred as a broadband-signal from 1–15 Hz, but peaked in the theta band (Fig. S2.10 panels B and D). The topographies exhibited a bimodal distribution with peaks both at frontopolar (FPz) and central (FCz, Cz, CPz) electrodes (Fig. S2.10 panels B and D). As visual inspection of Fig. S2.10 panel C shows, theta power increased in all conditions until 500 ms post cue onset and then bifurcated depending on the action: For NoGo actions, power decreased, while for Go actions, power kept rising and peaked at the time of the response. This resulted in higher broadband power for Go versus NoGo actions for about 575–1300 ms after cue onset (see Fig. S2.10 panel C; around -150–475 ms when response-locked, see Fig. S2.10 panel A). When looking at the cue-locked signal, the signal peaked earlier and higher for Go actions to Win than to Avoid cues on correct trials, but not on incorrect; hence, when testing for differences in broadband power between Win cues and Avoid cues, broadband power was not significantly different between Win and Avoid cues. This difference in latency and peak height of the ramping signal was not present in the response-locked signal, and the respective test of Win vs. Avoid cues not significant either.

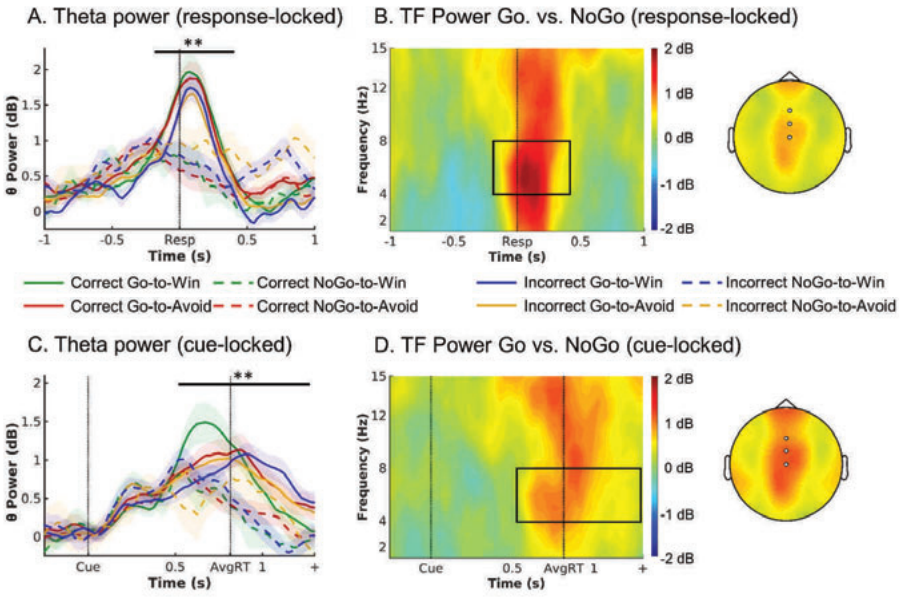


Figure 2.16. S2.10. EEG time-frequency power as a function of cue valence and action for both correct and incorrect trials.
A. Response-locked within trial time course of average theta power (4–8 Hz) over midfrontal electrodes (Fz/ FCz/ Cz) per cue condition (correct-trials only). Theta increased in all conditions relative to pre-cue levels, but to a higher level for Go than NoGo trials. There were no differences in theta peak height or latency between Go2Win and Go2Avoid trials. **B.** Left: Response-locked time-frequency power over midfrontal electrodes for Go minus NoGo trials. Go trials featured higher broadband TF power than NoGo trials. The broadband power increase for Go compared to NoGo trials is strongest in the theta range. Right: Topoplot for Go minus NoGo trials. The difference is strongest at FZ and FCz electrodes. **C-D.** Cue-locked within trial time course and time-frequency power. Theta increased in all conditions relative to pre-cue levels, but to a higher level for Go than NoGo trials, with earlier peaks for Go2Win than Go2Avoid trials. * $p < 0.05$. ** $p < 0.01$. Shaded error bars indicate (±SEM). Box in TF plots indicates the time frequency window where *t*-values > 2.

2.6.11 S2.11: Plots and tests of the evidence accumulation hypothesis

Previous research has suggested theta oscillations to reflect midfrontal mechanisms that lead to an elevation of the response threshold in basal-ganglia action selection mechanisms (Cavanagh et al. 2011; Cavanagh and Frank 2014; Cohen 2014; Frank et al. 2015). However, an alternative interpretation of past findings could be that theta reflects the subcortical action selection process itself (Bland and Oddie 2001; Caplan et al. 2003; DeCoteau et al. 2007a, 2007b; Womelsdorf et al. 2010). This reasoning would explain why in motor tasks, over the entire trial time course, theta oscillations strongly increase in any task condition, even in absence of conflict (Cohen and Cavanagh 2011; Swart et al. 2018). In conflict situations warranting elevated response thresholds, this process continues beyond normal levels and evolves for an extended time period, leading to the typical "conflict-related theta" reported in the literature (Murphy et al. 2015).

In fact, several characteristics of the theta signal we observed resembled an accumulating evidence process as evidenced by additional tests of systematic different in peak height and latency of the signal (O'Connell et al. 2012). In our case, features of the theta signal would be consistent with evidence selectively accumulated for making a Go action:

First, such a process should rise early (when Go is still a considered option) in all trials, but deactivate when the final response is NoGo, while it should keep rising when the final response is Go. This prediction is in line with our observations (see Fig. 2.3C in the main text).

Second, we found the theta signal to scale with reaction times, such that the signal peaked earlier on trials with earlier response times. This link would be expected when a signal causes the a response, such that the latency of the signal peak determines reaction times (O'Connell et al. 2012). To test this hypothesis in our data, we split up each participant's trials with Go actions (correct trials only) into three equally sized bins of fast, medium, and slow reaction times (tertials), separately for Win and Avoid trials (to account for inherent differences in reaction times between Win and Avoid trials). We then computed the average stimulus-locked signal in the theta range for each bin and determined the time point between 0.3 (fastest responses) and 1.3 s (slowest possible responses) at which the signal (first) reached its peak. We then used one-tailed paired-samples *t*-tests to test whether the signal peaked earlier in bins with faster reaction times, separately for Win and Avoid trials. Overall, the signal peaked earlier for Win trials (M = 0.697, SD = 0.175) than Avoid trials (M = 0.787, SD = 0.264), $t(35) = 2.335$, $p = 0.013$, $d = 0.39$. This difference was selective for the theta band (Fig. S2.11B panels A-C). For Win trials, indeed, the signal peaked significantly earlier for faster than for medium reaction times, $t(35) = 1.745$, $p = 0.045$, $d = 0.291$, and significantly earlier for medium than late reaction times, $t(35) = 2.577$, $p = 0.007$, $d = 0.430$ (see Fig. S2.11A panel A). For Avoid trials, the signal only peaked marginally significantly earlier for faster than for medium reaction times, $t(35) = 1.662$, $p = 0.053$, $d = 0.277$, but significantly earlier for medium than late reaction times, $t(35) = 5.220$, $p < 0.001$, $d = 0.870$. Conclusions were identical when using non-parametric permutation tests instead of *t*-tests. In sum, the peak latency of the stimulus-locked theta signal scaled with reaction times, as expected for a signal triggering actions.

Third, when response-locked, differences in peak latency and height between cue valence conditions and reaction time bins disappeared, in line the assumption of a fixed threshold that evidence must reach in order to trigger action release (O'Connell et al. 2012). To investigate

systematic differences in peak latency, we computed the average response-locked signal in the theta range for each bin for each participant and determined the time point between 0.5 s before and 0.5 s after the response at which the signal (first) reached its peak. We again compared bins within each valence condition using two-tailed *t*-tests. For Win trials, there were no significant differences in peak latency between fast and medium reaction times, $t(35) = -0.784$, $p = 0.438$, $d = -0.131$, medium and slow reaction times, $t(35) = 0.896$, $p = 0.376$, $d = 0.149$, or fast and slow reaction times, $t(35) = 0.135$, $p = 0.894$, $d = 0.023$ (see Fig. S2.11A panel B). Similarly for Avoid trials, there were no significant differences in peak latency between fast and medium reaction times, $t(35) = 0.014$, $p = 0.988$, d = 0.002, medium and slow reaction times, $t(35) = 0.347$, $p = 0.730$, $d = 0.058$, or fast and slow reaction times, $t(35) = 0.245$, $p = 0.808$, $d = 0.041$. Conclusions were identical when using non-parametric permutation tests instead of *t*-tests.

To investigate significant differences in peak height, we extracted the height of the theta signal at the peak latency within each bin for each participant. We again compared bins within each valence condition using two-tailed *t*-tests. For Win trials, were no significant differences in peak height between fast and medium reaction times, $t(35) = -1.138$, $p = 0.264$, $d = -0.190$, medium and slow reaction times, $t(35) = 1.003$, $p = 0.322$, $d = 0.167$, or fast and slow reaction times, $t(35) = 0.206$, $p = 0.838$, $d = 0.034$ (see Fig. S2.11A panel B). Similarly for Avoid trials, there were no significant differences in peak height between fast and medium reaction times, $t(35) = -0.017$, $p = 0.986$, $d = -0.003$, medium and slow reaction times, $t(35) = 0.195$, $p = 0.846$, $d = 0.033$, or fast and slow reaction times, $t(35) = 0.195$, $p = 0.846$, $d = 0.033$. Conclusions were identical when using non-parametric permutation tests instead of *t*-tests. Taken together, there were no significant differences in peak latency and height between different reaction times, in line with a fixed response threshold independent of response time or cue valence.

Fourth, previous research on EEG correlates of evidence accumulation in perceptual decision making has found that incorrect responses were elicited at systematically lower thresholds than correct responses (O'Connell et al. 2012), suggesting that trial-by-trial variation in the response threshold can cause erroneous action releases. To test this hypothesis in our data, we computed the response-locked theta signal separately for correct and incorrect Go actions on Win and Avoid trials, and then computed the peak height for each participant. We used one-tailed *t*-tests to test whether peak height was lower for incorrect than correct trials. This was indeed the case both on Win trials, $t(35) = 2.558$, $p = 0.008$, $d = 0.426$, and Avoid trials, $t(35) = 2.729$, $p = 0.005$, $d = 0.455$ (Fig. S2.11A panel C). As an alternative, we performed a cluster-based permutation test contrasting correct and incorrect responses. Both were indeed significantly different, p = 0.047, most dominantly from 125 ms before until 25 ms after around responses. In conclusion, we found evidence in line with the hypothesis that false-positive action releases occur at a systematically lower response threshold than true-positive action releases.

Fifth, our interpretation of theta as reflecting evidence accumulation is in line with previous research that has found perceptual and value-based evidence to be reflected in the theta band (Hunt et al. 2012; van Vugt et al. 2012). Also, a recent study observed both perceptual and value-based evidence encoded in the gamma band in topographies very similar to the one we found, with peaks in both frontopolar and centroparietal electrodes (Polanía et al. 2014). It is possible that feedforward activity encoded in the gamma band is nested in lower-frequency theta cycles reflecting top-down integration (Canolty et al. 2006; Maris et al. 2011; Landau et al. 2015). In

conclusion, our results are in line with theta power reflecting an evidence accumulation process for deciding whether to perform an active Go response.

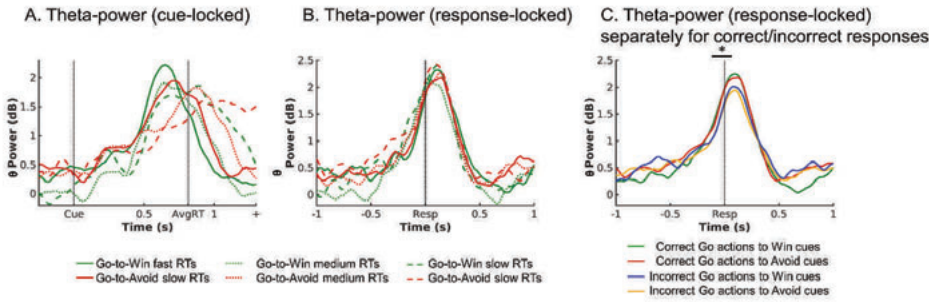


Figure 2.17. S2.11A. Plots displaying features of the theta signal akin to an evidence accumulation process for active Go responses. **A.** The stimulus-locked theta signal split into fast, medium, and slow reaction time bins separately for Win and Avoid trials. The signal peaks systematically earlier for earlier reaction times. **B.** The same signal response-locked. Differences in peak height and latency between reaction time bins are absent. **C.** Correct and incorrect Go actions separately for Win and Avoid trials. The theta peaks at significantly lower levels for incorrect compared to correct actions.

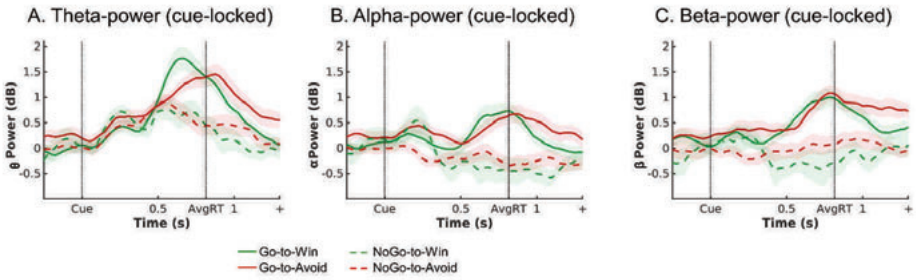


Figure 2.18. S2.11B. Plots displaying theta (A), alpha (B), and beta (C) power over midfrontal channels (Fz/ FCz/ Cz) split per cue-valence and performed action

(Correct trials only to avoid contamination error processing). Signal rise is strongest in the theta band. Also, the positively accelerating shape typically observed in evidence accumulation signals (Donner et al. 2009; O'Connell et al. 2012) is only observed in the theta band. Finally, differences in peak latency between Win and Avoid cues (in line with differences in reaction times between conditions) only arise in the theta band.

2.6.12 S2.12: Correlation of EEG power with head motion

Recently, (Fellner et al. 2016) reported that lower-frequency oscillations in simultaneous EEG-fMRI studies were affected by head motion: Realignment parameters strongly correlating with time-frequency power. These authors recorded simultaneous EEG and fMRI during encoding and retrieval phases of a memory task. They found difference in theta oscillations for remembered vs. forgotten items that had the opposite sign compared to when the same task was performed with EEG outside the MR scanner. Also, they found overall lower-frequency oscillations, especially in the theta range, to be strongly correlated with a summary measure of the six realignment parameters, casting doubt on the neural origin of theta effects measured in simultaneous EEG-fMRI recordings.

We leveraged our approach of fMRI-inspired EEG analysis approach (see main text) by computing the same summary statistic as (Fellner et al. 2016) based on the six realignment parameters for each volume, upsampling this signal to a TR of 0.140 s, and then downsampling the signal again into epochs of 2 s length relative to trial onset, yielding an indicator of overall head motion during each trial.

First, similar to (Fellner et al. 2016), we compared the head motion between trials with Go actions and trials with NoGo trials by computing the average head motion summary statistic for such trials for each participant and then performing a two-tailed paired-samples *t*-test. There was no significant difference in head motion between trials with Go and trials with NoGo actions, $t(35) = 1.614$, $p = 0.116$, $d = 0.269$. Similarly, there was no significant difference in head motion between Win and Avoid trials, $t(35) = -0.467$, $p = 0.643$, $d = -0.078$, nor between congruent and incongruent trials, $t(35) = -0.304$, $p = 0.763$, $d = -0.051$. These results suggest that head motion did not differ between experimental conditions.

Next, we used the head motion summary statistic as a trial-by-trial predictor of time-frequency power in multiple regression, similarly to the BOLD signal extracted from neural regions. When head motion was used as a sole regressor, we indeed observed a significant positive correlation with theta/ delta power ($p = 0.039$). However, this correlation was not spread out in time and frequency space as in (Fellner et al. 2016), but instead focused on theta/ delta power around 0.9–1.3 s. after cue onset—i.e., after the average response time (see Fig. S2.12 panel A). When entering BOLD signal from the selected regions as additional regressors, this pattern remained similar but became non-significant ($p = .089$).

Given that participants were instructed to perform Go actions only while the respective cue was visible (0–1.3 s after cue onset), we restricted our analyses of task-related neural signals to this period. However, head motion-related signals might occur even after this period. When performing fMRI-inspired EEG analyses on a window of 0–2 s after cue onset, we indeed observed a strong correlation ($p = 0.005$) of head motion with broadband time-frequency power after cue offset (after 1.3 s, see Fig. S2.12 panel B; especially so when including the five participants that were otherwise excluded from the fMRI-informed EEG analyses; see Fig. S2.12 panel C). This finding suggests that head-motion (and associated artifacts) might predominantly occur during the inter-stimulus interval when participants have performed any cue-related action and wait for the outcome. Note that the trials in (Fellner et al. 2016) were much longer (3 s) than in

our paradigm. Hence, head motion might be a particular problem in EEG-fMRI studies with paradigms featuring long trial durations, unlike the cue presentation phase in our paradigm.

Correlations of time-frequency power with BOLD and task factors remained significant even when head motion was included in the regression: First, we still observed a significant correlation of striatal BOLD with late theta/ delta power 0.5–1.0 s after cue onset (p = .045), suggesting that striatal BOLD predicts theta power independently of any head motion-related signals (see Fig. S2.12 panel D). Second, the correlations of left (p = .008) and right (p = .027) motor cortex (see S2.15) and vmPFC (p = .035) BOLD with time-frequency power remained significant. Third, inspecting the beta-map of the additional regressor Go vs. NoGo responses (coded as 1 and 0, regressor demeaned), which we included in all these analyses by default, revealed correlations with broadband signal around the time of responses (p = .026), a pattern very similar to the EEG-only analyses (see Fig. S2.12 panel E; compare to Fig. 2.3D in the main text). This finding suggests that the broadband signal associated with Go vs. NoGo responses in EEG-only analyses is not reducible to head motion artifacts either.

Finally, when using the trial-by-trial theta power as a regressor in an fMRI GLM (see S2.17), we observed theta correlates in action-related regions such as ACC, motor cortices, opercula, putamen, and cerebellum, unlike (Fellner et al. 2016) who observed correlates mostly in regions of the default-mode network. We thus conclude that the theta effects we observed constitute task-related neural signals rather than the head-motion related artifacts described by (Fellner et al. 2016).

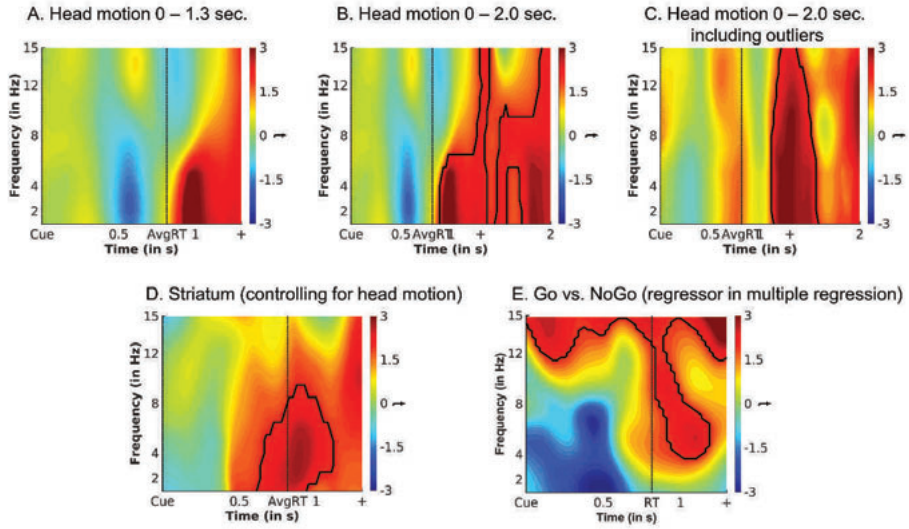


Figure 2.19. S2.12. Time-frequency correlates of head motion and task effects when controlled for head motion.
A. Head motion was quantified via relative displacement as a summary statistic of the volume-by-volume realignment parameters, which was first upsampled, then downsampled into a single value per trial, and used to predict average time-frequency power over midfrontal electrodes. Head motion correlated (though not significantly) with theta/ delta power around 1 sec. after stimulus onset. **B.** The same analyses on a time window of 0–2.0 revealed that head motion predominantly correlated with broadband time-frequency power after cue offset. **C.** Same plot as (B), but with the four participants included that are typically excluded due to out-of-range regression weights and strong head motion. **D.** The correlation of striatal BOLD with theta/ delta power around the time of responses remained unaltered when entering head motion as an additional regressor into the model. **E.** Using Go vs. NoGo responses (coded as 1 and 0, demeaned) as an additional regressor yielded again an increase in broadband time-frequency power for Go compared to NoGo responses (see main text Figure 3D), even when entering head motion as an additional regressor into the model. Areas surrounded by a black edge indicate clusters of $|t| > 2$ with $p < .05$ (cluster-corrected).

2.6.13 S2.13: Increase in midfrontal time-frequency power relative to baseline for each cue valence x performed action pairing

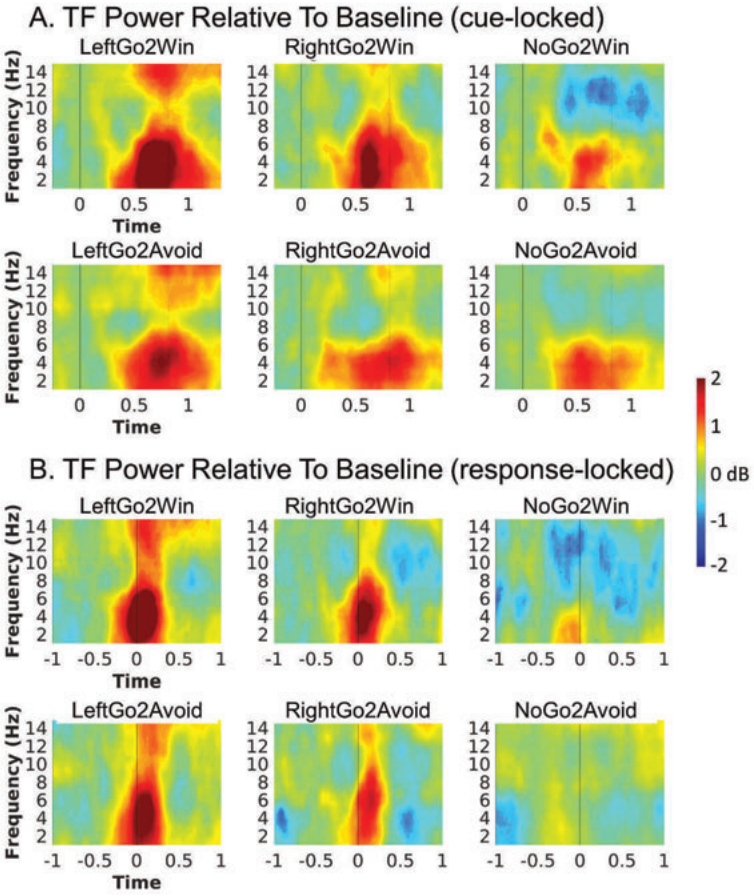


Figure 2.20. S2.13. Increase in midfrontal time-frequency power relative to baseline per condition.
Average time-frequency power over midfrontal electrodes (Fz/FCz/Cz) relative to baseline (mean signal -250–50 ms before cue onset), split up per cue valence (rows) and performed action (columns), both stimulus-locked (**A**) and response-locked (**B**). For stimulus-locked plots, vertical solid lines indicate cue onset, dashed vertical lines average response time. For response-locked plots, vertical solid lines indicate response time (i.e., 0 s). Time-frequency power in the delta/ theta range increases in every valence-action pairing, also for NoGo actions, which rules out that this signal is a mere motion artifact.

2.6.14 S2.14: Theta and beta power for left vs. right hand responses

The strong association of broadband power with motor activity opened the possibility that this signal was potentially an artifact of EEG acquisition (e.g. head movement in the scanner) rather than a neural signal (Fellner et al. 2016). If the signal constituted an artifact, one would expect it to be symmetrical for both hands. One the other hand, if it was a neural signal reflecting the level of evidence accumulated before initiating a response, one might expect the signal to be sensitive to differences in evidence thresholds between hands. Given that all our participants were right-handed, one might expect that responses of the left (non-dominant) hand were less easily initiated and required a higher level of evidence to be selected than responses of the right (dominant) hand. A broadband permutation test (stimulus-locked: $p = .020$; response-locked: $p = .006$) indicated that power in the theta band (around -250–25 ms relative to responses, see Fig. S2.14) and in the beta-band (around -150–675 ms relative to responses (see Fig. S2.14 panels B and D) was in fact higher for left-hand than right-hand responses.

This modulation of the theta signal by handedness corroborates the interpretation that the observed theta synchronization is of neural origin. It might reflect a bias towards right-hand responses, such that left-hand responses require a higher level of evidence to be initiated than right-hand responses. Such a right-hand bias might also explain why synchronization in the beta band was higher for left than right hand responses: Given that beta synchronization is typically found for motor inhibition (Wessel, Ghahremani, et al. 2016; Wessel et al. 2019), higher beta on trials with left-hand responses might reflect that the right hand needed to be actively suppressed on these trials (see also S2.15). In fact, we also observed that left-hand responses ($M = 0.772$) were overall slower than right hand responses ($M = 0.745$), $\chi^2(1) = 6.709$, $p = .010$, which further corroborates the interpretation that left-hand responses might have been harder to perform than right-hand responses.

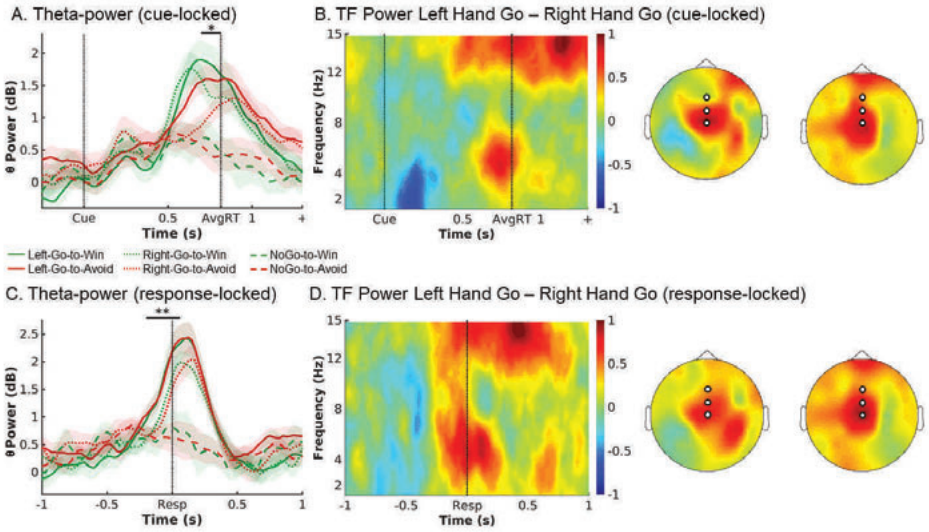


Figure 2.21. S2.14. Theta and beta power split up for left vs. right hand responses.
A. Trial time course of average (±SEM) theta power (4–8 Hz) over midfrontal electrodes (Fz/FCz/Cz) per split up for left- and right-hand responses (correct-trials only; stimulus-locked). Theta increased to a higher extend for left-hand than right-hand responses. * $p < 0.05$. **B.** Left: Time-frequency power over midfrontal electrodes for left-hand minus right-hand responses trials. Left-hand responses were associated with high theta and beta power compared to right-hand responses. Right: Topoplot for left-hand minus right-hand responses in the theta (left) and beta (right) range. **C-D.** Same data when response-locked. Differences in theta peak latency between Go2Win and Go2Avoid trials disappear. ** $p < 0.01$.

2.6.15 S2.15: Supplementary fMRI-inspired EEG results in time-frequency space

In addition to EEG correlates of BOLD signal in the striatum, ACC and vmPFC (see main text), we also observed correlates for BOLD in left and right motor cortex. Both left motor cortex (two separate clusters, both $p = .030$ and $p = .030$, cluster-corrected) and right motor cortex ($p = .001$ cluster-corrected) exhibited overlapping, but oppositely signed correlates in the alpha/beta band, with left motor cortex correlating negatively with midfrontal beta power (around 12–15 Hz, 0.6–1.3 s, Fig. S2.15 panel B), while right motor cortex correlated positively with alpha/beta power (around 10–15 Hz, 0.6–1.3 s, Fig. S2.15 panel C). These findings mirror the observation of higher theta and beta power for left hand (i.e., right motor cortex) compared to right hand (i.e., left motor cortex) responses (see Supplementary Material S2.9), again suggesting that executing a left-hand response might have required an active suppression (associated with increased beta power) of the right hand. These results corroborate that theta power does not reflect motor preparation/ execution signals from the motor cortices, but signals from distinct regions. Furthermore, these associations replicate numerous intracranial and source-localization studies (Sanes and Donoghue 1993; Salmelin, Forss, et al. 1995; Salmelin, Hämäläinen, et al. 1995; Stolk et al. 2019) and previous EEG-fMRI studies (Jurkiewicz et al. 2006; Ritter et al. 2009) reporting beta oscillations in motor cortices. The presence of this well-established BOLD-EEG association corroborates the robustness of the data and analysis.

We performed a range of follow-up analyses to check for the robust of our results. We reached similar results and identical conclusions when a) performing regressions with each region as the sole predictor, b) including a summary measure of the realignment parameters as a proxy for head motion into the regression (Fellner et al. 2016), and c) when fitting HRFs for all trials of a certain block within a single GLM instead of separately for each trial.

We lastly aimed to test whether distinct striatal subregions with opposite valence coding, i.e., left putamen (Win > Avoid) and bilateral medial caudate (Avoid > Win), showed distinct time-frequency correlates. When using BOLD from those subregions instead of overall striatal BOLD as regressors, left putamen BOLD did not exhibit a significant association with time-frequency power ($p = .218$), while medial caudate BOLD did significantly correlate with delta/theta power around 825–1,2500 ms post-stimulus ($p = .011$). The cluster of significant correlations observed for medial caudate was highly similar to the cluster observed as a correlate of the entire striatum. Descriptively, both ROIs showed clusters of positive correlations with theta/ delta power around the time of responses, and slightly earlier so for the left putamen than for medial caudate. This finding would be in line with the idea of left putamen more strongly driving Go responses on Win trials, which showed shorter RTs, and medial caudate rather driving Go responses on Avoid trials, which showed longer RTs. However, as clusters associated with those regions were small and permutation tests not significant, this descriptive finding should be interpreted with caution.

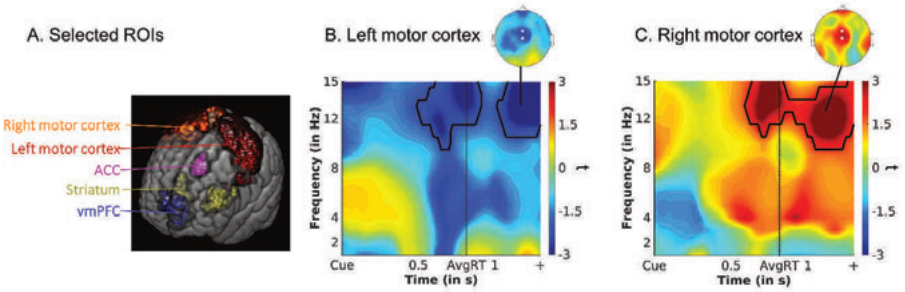


Figure 2.22. S2.15. Supplementary fMRI-inspired EEG results in time-frequency space.
A. Selected ROIs from which trial-by-trial HRF amplitudes were used as predictors in a multiple regression of midfrontal EEG time-frequency power. **B-C.** Unique temporal correlation of BOLD signal in (**B**) left and (**C**) right motor cortex to average EEG time-frequency power over midfrontal electrodes (FCz/Cz). Group-level *t*-maps display the modulation of the EEG time-frequency power by trial-by-trial BOLD signal in the selected ROIs. Midfrontal beta power correlates negatively with BOLD in left motor cortex (more active for right hand responses), but positively with BOLD in right motor cortex (more active for left hand responses), putatively indexing response conflict and inhibitory processes when left hand responses were executed. Areas surrounded by a black edge indicate clusters of $|t| > 2$ with $p < .05$ (cluster-corrected). Topoplots indicate the topography of the respective cluster.

2.6.16 S2.16: Supplementary fMRI-inspired EEG results in time space (ERPs)

Given that the time-frequency correlate of trial-by-trial vmPFC BOLD occurred very early after cue onset and was extended in frequency space, we hypothesized that vmPFC BOLD might be correlated with evoked rather than induced activity, which, when analyzed in time-frequency space, smeared across frequencies. We used the same approach for fMRI-informed EEG analyses as reported in the main text, but with the voltage signal (time-domain) instead of time-frequency power as dependent variable. We again used BOLD signal from striatum, ACC, left and right motor cortex, and vmPFC as simultaneous predictors in one single multiple regression.

When restricting analyses to midfrontal electrodes (FCz/ Cz), we found no significant modulation of EEG voltage by vmPFC BOLD (p = .260; see Fig. S2.16 panels A and C). However, when considering a broader frontal ROI (F1/F3/FCz/FC1/FC3/ Cz/C1/C3), vmPFC appeared to attenuate the amplitude of the P2 component over left frontal electrodes (two clusters above threshold: p = .021 around 213–269 ms; p = .003 around 349 – 410 ms; see Fig. S2.16 panels B and D). The topography of EEG voltage modulation did not exactly match with the topography of the time-frequency power modulation, but was rather restricted to left frontal electrodes (see Fig. S2.16 panels C and E). Interestingly, these electrodes also showed the peak modulation of the P2 by Go compared to NoGo actions (see S2.7). In conclusion, we found inconclusive evidence regarding whether broadband power decreases associated with vmPFC BOLD were reducible to evoked activity (modulation of the P2) or not.

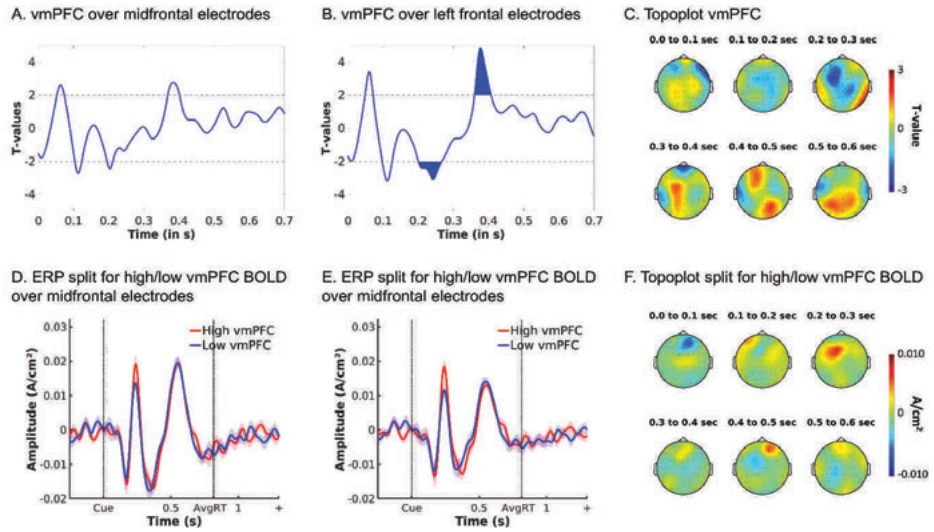


Figure 2.23. S2.16. Modulation of EEG voltage by vmPFC BOLD signal.
A. Average EEG voltage over midfrontal electrodes (FCz/ Cz) was not significantly modulated by vmPFC BOLD, while (**B**) EEG voltage over left frontal electrodes (F1/ F3/ FCz/ FC1/ FC3/ FC5/ Cz/ C1/ C3) was. Filled areas indicate clusters of $|t| > 2$ with $p < .05$ (cluster-corrected). **C.** Topoplots displaying *t*-values over the entire scalp in steps of 100 ms from 0 to 800 ms. The strongest modulation of frontal EEG voltage by vmPFC BOLD occurred over left frontal electrodes. This pattern does not fully match the topography of broadband power decreased associated with vmPFC BOLD (see Fig. 2.4C in the main text). **D.** For plotting purposes, we sorted trials according to the trial-by-trial HRF amplitude in the vmPFC ROI and plotted the 33% of trials with highest vmPFC BOLD signal vs. the 33% trials with lowest vmPFC BOLD signal. This contrast indicates no strong difference a midfrontal electrodes (FCz/ Cz), but (**E**) does indicate an attenuation of the P2 component through high vmPFC BOLD over left frontal electrodes (F1/ F3/ FCz/ FC1/ FC3/ FC5/ Cz/ C1/ C3). **F.** Topoplots displaying voltage for high vmPFC BOLD minus low vmPFC BOLD trials over the entire scalp in steps of 100 ms from 0 to 800 ms. The strongest modulation of frontal EEG voltage by vmPFC BOLD occurred over left frontal electrodes.

2.6.17 S2.17: EEG-informed fMRI analyses

For the EEG-inspired fMRI analyses, we added trial-by-trial summary measures of conflict-related alpha power and action-related theta power to our GLM. These measures were created by using the 3-D (time-frequency-channel) *t*-map obtained when contrasting incongruent vs. congruent actions (Mask 1; stimulus-locked) and Go vs. NoGo actions (Mask 2; response-locked) over midfrontal channels (Fz/ FCz/ Cz) as a linear filter. We extracted those maps and retained all voxels with $t > 2$. We did not enforce strict frequency band cutoffs, but rather extracted the entire cluster of *t*-values above threshold. Restricting the action contrast *t*-map to the theta range or using the stimulus-locked rather than the response-locked map led to highly similar results and identical conclusions. These masks were applied to the trial-by-trial time-frequency data to create weighted summary measures of the average power in the identified clusters in each trial. Both resultant time series correlated only weakly (mean correlation across participants: $r = .105$). They were entered as parametric modulators on top of the task regressors as described above, with each regressor entering a separate parametric contrast.

The EEG alpha regressor correlated significantly negatively with BOLD in two clusters in left middle frontal gyrus/ lateral frontal pole ($z_{max} = 3.99$, p = 0.000506, xyz = [-38 34 32]) and in right supramarginal gyrus ($z_{max} = 4.45$, p = 0.00619, xyz = [54 -42 38]). As Fig. S2.17 panel A indicates, sub-threshold, the same areas in the respective other hemisphere also correlated negatively with trial-by-trial alpha, as did extensive areas in medial parietal/ occipital cortex. Overall, midfrontal alpha appeared to correlate negatively with extended areas that were part of the fronto-parietal and dorsal attention resting-state networks. Notably, no region correlated positively with midfrontal alpha.

The EEG theta regressor correlated significantly positively with BOLD in pre-SMA/ ACC/ left precentral and postcentral gyrus ($z_{max} = 4.49$, p = 1.08e-15, xyz = [14 -10 64]), right precentral and postcentral gyrus ($z_{max} = 4.34$, p = 1.14e-08, xyz = [22 -34 74]), left ($z_{max} = 4.56$, p = 0.000213, xyz = [-54 -24 22]) and right ($z_{max} = 4.28$, p = 7.75e-07, xyz = [62 -24 30]) supramarginal gyrus/ operculum, left ($z_{max} = 4.82$, p = 0.000279, xyz = [-48 4 6]) and right ($z_{max} = 4.07$, p = 5.3e-05, xyz = [26 2 10]) striatum and operculum, and bilateral cerebellum ($z_{max} = 4.67$, p = 3.58e-07, xyz = [10 -54 -14]) (see Fig. S2.17 panel B). These regions also tended to be more active for Go than NoGo responses, corroborating the notion of theta reflecting evidence for active responses.

Notably, correlations with trial-by-trial theta power were not restricted to the striatum, but also occurred for other motor regions such as the ACC and motor cortices. These differences to the results of the fMRI-inspired EEG analyses might be attributable to methodological differences between both approaches: First, if theta power reflects global trial-by-trial brain activity associated with motion, this signal property will lead to correlations with BOLD in several motor regions in EEG-inspired fMRI analyses. In contrast, in fMRI-inspired analyses, such variance will be shared among regressors and thus be attributed to neither of them. Second, for EEG-inspired fMRI analyses, we created trial-by-trial indicators of theta power within a broad time window, which potentially mixes distinct events in theta that reflect activity in different brain regions. Thus, this approach might lead to correlations with BOLD in several brain regions that actually perform different computations at different time points. Following this reasoning, fMRI-inspired analyses have the advantage of a) identifying which regions uniquely predict time-frequency power beyond

variance shared among regions, and b) unmixing different regions predicting time-frequency power at different time points.

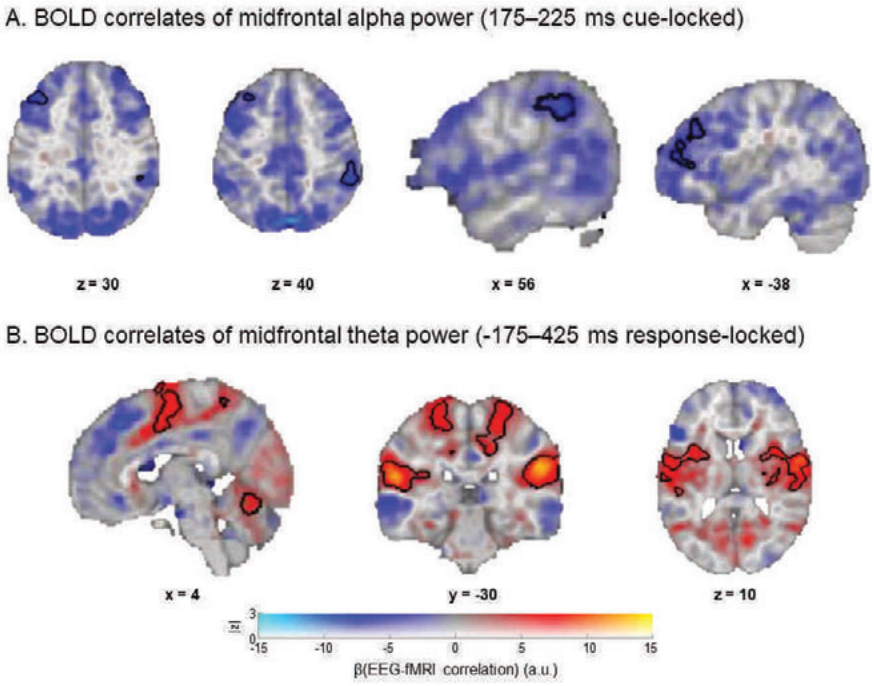
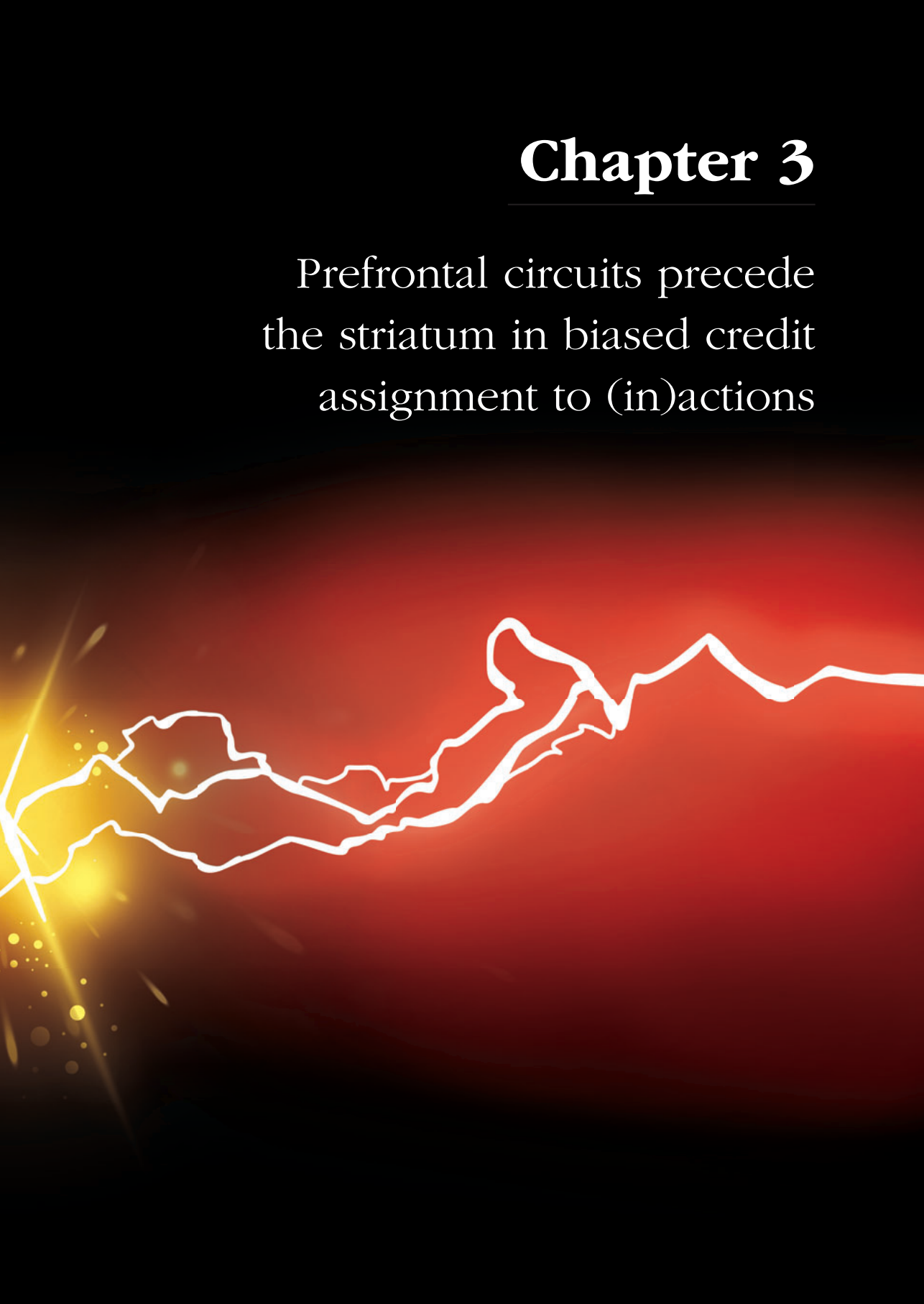


Figure 2.24. S2.17. Trial-by-trial time-frequency power as a predictor of BOLD in an fMRI GLM.
A. Trial-by-trial midfrontal alpha power correlated significantly negatively with BOLD in two clusters in left middle frontal gyrus and in right supramarginal gyrus. Sub-threshold, it correlated negatively with extended regions in fronto-parietal and dorsal attention network. **B.** Trial-by-trial midfrontal theta power correlated significantly positively with BOLD in pre-SMA, ACC, bilateral precentral and postcentral gyrus, superior parietal lobule, precuneous, bilateral operculum, bilateral putamen, and bilateral cerebellum.

Chapter 3

Prefrontal circuits precede the striatum in biased credit assignment to (in)actions



3 PREFRONTAL CIRCUITS PRECEDE THE STRIATUM IN BIASED CREDIT ASSIGNMENT TO (IN)ACTIONS

3.1 ABSTRACT

Actions are biased by the outcomes they can produce: Humans are more likely to show action under reward prospect, but hold back under punishment prospect. Such motivational biases derive not only from biased response selection, but also from biased learning: humans tend to attribute rewards to their own actions, but are reluctant to attribute punishments to having held back. The neural origin of these biases is unclear; in particular, it remains open whether motivational biases arise primarily from the architecture of subcortical regions or also reflect cortical influences, the latter being typically associated with increased behavioral flexibility. Simultaneous EEG-fMRI allowed us to track which regions encoded biased prediction errors in which order. Biased prediction errors occurred in cortical regions (dACC, pgACC, PCC) before subcortical regions (striatum). These results highlight that biased learning is not a mere feature of the basal ganglia, but arises through prefrontal cortical contributions, revealing motivational biases to be a potentially flexible, sophisticated mechanism.

3.2 Introduction

Human action selection is biased by potential action outcomes: reward prospect drives us to invigorate action, while threat of punishment holds us back (Dayan et al. 2006; Guitart-Masip, Duzel, et al. 2014; Swart et al. 2017). These motivational biases have been evoked to explain why humans are tempted by reward-related cues signaling the chance to gain food, drugs, or money, as they elicit automatic approach behavior. Conversely, punishment-related cues suppress action and lead to paralysis, which may even lie at the core of mental health problems such as phobias and mood disorders (Huys et al. 2016; Mkrtchian, Aylward, et al. 2017). While such examples highlight the potential maladaptiveness of biases in some situations, they confer benefits in other situations: Biases could provide sensible "default" actions before context-specific knowledge is acquired (Dayan et al. 2006; Huys et al. 2011). They may also provide ready-made alternatives to more demanding action selection mechanisms, especially when speed has to be prioritized (Boureau et al. 2015).

Previous research has assumed that motivational biases arise because the valence of prospective outcomes influences action selection (Guitart-Masip, Huys, et al. 2012). However, we have recently shown that not only action selection, but also the updating of action values based on obtained outcomes is subject to valence-dependent biases (Swart et al. 2017, 2018; de Boer et al. 2019): humans are more inclined to ascribe rewards to active responses, but have problems with attributing punishments to having held back. On the one hand, such biased learning might be adaptive in combining the flexibility of instrumental learning with somewhat rigid "priors" about typical action-outcome relationships. Exploiting lifetime (or evolutionary) experience might lead to learning that is faster and more robust to environmental "noise". On the other hand, biases might be responsible for phenomena of "animal superstition" like negative auto-maintenance, with rats and pigeons showing vigorous behavior under reward availability even when their behavior prevents or delays reward delivery—failing to ever attribute reward delivery to having held back(Brown and Jenkins 1968; Williams and Williams 1969; Dayan et al. 2006). While reward attainment can lead to an illusory sense of control over outcomes, control is underestimated under threat of punishment: Humans find it hard to comprehend how inactions can cause negative outcomes, which makes them more lenient in judging harms caused by others' inactions (Ritov and Baron 1990; Zeelenberg et al. 2000). Taken together, also credit assignment is subject to motivational biases, with enhanced credit for rewards given to actions, but diminished credit for punishments given to inactions.

While evident in behavior, the neural mechanisms subserving such biased credit assignment remain elusive. Previous fMRI studies have addressed correlates of motivational response biases (Guitart-Masip, Fuentemilla, et al. 2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012) and we have recently found evidence for valence signals from vmPFC and ACC biasing action selection processes in the striatum (Algermissen et al. 2022). The same regions might be involved in motivational learning biases, given the prominent role of the basal ganglia system not only in action selection, but also learning. Influential computational models of basal ganglia function (Frank 2005; Collins and Frank 2014) (henceforth called "asymmetric pathways model") predict such motivational learning biases: Positive prediction errors, elicited by rewards, lead to long-term potentiation in the striatal direct "Go" pathway (and long term depression in the indirect pathway), allowing for a particularly effective acquisition of Go responses after rewards.

Conversely, negative prediction errors, elicited by punishments, lead to long term potentiation in the "NoGo" pathway, impairing the unlearning of NoGo responses after punishments. This account suggests that motivational biases arise within the same pathways involved in standard reinforcement learning (RL). An alternative candidate model is that biases arise through the modulation of these RL systems by external areas that also track past actions, putatively the prefrontal cortex (PFC). Past research has suggested that standard RL can be biased by information stored in PFC, such as explicit instructions (Doll et al. 2009; Atlas et al. 2016) or cognitive map-like models of the environment (Daw et al. 2011; Lee et al. 2014; Piray et al. 2016). Most notably, the anterior cingulate cortex (ACC) has been found to reflect the impact of explicit instructions (Atlas et al. 2016) and of environmental changes (Behrens et al. 2007; Meder et al. 2017) on prediction errors.

Both candidate models predict that BOLD signal in striatum should be better described by biased compared with "standard" prediction errors. In addition, the model proposing a prefrontal influence on striatal processing makes a notable prediction about the timing of signals: information about the selected action and the obtained outcome should be present first in prefrontal circuits to then later affect processes in the striatum. While fMRI BOLD recordings allow for unequivocal access to striatal activity, the sluggish nature of the BOLD signal prevents clear inferences about temporal precedence of signals from different regions. We thus combined BOLD with simultaneous EEG recordings which allowed us to precisely characterize learning signals in both space and time.

The key question is whether biased credit assignment arises directly from biased RL through the asymmetric pathways in the striatum, or whether striatal RL mechanisms are biased by external prefrontal sources, with the dACC as likely candidate. To this end, participants performed a motivational Go/ NoGo learning task that is well-established to evoke motivational biases (Swart et al. 2017, 2018; van Nuland et al. 2020). We expected to observe biased PEs in striatum and frontal cortical areas. By simultaneously recording fMRI and EEG and correlating trial-by-trial BOLD signal with EEG time-frequency power, we were able to time-lock the peaks of EEG-BOLD correlations for regions reflecting biased PEs and infer their relative temporal precedence. We focused on two well-established electrophysiological signatures of RL, namely theta and delta power (Cavanagh et al. 2010; Cohen, Wilmes, et al. 2011; van de Vijver et al. 2011; Talmi et al. 2013; Bernat et al. 2015; Cavanagh 2015) as well as beta power (van de Vijver et al. 2011; Marco-Pallarés et al. 2015) over midfrontal electrodes.

3.3 Results

Thirty-six participants performed a motivational Go/ NoGo learning task (Swart et al. 2017, 2018) in which required action (Go/ NoGo) and potential outcome (/ punishment) were orthogonalized (Fig. 3.1A-D). They learned by trial-and-error for each of eight cues whether to perform a left button press (Go_{LEFT}), right button press (Go_{RIGHT}), or no button press (NoGo), and whether a correct action increased the chance to win a reward (Win cues) or to avoid a punishment (Avoid cues). Correct actions led to 80% positive outcomes (reward, no punishment), with only 20% positive outcomes for incorrect actions. Participants performed two sessions of 320 trials with separate cue sets, which were counterbalanced across participants.

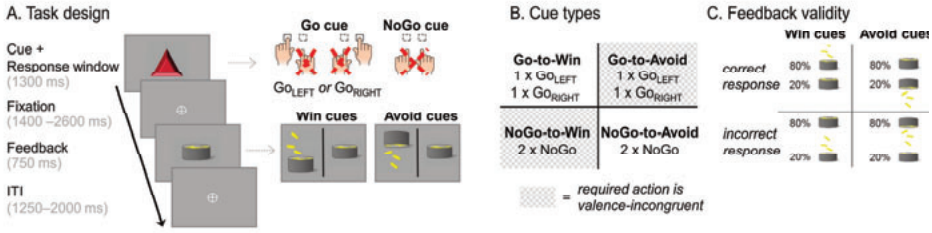


Figure 3.1. Motivational Go/ NoGo learning task design.
A. On each trial, a Win or Avoid cue appeared; valence of the cue was not signaled but should be learned. Cue offset was also the response deadline. Response-dependent feedback followed after a jittered interval. Each cue had only one correct action (Go_{LEFT}, Go_{Right}, or NoGo), which was followed by the positive outcome 80% of the time. For Win cues, actions could lead to rewards or neutral outcomes; for Avoid cues, actions could lead to neutral outcomes or punishments. Rewards and punishments were represented by money falling into/ out of a can. **B.** There were eight different cues, orthogonalizing cue valence (Win versus Avoid) and required action (Go versus NoGo). The motivationally incongruent cues (for which the motivational action tendencies were incongruent with the instrumental requirements) are highlighted in gray. **C.** Feedback was probabilistic: Correct actions to Win cues led to rewards in 80% of cases, but neutral outcomes in 20% of cases. For Avoid cues, correct actions led to neutral outcomes in 80% of cases, but punishments in 20% of cases. For incorrect actions, these probabilities were reversed.

3.3.1 Regression analyses of behavior

We performed regression analyses to test whether a) responses were biased by the valence of prospective outcomes (Win/ Avoid), reflecting biased responding and/ or learning, and b) whether response repetition after positive vs. negative outcomes was biased by whether a Go vs. NoGo response was performed, selectively reflecting biased learning.

For the first purpose, we analyzed choice data (Go/ NoGo) using mixed-effects logistic regression that included the factors required action (Go/ NoGo; note that this approach collapses across Go_{LEFT} and Go_{RIGHT} responses), cue valence (Win/ Avoid), and their interaction (also reported in)(Algermissen et al. 2022). Participants learned the task, i.e., they performed more Go responses towards Go than NoGo cues (main effect of required action: $b = 0.815$, $SE = 0.113$, $\chi^2(1) = 32.008$, $p < .001$). In contrast to previous studies (Swart et al. 2017, 2018), learning did not asymptote (Fig. 3.2A), which provided greater dynamic range for the biased learning effects to surface. Furthermore, participants showed a motivational bias, i.e., they performed more Go responses to Win than Avoid cues (main effect of cue valence, $b = 0.423$, $SE = 0.073$, $\chi^2(1) = 23.695$, $p < .001$). Replicating other studies with this task, there was no significant interaction between required action and cue valence ($b = 0.030$, $SE = 0.068$, $\chi^2(1) = 0.196$, $p = .658$, Fig. 3.2A-B), i.e., there was no evidence for the effect of cue valence (motivational bias) differing in size between Go or NoGo cues.

Secondly, as a proxy of (biased) learning, we analyzed cue-based response repetition (i.e., the probability of repeating a response on the next encounter of the same cue) as a function of outcome valence (positive vs negative outcome), performed action (Go vs. NoGo), and outcome salience (salient: reward or punishment vs. neutral: no reward or no punishment). As expected, participants were more likely to repeat the same response following a positive outcome (main effect of outcome valence: $b = 0.504$, $SE = 0.053$, $\chi^2(1) = 45.595$, $p < .001$). Most importantly, after salient outcomes, participants adjusted their responses to a larger degree following Go

responses than NoGo responses, revealing the presence of a learning bias (Fig. 3.2C; interaction of valence x action x salience: $b = 0.248$, $SE = 0.048$, $\chi^2(1) = 19.732$, $p < .001$). When selectively analyzing trials with salient outcomes only, rewards (compared to punishments) led to a higher proportion of choice repetitions following Go relative to NoGo responses (valence x response: $b = 0.308$, $SE = 0.064$, $\chi^2(1) = 17.798$, $p < .001$; valence effect for Go only: $b = 1.276$, $SE = 0.115$, $\chi^2(1) = 53.932$, $p < .001$; valence effect for NoGo only: $b = 0.637$, $SE = 0.127$, $\chi^2(1) = 18.228$, $p < .001$; see full results in S3.2).

Taken together, these results suggested that behavioral adaptation following rewards and punishments was biased by the type of action that led to this outcome (Go or NoGo). However, this analysis only considered behavioral adaptation on the next trial, and could not pinpoint the precise algorithmic nature of this learning bias. More importantly, it did not provide trial-by-trial estimates of action values as required for model-based fMRI and EEG analyses to test for regions or time points that reflected biased learning. We thus analyzed the impact of past outcomes on participants' choices using computational RL models.

3.3.2 Computational modeling of behavior

In line with previous work (Swart et al. 2017, 2018), we fitted a series of increasingly complex RL models. We started with a simple Rescorla Wagner model featuring learning rate and feedback sensitivity parameters (M1). We next added a Go bias, capturing participants' overall propensity to make Go responses (M2), and a Pavlovian response bias (M3), reflecting participants' propensity to adjust their likelihood of emitting a Go response in response to Win vs. Avoid cues (Swart et al. 2017). Alternatively, we added a learning bias (M4), amplifying the learning rate after rewarded Go responses and dampening it after punished NoGo responses (Swart et al. 2017), in line with the asymmetric pathways model. In the final model (M5), we added both the response bias and the learning bias. For the full model space (M1-M5) and model definitions, see the Methods section.

Model comparison showed clear evidence in favor of the full asymmetric pathways model featuring both response and learning biases (M5; model frequency: 86.43%, protected exceedance probability: 100%, see Fig. 3.2D, H; for model parameters and fit indices, see S3.3; for parameter recovery analyses, see S3.4). Posterior predictive checks involving one-step-ahead predictions and model simulations showed that this model captured key behavioral features (Fig. 3.2E, F), including motivational biases and a greater behavioral adaptation after Go responses followed by salient outcomes than after NoGo responses followed by salient outcomes (Fig. 3.2G). This pattern could not be captured by an alternative learning bias model based on the idea that active responses generally enhance credit assignment (Cockburn et al. 2014) (see S3.5).

One feature of the behavioral data that was not well captured by the asymmetric pathways model was a high tendency of participants to repeat responses ("stay") to the same cue irrespective of outcomes (see Fig. 3.2C and G). This tendency was stronger for Win than Avoid cues. We explored three additional models featuring supplementary mechanisms to account for this behavioral pattern (see S3.6). All these models fitted the data well and captured the propensity of staying better than M5; however, these models overestimated the proportion of incorrect Go responses. Model-based fMRI analyses based on these models led to results largely identical to those obtained with M5 (see S3.6). We thus focused on M5, which relied on only a single

mechanism (i.e., biased learning from rewarded Go and punishment NoGo actions), while additional models that tried to better capture the pattern of response repetition needed to use several mechanisms, leaving ambiguity about which mechanism drove neural correlates of biased prediction errors. See S3.6 for further details.

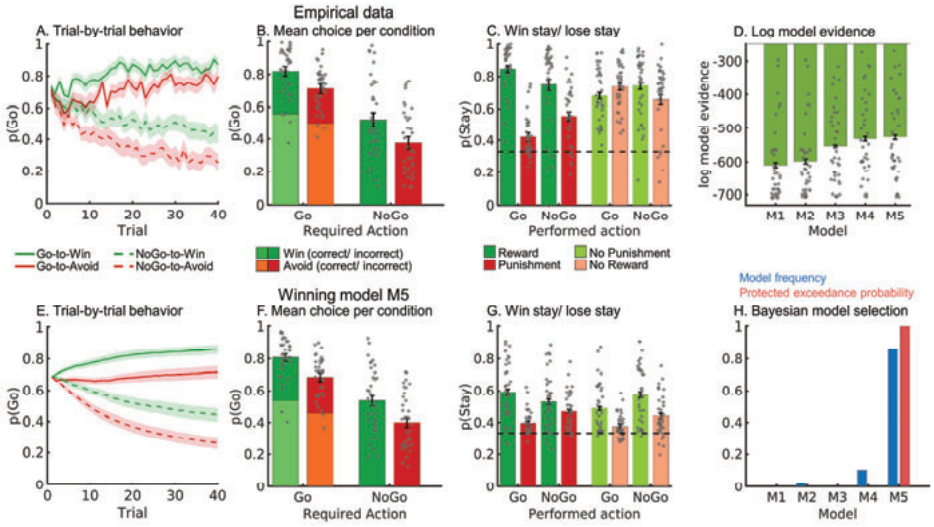


Figure 3.2. Behavioral performance.
A. Trial-by-trial proportion of Go responses (±SEM across participants) for Go cues (solid lines) and NoGo cues (dashed lines). The motivational bias was already present from very early trials onwards, as participants made more Go responses to Win than Avoid cues (i.e., green lines are above red lines). Additionally, participants clearly learn whether to make a Go response or not (proportion of Go responses increases for Go cues and decreases for NoGo cues). **B**. Mean (±SEM across participants) proportion Go responses per cue condition (points are individual participants' means). **C.** Probability to repeat a response ("stay") on the next encounter of the same cue as a function of action and outcome. Learning was reflected in higher probability of staying after positive outcomes than after negative outcomes (main effect of outcome valence). Biased learning was evident in learning from salient outcomes, where this valence effect was stronger after Go responses than NoGo responses. Dashed line indicates chance level choice (p_{Stay} = 0.33). **D.** Log-model evidence favors the asymmetric pathways model (M5) over simpler models (M1-M4). **E-G.** Trial-by-trial proportion of Go responses, mean proportion Go responses, and probability of staying based on one-step-ahead predictions using parameters (hierarchical Bayesian inference) of the winning model (asymmetric pathways model, M5). **H.** Model frequency and protected exceedance probability indicate best fit for model M5 (asymmetric pathways model), in line with log model evidence.

3.3.3 fMRI: Basic quality control analyses

First, we performed a GLM as a quality-check to test which regions encoded positive (rewards, no punishments) vs. negative (no reward/ punishment) outcomes in a "model-free" way, independent of any model-based measure derived from a RL model (for full description of the GLM regressors and contrasts, see S3.8). Positive outcomes elicited a higher BOLD response in regions including ventromedial PFC (vmPFC), ventral striatum, and right hippocampus, while negative outcomes elicited higher BOLD in bilateral dorsolateral PFC (dlPFC), left ventrolateral PFC, and precuneous (Fig. 3.3A, see full report of significant clusters in S3.9).

We also assessed which regions encoded Go vs. NoGo as well as Go_{LEFT} vs. Go_{RIGHT} responses. There was higher BOLD for Go than NoGo responses at the time of response in dorsal

ACC (dACC), striatum, thalamus, motor cortices, and cerebellum, while BOLD was higher for NoGo than Go responses in right IFG (Fig. 3.6C left panel; see S3.9)(Algermissen et al. 2022). For lateralized Go responses, there was higher BOLD signal in contralateral motor cortex and operculum as well as ipsilateral cerebellum when contrasting hand responses against each other (Fig. 3.6C, right panel). These results are in line with previous results on outcome processing and response selection and thus assure the general data quality.

3.3.4 fMRI: Biased learning in prefrontal cortex and striatum

To test which brain regions were involved in biased learning, we performed a model-based GLM featuring the trial-by-trial PE update as a parametric regressor (see GLM notation in S3.8). We used the group-level parameters of the best fitting computational model (M5) to compute trial-by-trial belief updates (i.e., prediction error * learning rate) for every trial for every participant. In assessing neural signatures of biased learning, we faced the complication that standard (Rescorla-Wagner learning in M1) and biased PEs (winning model M5) were highly correlated. A mean correlation of 0.92 across participants (range 0.88–0.95) made it difficult to neurally distinguish biased from standard learning. To circumvent this collinearity problem, we decomposed the biased PE (computed using model M5) into the standard PE (computed using model M1) plus a difference term (Wittmann et al. 2008; Daw et al. 2011):

$$PE_{BIAS} = PE_{STD} + PE_{DIF}$$

A neural signature of biased learning should significantly—and with the same sign—encode both components of this biased PE term. Standard PEs and the difference term were uncorrelated (mean correlation of -0.02 across participants; range -0.33–0.24). We tested for biased prediction errors PE_{BIAS} by testing which regions significantly encoded the conjunction of both its components. In other words, we assessed which regions significantly encoded both the standard prediction errors PE_{STD} and the difference to biased PEs PE_{DIF}. Significant encoding of both components (with the same sign) provided strong evidence for encoding of biased prediction errors PE_{BIAS}.

While PE_{STD} was encoded in a range of cortical and subcortical regions (Fig. 3.3B, S3.9) previously reported in the literature (Bartra et al. 2013), significant encoding of both PE_{STD} and PE_{DIF} (conjunction) occurred in striatum (caudate, nucleus accumbens), dACC (area 23/24), perigenual ACC (pgACC; area 32d bordering posterior vmPFC), posterior cingulate cortex (PCC), left motor cortex, left inferior temporal gyrus, and early visual regions (Fig. 3.3C; see full report of significant clusters in S3.9). Thus, BOLD signal in these regions was better described (i.e., more variance explained) by biased learning than by standard prediction error learning.

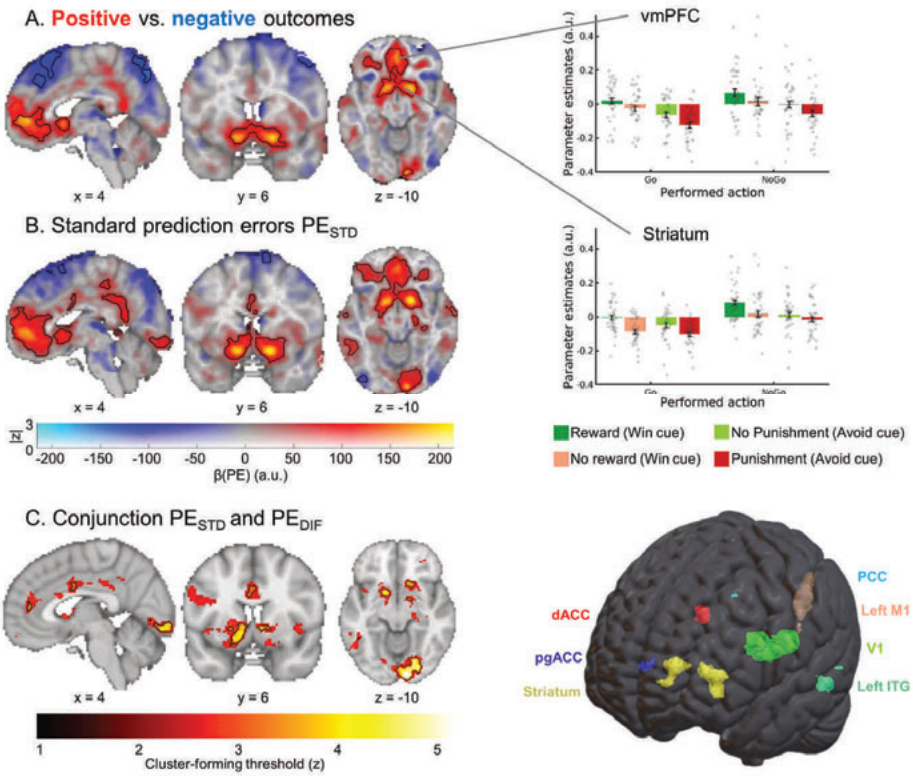


Figure 3.3. BOLD signal reflecting biased outcome processing.
BOLD effects displayed using a dual-coding visualization: color indicates the parameter estimates and opacity the associated z-statistics. Significant clusters are surrounded by black edges. **A.** Significantly higher BOLD signal for positive outcomes (rewards, no punishments) compared with negative outcomes (no rewards, punishments) was present in a range of regions including bilateral ventral striatum and vmPFC. Bar plots show mean parameter estimates per condition (±SEM across participants; dots indicating individual participants) **B.** BOLD signals correlated positively to "standard" RL prediction errors in several regions, including the ventral striatum, dACC, vmPFC, and PCC. **C.** Left panel: Regions encoding both the standard PE term and the difference term to biased PEs (conjunction) at different cluster-forming thresholds ($1 < z < 5$, color coding; opacity constant). Clusters significant at a threshold of $z > 3.1$ are surrounded by black edges. In bilateral striatum, dACC, pgACC, PCC, left motor cortex, left inferior temporal gyrus, and primary visual cortex, BOLD was significantly better explained by biased learning than by standard learning. Right panel: 3D representation with all seven regions encoding biased learning (and used in fMRI-informed EEG analyses).

3.3.5 EEG: Biased learning in midfrontal delta, theta, and beta power

Similar to the fMRI analyses, we next tested whether midfrontal power encoded biased PEs rather than standard PEs. While fMRI provides spatial specificity of where PEs are encoded, EEG power provides temporal specificity of when signals encoding prediction errors occur (Cohen, Wilmes, et al. 2011; Marco-Pallarés et al. 2015). In line with our fMRI analysis, we used the standard PE term PE_{STD} and the difference to the biased PE term PE_{DIF} as trial-by-trial regressors for EEG power at each channel-time-frequency bin for each participant and then performed cluster-based permutation tests across the *b*-maps of all participants. Note that differently from BOLD signal, EEG signatures of learning typically do not encode the full

prediction error. Instead, PE sign (positive vs. negative outcomes) and PE magnitude (saliency, surprise) have been found encoded in the theta and delta band, respectively, but with opposite signs, partially occluding each other (Talmi et al. 2013; Bernat et al. 2015; Cavanagh 2015). When testing for parametric correlates of PE magnitude, we controlled for PE sign, which implies that we tested for correlations with the absolute PE magnitude. Note that PE sign was identical for standard and biased PEs; only PE magnitude distinguished both learning models.

Both midfrontal theta and beta power reflected outcome valence, i.e., the PE sign: Theta power was higher for negative (non-reward and punishment) than for positive (reward and non-punishment) outcomes (225–475 ms, p = .006; Fig. 3.4A-B), while beta power was higher for positive than for negative outcomes (300–1,250 ms, p = .002; Fig. 3.4A, C). Differences in theta power were clearly strongest over frontal channels, while differences in the beta range were more diffuse, spreading over frontal and parietal channels (Fig. 3.4B-C). All results held when the condition-wise ERP was removed from the data (see S3.10), suggesting that differences between conditions were due to induced (rather than evoked) activity (for results in the time domain, see S3.11).

When testing for correlates of PE magnitude, we controlled for PE sign given that previous studies have reported TF correlates of both PE sign and PE magnitude in a similar time and frequency range, but with opposite signs, partially occluding each other (Talmi et al. 2013; Bernat et al. 2015; Cavanagh 2015). Midfrontal delta power was indeed positively correlated with the PE_{BIAS} term (225–475 ms; p = .017; Fig. 3.4D). In contrast, this correlation was not significant for the PE_{STD} term (p = 0.074, Fig. 3.4E) nor for the PE_{DIF} term (p = 0.185; Fig. 3.4F). This result does not imply that the PE_{BIAS} term explained delta power significantly better than the PE_{STD} term; it only implied significant encoding of the PE_{BIAS} term as suggested by the model that best fitted the behavioral data, with no significant evidence for a similar encoding of the conventional PE_{STD} term. For a similar observation in the time-domain EEG signal, see S3.12. Beyond delta power, beta power correlated positively, though not significantly with PE_{STD} (p = 0.110, Fig. 3.4E) and significantly negatively with PE_{DIF} (p = .001, 425–850 ms). Given these oppositely-signed correlations of its constituents, the PE_{BIAS} term did not significantly correlate with beta power (p = 0.550, Fig 4D).

In sum, both midfrontal theta power (negatively) and beta power (positively) encoded PE sign. In addition, delta power encoded PE magnitude (positively). This encoding was only significant for biased PEs, but not standard PEs. Taken together, as was the case for BOLD signal, midfrontal EEG power also reflected biased learning. As a next step, we tested whether the identified EEG phenomena were correlated with trial-by-trial BOLD signal in identified regions. Crucially, this allowed us to test whether EEG correlates of cortical learning precede EEG correlates of subcortical learning.

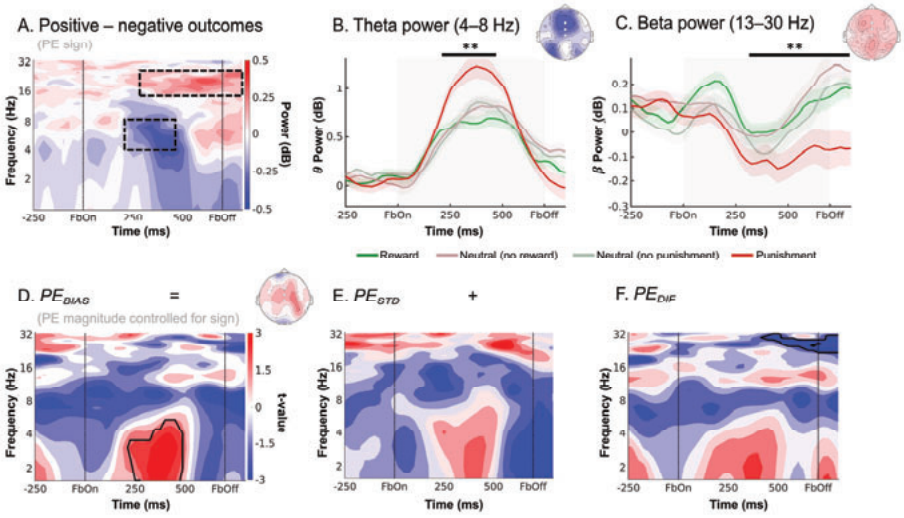


Figure 3.4. EEG time-frequency power over midfrontal electrodes (Fz/ FCz/ Cz). reflecting biased outcome processing.
A. Time-frequency plot (logarithmic y-axis) displaying higher theta (4–8 Hz) power for negative (non-reward and punishment) outcomes and higher beta power (16–32 Hz) for positive (reward and non-punishment) outcomes. Black square dot boxes indicate clusters above threshold that drive significance in a-priori defined frequency ranges. **B.** Theta power transiently increases for any outcome, but more so for negative outcomes (especially punishments) around 225–475 ms after feedback onset. Black horizontal lines indicate the time range for which the cluster driving significance was above threshold. (C) Beta power was higher for positive than negative outcomes over a long time period around 300–1,250 ms after feedback onset. **D-F.** Correlations between midfrontal EEG power and trial-by-trial PE magnitudes controlling for PE sign (i.e., the PE magnitude or "absolute" PE). Plots display correlates of two different operationalizations of the PE magnitude under different model assumptions (while panels A–C show the PE sign/ outcome valence, which was common to all models). Solid black lines indicate clusters above threshold. Biased PEs were significantly positively correlated with midfrontal delta power (**D**). The correlations of delta with the standard PEs (**E**) and the difference term to biased PEs (**F**) were positive as well, though not significant. Beta power only significantly encoded the difference term to biased PEs (F). ** $p < 0.01$.

3.3.6 Combined EEG-fMRI: Prefrontal cortex signals precede striatum during biased outcome processing

The observation that also cortical areas (dACC, pgACC, PCC) show biased PEs is consistent with the "external model" of cortical signals biasing learning processes in the striatum. However, this model makes the crucial prediction that these biased learning signals should be present first in cortical areas and only later in the striatum. Next, we used trial-by-trial BOLD signal from those regions encoding biased PE to predict midfrontal EEG power. By determining the time points at which different regions correlated with EEG power, we were able to infer the relative order of biased PE processing across cortical and subcortical regions, revealing whether cortical processing preceded striatal processing. We used trial-by-trial BOLD signal from the seven regions encoding biased PEs, i.e., striatum, dACC, pgACC, PCC, left motor cortex, left ITG, and primary visual cortex (see masks in S3.7) as regressors on average EEG power over midfrontal electrodes (Fz/ FCz/ Cz; see S3.13 for a graphical illustration of this approach). We performed analyses with and without PEs included in the model, which yielded identical results and suggested that EEG-fMRI correlations were not merely driven by both modalities reflecting PEs (as a "common cause"). Instead, EEG-fMRI correlations reflected incremental variance explained in EEG power that was

afforded by the BOLD signal in selected regions (even beyond variance explained by the PEs). As the timeseries of all seven regions were included in one single regression, their regression weights reflected each region's unique contribution, controlling for any shared variance. In line with the "external model", BOLD signal from prefrontal cortical regions correlated with midfrontal EEG power earlier after outcome onset than did striatal BOLD signal:

First, dACC BOLD was significantly negatively correlated with alpha/ theta power early after outcome onset (100–575 ms, 2–17 Hz, p = .016; Fig. 3.5A). This cluster started in the alpha/ theta range and then spread into the theta/delta range (henceforth called "lower alpha band power"). It was not observed in the EEG-only analyses reported above.

Second, while pgACC BOLD did not correlate significantly with midfrontal EEG power (p = .184), BOLD in PCC was negatively correlated with theta/ delta power (Fig. 3.5B; 175–500 ms, 1–6 Hz, p = .014). This finding bore resemblance in terms of time-frequency space to the cluster of (negative) PE sign encoding in the theta band and (positive) PE magnitude encoding in the delta band identified in the EEG-only analyses (Fig. 3.4A). Complementarily to the fMRI-informed EEG analyses, we also performed EEG-informed fMRI analyses using the trial-by-trial EEG signal in the cluster identified in the EEG-only analyses (see Fig. 3.4 A, B) to predict BOLD signal across the brain (see S3.13 for a graphical illustration of this approach). Both analysis approaches were blind to each other's results. Using trial-by-trial power in the midfrontal theta/delta band cluster identified in the EEG-only analyses (Fig. 3.4A, B), we observed significant clusters of negative EEG-BOLD correlation in vmPFC and PCC (Fig. 3.5F; S3.16). We thus discuss vmPFC and PCC together in the following.

Third, there was a significant positive correlation between striatal BOLD and midfrontal beta/ alpha power (driven by a cluster at 100–800 ms, 7–23 Hz, p = .010; Fig. 3.5C). This finding bore resemblance in time-frequency space to the cluster of positive PE sign encoding in beta power identified in the EEG-only analyses (Fig. 3.4A, C). Again, to substantiate this link, we used trial-by-trial midfrontal beta power in the cluster identified in the EEG-only analyses (see Fig. 3.4A, C) to predict BOLD signal across the brain. Clusters of positive EEG-BOLD correlations in right dorsal caudate (and left parahippocampal gyrus) as well as clusters of negative correlations in bilateral dorsolateral PFC (dlPFC) and supramarginal gyrus (SMG; Fig. 3.5G; see S3.16) confirmed the positive striatal BOLD-beta power association. Given that the striatum is unlikely to be the source of midfrontal beta power over the scalp, these results furthermore suggest dlPFC and SMG as likely candidate sources.

Finally, regarding the other three regions that showed a significant BOLD signature of biased PEs, BOLD in left motor cortex was significantly negatively correlated with midfrontal beta power (p = .002; around 0–625 ms; see S3.14). There were no significant correlations between midfrontal EEG power and left inferior temporal gyrus or primary visual cortex BOLD (see S3.14). All results were robust to different analysis approaches including shorter trial windows, different GLM specifications, inclusion of task-condition and fMRI motion realignment regressors, and individual modelling of each region. TF results were not reducible to phenomena in the time domain (see S3.15).

In sum, there were negative correlations between dACC BOLD and midfrontal lower alpha band power early after outcome onset, negative correlations between PCC BOLD and midfrontal

theta/ delta power at intermediate time points, and positive correlations between striatal BOLD and midfrontal beta power at late time points (Fig. 3.5D, H). These results are consistent with an "external model" of motivational biases arising from early cortical processes biasing later learning processes in the striatum.

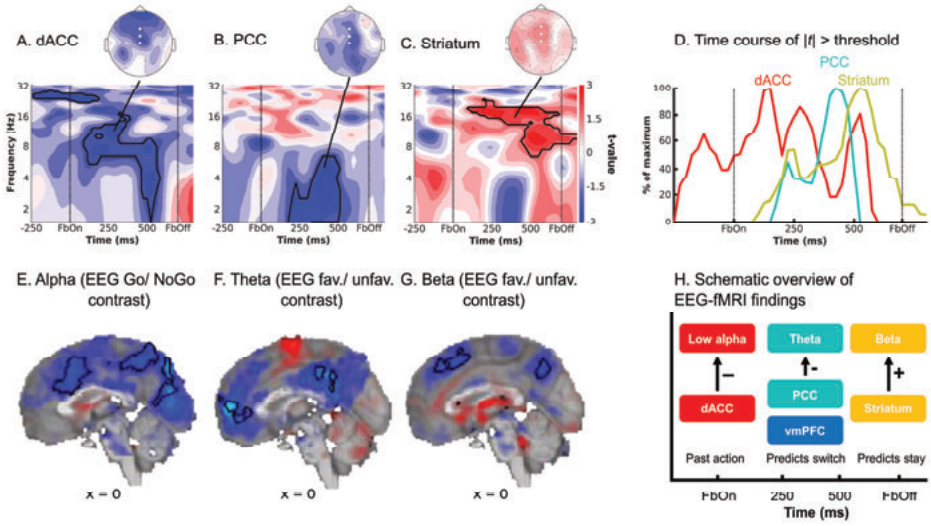


Figure 3.5. fMRI-informed EEG analyses.
Unique temporal contributions of BOLD signal in (**A**) dACC, (**B**) PCC, and (**C**) striatum to average EEG power over midfrontal electrodes (Fz/ FCz/ Cz). Group-level *t*-maps display the modulation of the EEG power by trial-by-trial BOLD signal in the selected ROIs. dACC BOLD correlated negatively with early alpha/ theta power, PCC BOLD negatively with theta/ delta power, and striatal BOLD positively with beta/ alpha power. Areas surrounded by a black edge indicate clusters of $|t| > 2$ with $p < .05$ (cluster-corrected). Topoplots indicate the topography of the respective cluster. **D**. Time course of dACC, PCC, and striatal BOLD correlations, normalized to the peak of the time course of each region. dACC-lower alpha band correlations emerged first, followed by (negative) PCC-theta correlations and finally positive striatum-beta correlations. The reverse approach using lower alpha (**E**), theta (**F**) and beta (**G**) power as trial-by-trial regressors in fMRI GLMs corroborated the fMRI-informed EEG analyses: Lower alpha band power correlated negatively with the dACC BOLD, theta power negatively with vmPFC and PCC BOLD, and beta power positively with striatal BOLD. **H.** Schematic overview of the main EEG-fMRI results: dACC encoded the previously performed response and correlated with early midfrontal lower alpha band power. vmPFC/ PCC (correlated with theta power) and striatum (correlated with beta power) both encoded outcome valence, but had opposite effects on subsequent behavior. Note that activity in these regions temporally overlaps; boxes are ordered in temporal precedence of peak activity.

3.3.7 dACC BOLD and midfrontal lower alpha band power encode the previously performed action during outcome presentation

While the clusters of EEG-fMRI correlation in the theta/ delta and beta range matched the clusters identified in EEG-only analyses, the cluster of negative correlations between dACC BOLD and early midfrontal lower alpha band power was novel and did not match our expectations. Given that these correlations arose very soon after outcome onset, we hypothesized that dACC BOLD and midfrontal lower alpha band power might reflect a process occurring even before outcome onset, such as the maintenance ("memory trace") of the previously performed response to which credit may later be assigned. We therefore assessed whether information of the

previous response was present in dACC BOLD and in the lower alpha band around the time of outcome onset.

First, we tested for BOLD correlates of the previous response at the time of *outcomes* (eight outcome-locked regressors for every Go/ NoGo x reward/ no reward/ no punishment/ punishment combination) while controlling for motor-related signals at the time of the *response* (response-locked regressors for left-hand and right-hand button presses). At the time of outcomes, there was higher BOLD signal for NoGo than Go responses across several cortical and subcortical regions, peaking in both the dACC and striatum (Fig. 3.6D). This inversion of effects—higher BOLD for Go than NoGo responses at the time of response (see quality checks), but the reverse at the time of outcome—was also observed in the upsampled raw BOLD and was independent of the response of the next trial (S3.17). In sum, large parts of cortex, including the dACC, encoded the previously performed response at the moment outcomes were presented, in line with the idea that the dACC maintains a "memory trace" of the previously performed response.

Second, we tested for differences between Go and NoGo responses at the time of outcomes in midfrontal broadband EEG power. Power was significantly higher on trials with Go than on trials with NoGo responses, driven by clusters in the lower alpha band (spreading into the theta band; around 0.000–0.425 sec., 1–11 Hz, p = .012) and in the beta band (around 0.200–0.450 sec., 18–27 Hz, p = .022; Fig. 3.6A, B). The first cluster matched the time-frequency pattern of dACC BOLD-alpha power correlations (Fig. 3.5A).

If this activity cluster contained a signature of the previously performed response, it might have been present throughout the delay between cue offset and outcome onset. When repeating the above permutation test including the last second before outcome onset, there were significant differences again, driven by a sustained cluster in the beta band (-1–0 sec., 13–33 Hz, p = .002) and two clusters in the alpha/ theta band (Cluster 1: -1.000– -0.275 sec., 1–10 Hz, p = 0.014; Cluster 2: -0.225–0.425 sec., 1–11 Hz, p = .022; Fig. 3.6B). These findings suggest that lower alpha band power might reflect a sustained memory of the previously performed response. Supplemental analyses (S3.17) yielded that this Go-NoGo trace during outcome processing did not change over the time course of the experiment, suggesting that it did not reflect typical fatigue/ time-on task effects often observed in the alpha band.

Again, we performed the reverse EEG-fMRI analysis using trial-by-trial power in the identified lower alpha band cluster (Fig. 3.6B) as an additional regressor in the quality-check fMRI GLM. Clusters of negative EEG-BOLD occurred correlation in a range of cortical regions, including dACC and precuneous (Fig. 3.5E; see S3.16). In sum, both dACC BOLD signal and midfrontal lower alpha band power contained information about the previously performed response, consistent with the idea that both signals reflect a "memory trace" of the response to which credit is assigned once an outcome is obtained.

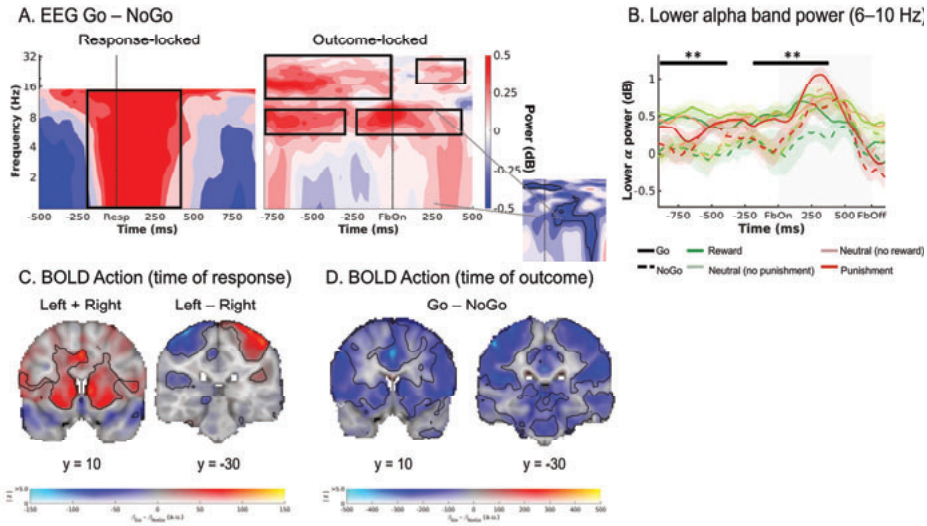


Figure 3.6. Exploratory follow-up analyses on dACC BOLD signal and midfrontal lower alpha band power.
A. Midfrontal time-frequency response-locked (left panel) and outcome-locked (right panel). Before and shortly after outcome onset, power in the lower alpha band was higher on trials with Go actions than on trials with NoGo actions. The shape of this difference resembles the shape of dACC BOLD-EEG TF correlations (small plot; note that this plot depicts BOLD-EEG correlations, which were negative). Note that differences between Go and NoGo trials occurred already before outcome onset in the alpha and beta range, reminiscent of delay activity, but were not fully sustained throughout the delay between response and outcome. **B.** Midfrontal power in the lower alpha band per action x outcome condition. Lower alpha band power was consistently higher on trials with Go actions than on trials with NoGo actions, starting already before outcome onset. **C.** BOLD signal differences between Go and NoGo actions (left panel) and left vs. right hand responses (right panel) at the time or responses. Response-locked dACC BOLD signal was significantly higher for Go than NoGo actions. **D.** BOLD signal differences between Go and NoGo actions at the time of outcomes. Outcome-locked dACC BOLD signal (and BOLD signal in other parts of cortex) was significantly lower on trials with Go than on trials with NoGo actions.

3.3.8 Striatal and vmPFC/ PCC BOLD differentially relate to action policy updating

EEG correlates of PCC BOLD and striatal BOLD occurred later than for the dACC BOLD, and overlapped with classical feedback-related midfrontal theta and beta power responses. We hypothesized that those neural signals might be more closely related to updating of action policies (i.e., which action to perform for each cue) and might thus predict the next response to the same cue (Frank, Woroch, et al. 2005; Cavanagh et al. 2010). We thus used the trial-by-trial BOLD responses in dACC, vmPFC, PCC, and striatum to predict whether participants would repeat the same response on the next trial with the same cue ("stay") or switch to another response ("shift"). Mixed-effects logistic regression yielded that dACC BOLD did not significantly predict response repetition (b = -0.019, SE = 0.016, $\chi^2(1)$ = 1.294, p = .255). In contrast, BOLD in PCC/ vmPFC and striatum did predict response repetition, though in opposite directions: Participants were significantly more likely to *repeat* the same response when striatal BOLD was high (b = 0.067, SE = 0.024, $\chi^2(1)$ = 9.051, p = .003), but more likely to *switch* to another response when vmPFC BOLD (b = -0.065, SE = 0.020, $\chi^2(1)$ = 8.765, p = .003) or PCC BOLD (b = -0.036, SE = 0.016, $\chi^2(1)$ = 3.691, p = .030; Fig. 3.5H) was high (for plots, see S3.18). Similarly, high pgACC BOLD predicted a higher likelihood of switching, associating it with circuit formed by vmPFC and PCC (b = -0.076,

SE = 0.017, $\chi^2(1) = 15.559$, $p < .001$) We also inspected the raw upsampled HRF shapes per region per condition, confirming that differential relationships were not driven by differences in HRF shapes across regions.

We also tested whether trial-by-trial midfrontal lower alpha band, theta, or beta power (within the clusters identified in the EEG-only analyses) predicted action policy updating. Participants were significantly more likely to repeat the same response when beta power was high ($b = 0.145$, *SE* = 0.041, $\chi^2(1) = 11.886$, $p < .001$), but more likely to switch when theta power was high ($b = -0.099$, *SE* = 0.047, $\chi^2(1) = 4.179$, $p = .041$). Notably, unlike its BOLD correlate in ACC, lower alpha band power did predict response repetition, with more repetition when alpha power was high ($b = .0.179$, SE = 0.052, $\chi^2(1) = 10.711$, $p = .001$; for plots, see S3.18).

In sum, high striatal BOLD and midfrontal beta power predicted that the same response would be repeated on the next encounter of a cue, while high vmPFC and PCC BOLD and high theta power predicted that participants would switch to another response. Thus, although both striatal and vmPFC/ PCC BOLD positively encoded biased prediction errors, these two sets of regions had opposite roles in learning: while the striatum reinforced previous responses, vmPFC/ PCC triggered the shift to another response strategy (Fig. 3.5H).

3.4 Discussion

We investigated neural correlates of biased learning for Go and NoGo responses. In line with previous research (Swart et al. 2017, 2018), participants' behavior was best described by a computational model featuring faster learning from rewarded Go responses and slower learning from punished NoGo responses. Neural correlates of biased PEs were present in BOLD signals in several regions, including ACC, PCC, and striatum. These regions exhibited distinct midfrontal EEG power correlates. Most importantly, correlates of prefrontal cortical BOLD preceded correlates of striatal BOLD: Trial-by-trial dACC BOLD correlated with lower alpha band power immediately after outcome onset, followed by PCC (and vmPFC) BOLD correlated with theta power, and finally, striatal BOLD correlated with beta power. These results suggest that the architecture of the asymmetric striatal pathways might not be the only neural structure that gives rise to motivational learning biases; instead, the PFC might critically contribute to these biases.

The observation that both PFC and striatal BOLD signal reflected biased PEs might be explained by three different models. One model assumes that both PFC and striatal processes arrive at biased learning independently of each other, which is highly unlikely given strong recurrent connections between both regions (Haber 2003; Frank 2005; Collins and Frank 2014). Another model incorporates such interconnections, but assumes the that striatum leads the PFC. While such a model is in line with past animal studies (Pasupathy and Miller 2005) and modeling work (Wang et al. 2018), it would predict EEG correlates of the PFC to trail after EEG correlates of the striatum—or at least to occur with considerable delay after outcome onset. This model is not supported by our findings, which showed EEG correlates of PFC regions soon after outcome onset, preceding striatal EEG correlates. The only model consistent with our data assumes recurrent connections between PFC and striatum, but with the PFC leading the striatum. Hence, these results are in line with a model of PFC biasing striatal outcome processing, giving rise to motivational learning biases in behavior.

The dominant idea about the origin of motivational biases has been that these biases are an emergent feature of the asymmetric direct/ indirect pathway architecture in the basal ganglia (Collins and Frank 2014; Guitart-Masip, Duzel, et al. 2014). We find that these biases are present first in prefrontal cortical areas, notably dACC and PCC, which argues against biases being purely driven by subcortical circuits. Rather, motivational learning biases might be an instance of sophisticated, even "model-based" learning processes in the striatum instructed by the prefrontal cortex (Sharpe et al. 2017, 2019). An influence of PFC on striatal RL has prominently been observed in the case of model-based vs. model-free learning (Lee et al. 2014; Piray et al. 2016) and has been stipulated as a mechanism of how instructions can impact RL (Doll et al. 2009; Atlas et al. 2016). Although there are reports of striatal processes preceding prefrontal processes within learning tasks (Pasupathy and Miller 2005; Antzoulatos and Miller 2014), the opposite pattern of PFC preceding striatum has been observed as well (Seo et al. 2012) and a causal impact of PFC on striatal learning is well established (van Schouwenburg et al. 2012; Howard et al. 2020). In particular, we have previously observed that motivational response biases during action selection arise from early prefrontal inputs to the striatum, as well(Algermissen et al. 2022). Prefrontal influences on striatal processes might thus be a common momentum to both motivational response and learning biases.

The particular subregion of PFC showing the earliest EEG correlates was the dACC. This observation is in line with an earlier EEG-fMRI study reporting dACC to be part of an early valuation system preceding a later system comprising vmPFC and striatum (Fouragnan et al. 2015). The dACC has been suggested to encode models of agents' environment (Alexander and Brown 2011, 2018) that are relevant for interpreting outcomes, with BOLD in this region scaling with the size of PEs (Behrens et al. 2007; Meder et al. 2017) and indexing how much should be learned from new outcomes. We hypothesize that, at the moment of outcome, dACC maintains a "memory trace" of the previously performed response (Enel et al. 2020) which might modulate the processing of outcomes as soon as they become available (Shadmehr et al. 2010; Vyas et al. 2020). Notably, dACC exhibited stronger BOLD signal for Go than NoGo responses at the time of participants' response, but this pattern reversed at the time of outcomes. This reversal rules out the possibility that response-locked BOLD signal simply spilled over into the time of outcomes. Future research will be necessary to corroborate such a motor "memory trace" in dACC. In sum, the dACC might be in a designated position to inform subsequent outcome processing in downstream regions by modulating the learning rate as a function of the previously performed response and the obtained outcome. Rather than striatal circuits being sufficient for the emergence of motivational biases, the more "flexible" PFC seems to play an important role in instructing downstream striatal learning processes.

Striatal, dACC and PCC BOLD encoded biased PEs. In line with previous research, striatal BOLD positively linked to midfrontal beta power (Sadaghiani et al. 2010; Andreou et al. 2017), which positively encoded PE sign (Marco-Pallarés et al. 2008, 2015; van de Vijver et al. 2011). PCC and vmPFC BOLD negatively linked to midfrontal theta/ delta power (Scheeringa et al. 2008, 2009; Algermissen et al. 2022), which encoded PE sign negatively, but PE magnitude positively. Notably, theta/ delta power correlates of vmPFC/ PCC BOLD preceded beta power correlates of striatal BOLD in time, which aligns with previous findings of motivational response biases being first visible in the vmPFC BOLD before they impact striatal action selection (Algermissen

et al. 2022). Note however that EEG correlates of striatal BOLD during outcome processing were in the beta band, while we previously found correlates of striatal BOLD during action selection in the theta band (Algermissen et al. 2022). This dissociation suggests important differences in the roles of the striatum during these two processes. Also, these findings highlight that EEG-fMRI correlations in this study likely did not reflect resting-state associations, but signatures of task-induced events that were specific to the trial phase. In sum, while these EEG-fMRI findings on outcome processing relate to EEG-fMRI findings on action selection in that prefrontal signals precede striatal signals, EEG correlates were markedly dissimilar between both processes, highlighting their specificity to certain cognitive operations.

Positive encoding of prediction errors in striatal BOLD signal is a well-established phenomenon (Bartra et al. 2013; Fouragnan et al. 2018). Striatal BOLD was better described by biased PEs than by standard PEs, corroborating the presence of motivational learning biases also in striatal learning processes. Notably, EEG correlates of striatal BOLD peaked rather late, suggesting that these processes are informed by early sources in PFC which are connected to the striatum via recurrent feedback loops (Haber 2003; Frank 2005). Positive prediction errors increase the value of a performed action and thus strengthen action policies. Hence, it is not surprising that high striatal BOLD signal and midfrontal beta power predicted action repetition (Engel and Fries 2010; Feingold et al. 2015).

In contrast to striatal learning signals, the PCC and vmPFC BOLD as well as midfrontal theta and delta power signals were more complicated: Theta encoded PE sign, delta encoded PE magnitude. Both correlates showed opposite polarities. This observation is in line with previous literature suggesting that midfrontal theta and delta power might reflect the "saliency" or "surprise" aspect of PEs (Talmi et al. 2013; Hauser et al. 2014; Cavanagh 2015). Surprises have the potential to disrupt an ongoing action policy (Wessel and Aron 2017) and motivate a shift to another policy, which might explain why these signals predicted switching to another response (Domenech et al. 2020; Trudel et al. 2021). Notably, this EEG surprise signal was only significantly correlated with the biased (but not the standard) PE term, corroborating that the surprise attributed to outcomes depends on the previously performed response in line with motivational learning biases. In sum, both vmPFC and striatum encode biased PEs, though with different consequences for future action policies.

Taken together, distinct brain regions processed outcomes in a biased fashion at distinct time points with distinct EEG power correlates. Simultaneous EEG-fMRI recordings allowed us to infer when those regions reached their peak activity (Hauser et al. 2015). However, the correlational nature of BOLD-EEG links precludes strong statements about these regions actually generating the respective power phenomena. Alternatively, activity in those regions might merely modulate the amplitude of time-frequency responses originating from other sources. Furthermore, while the observed associations align with previous literature (Scheeringa et al. 2008, 2009; Sadaghiani et al. 2010; Andreou et al. 2017; Algermissen et al. 2022), the considerable distance of the striatum to the scalp raises the question whether scalp EEG could in principle reflect striatal activity, at all (Cohen, Cavanagh, et al. 2011; Foti et al. 2011). Intracranial recordings have observed beta oscillations during outcome processing in the striatum before (Feingold et al. 2015; Amemori et al. 2018, 2020). Also, our analysis controlled for BOLD signal in motor cortex, an alternative candidate source for beta power, suggesting that late midfrontal beta power did not merely reflect

motor cortex beta. Even if the striatum is not the generator of the beta oscillations over the scalp, their true (cortical) generator might be tightly coupled to the striatum and thus act as a "transmitter" of striatal beta oscillations. In fact, the analyses using trial-by-trial beta power to predict BOLD yielded significant clusters in dlPFC and SMG, two candidate regions for such a "transmitter".

We observed EEG correlates of striatal BOLD at a rather late time point after outcome onset. While we conclude that biased outcome processing occurs much earlier in cortical regions than the striatum, it is possible that the modulating influence of the striatum on cortical beta synchronization just takes time to surface. However, speaking against this, some single studies have reported maximal correlations between striatal LFPs and scalp EEG at a time lag of 0(Cohen et al. 2009). Regardless, even in the presence of a non-zero lag, our main conclusion would hold: Biased learning is present in cortical regions early after outcome onset, which cannot be a consequence of striatal input, but must constitute an independent origin of motivational learning biases.

Finally, the correlational nature of the study prevents strong statements over any causal interactions between the observed regions. We assume here that a region showing an earlier midfrontal EEG correlate influences other regions showing later midfrontal EEG correlates, and such an influence is plausible given findings of feedback loops between prefrontal regions and the striatum (Haber 2003). Future studies targeting those regions via selective causal manipulations will be necessary to test for the causal role of PFC in informing striatal learning.

In conclusion, biased learning—increased credit assignment to rewarded action, decreased credit assignment to punished inaction—was visible both in behavior and in BOLD signal in a range of regions. EEG correlates of prefrontal cortical regions, notably dACC and pgACC, *preceded* correlates of the striatum, consistent with a model of the PFC biasing RL in the striatum. The dACC appeared to hold a "motor memory trace" of the past response, biasing early outcome processing. Subsequently, biased learning was also present in vmPFC/ PCC and striatum, with opposite roles in adjusting vs. maintaining action policies. These results refine previous views on the neural origin of these learning biases, which might not only rely on parts of the brain associated with rigid, habit-like responding, but rather incorporate sophisticated, even "model-based" processes relying on frontal inputs that are associated with counterfactual reasoning and increased behavioral flexibility (Boorman et al. 2009; Fouragnan et al. 2019). The PFC is typically believed to facilitate goal-directed over instinctive processes. Hence, PFC involvement into biased learning suggests that these biases are not necessarily agents' inescapable "fate", but rather likely act as global "priors" that facilitate learning of more local relationships. They allow for combining "the best of both worlds"—long-term experience with consequences of actions and inactions together with flexible learning from rewards and punishments.

3.5 MATERIALS AND METHODS

3.5.1 Participants

Thirty-six participants (M_{age} = 23.6, SD_{age} = 3.4, range 19–32; 25 women; all right-handed; all normal or corrected-to-normal vision) took part in a single 3-h data collection session, for which they received €30 flat fee plus a performance-dependent bonus (range €0–5, M_{bonus} = €1.28, SD_{bonus}

= 1.54). The study was approved by the local ethics committee (CMO2014/288; Commissie Mensengeboden Onderzoek Arnhem-Nijmegen) and all participants provided written informed consent. Exclusion criteria comprised claustrophobia, allergy to gels used for EEG electrode application, hearing aids, impaired vision, colorblindness, history of neurological or psychiatric diseases (including heavy concussions and brain surgery), epilepsy and metal parts in the body, or heart problems. Sample size was based on previous EEG studies with a comparable paradigm (Cavanagh et al. 2013; Swart et al. 2018).

Behavioral and modeling results include all 36 participants. The following participants were excluded from analyses of neural data: For two participants, fMRI functional-to-standard image registration failed; hence, all fMRI-only results are based on 34 participants (M_{age} = 23.47, 25 women). Four participants exhibited excessive residual noise in their EEG data (> 33% rejected trials) and were thus excluded from all EEG analyses; hence, all EEG-only analyses are based on 32 participants (M_{age} = 23.09, 23 women). For combined EEG-fMRI analyses, we excluded the above-mentioned six participants plus one more participant whose regression weights for every regressor were about ten times larger than for other participants, leaving 29 participants (M_{age} = 23.00, 22 women). Exclusions were in line with a previous analysis of this data set (Algermissen et al. 2022). fMRI- and EEG-only results held when analyzing only those 29 participants (see S3.1).

3.5.2 Task

Participants performed a motivational Go/ NoGo learning task (Swart et al. 2017, 2018) administered via MATLAB 2014b (MathWorks, Natick, MA, United States) and Psychtoolbox-3.0.13. On each trial, participants saw a gem-shaped cue for 1300 ms which signaled whether they could potentially win a reward (Win cues) or avoid a punishment (Avoid cues) and whether they had to perform a Go (Go cue) or NoGo response (NoGo cue). They could press a left (Go_{LEFT}), right (Go_{RIGHT}), or no (NoGo) button while the cue was presented. Only one response option was correct per cue. Participants had to learn both cue valence and required action from trial-and-error. After a variable inter-stimulus-interval of 1,400–1,600 ms, the outcome was presented for 750 ms. Potential outcomes were a reward (symbolized by coins falling into a can) or neutral outcome (can without money) for Win cues, and a neutral outcome or punishment (symbolized by money falling out of a can) for Avoid cues. Feedback validity was 80%, i.e., correct responses were followed by positive outcomes (rewards/ no punishments) on only 80% of trials, while incorrect responses were still followed by positive outcomes on 20% of trials. Trials ended with a jittered inter-trial interval of 1250–2000 ms, yielding total trial lengths of 4700–6650 ms.

Participants gave left and right Go responses via two button boxes positioned lateral to their body. Each box featured four buttons, but only one button per box was required in this task. When participants accidentally pressed a non-instructed button, they received the message "Please press one of the correct keys" instead of an outcome. In the analyses, these responses were recoded into the instructed button on the respective button box. In the fMRI GLMs, such trials were modeled with a separate regressor.

Before the task, participants were instructed that each cue could be followed by either reward or punishment, that each cue had one optimal response, that feedback was probabilistic, and that the rewards and punishments were converted into a monetary bonus upon completion of the study. They performed an elaborate practice session in which they got familiarized first with each

condition separately (using practice stimuli) and finally practiced all conditions together. They then performed 640 trials of the main task, separated into two sessions of 320 trials with separate cue sets. Introducing a new set of cues allowed us to prevent ceiling effects in performance and investigate continuous learning throughout the task. Each session featured eight cues that were presented 40 times. After every 100–110 trials (~ 6 min.), participants could take a self-paced break. The assignment of the gems to cue conditions was counterbalanced across participants, and trial order was pseudo-random (preventing that the same cue occurred on more than two consecutive trials).

3.5.3 Behavior analyses

We used mixed-effects logistic regression (as implemented in the R package *lme4*) to analyze behavioral responses (Go vs. NoGo) as a function of required action (Go/ NoGo), cue valence (Win/ Avoid), and their interaction. We included a random intercept and all possible random slopes and correlations per participant to achieve a maximal random-effects structure (Barr et al. 2013). Sum-to-zero coding was employed for the factors. Type 3 *p*-values were based on likelihood ratio tests (implemented in the R package *afex*). We used a significance criterion of $\alpha = .05$ for all the analyses.

Furthermore, we used mixed-effects logistic regression to analyze "stay behavior", i.e., whether participants repeated an action on the next encounter of the same cue, as a function of outcome valence (positive: reward or no punishment/ negative: no reward or punishment), outcome salience (salient: reward or punishment/ neutral: no reward or no punishment), and performed action (Go/ NoGo). We again included all possible random intercepts, slopes, and correlations.

3.5.4 Computational modeling

We fit a series of increasingly complex RL models to participants' choices to decide between different algorithmic explanations for the emergence of motivational biases in behavior. We employed the same set of nested models as in previous studies using this task (Swart et al. 2017, 2018). For tests of alternative biases specifications, see S3.5 and S3.6.

3.5.4.1 Model space

To determine whether a Pavlovian response bias, a learning bias, or both biases jointly predicted behavior best, we fitted a series of increasing complex computational models. In each trial (t), choice probabilities for all three response options (a) given the displayed cue (s) were computed from their action weights (modified Q-values) using a softmax function:

$$p(a_t|s_t) = \frac{\exp\left(w(a_t,s_t)\right)}{\sum_a \exp\left(w(a,s_t)\right)} \quad (1)$$

After each response, action values were updated with the prediction error based on the obtained outcome $r \in \{-1; 0; 1\}$. As the starting model (M1), we fitted an standard delta-learning model (Rescorla and Wagner 1972) in which action values were updated with prediction errors, i.e., the deviation between the experienced outcome and expected outcome. This model contained two free parameters: the learning rate (ε) scaling the updating term and the feedback sensitivity (ϱ) scaling the received outcome (i.e., higher feedback sensitivity led to choices more

strongly guided by value difference, akin to the role of the inverse temperature parameter frequency used in reinforcement learning models):

$$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \varepsilon(\rho r - Q_{t-1}(a_t, s_t)) \quad (2)$$

In this model, choice probabilities were fully determined by action values, without any bias. We initialized action values Q_0 to the mean of the two possible outcomes for each cue type, with outcomes scaled by the feedback sensitivity parameter ϱ (Win cues: mean of $1*\varrho$ and $0*\varrho$; Avoid cues: mean of $0*\varrho$ and $-1*\varrho$). Unlike previous versions of the task (Swart et al. 2017), cue valences were not instructed, but had to be learned from outcomes, as well (Swart et al. 2018). Thus, until experiencing the first non-neutral outcome (reward or punishment) for a cue, participants could not know its valence and thus not learn from neutral feedback. Hence, for these early trials, action values were multiplied with zero when computing choice probabilities. After the first encounter of a valenced outcome, action values were "unmuted" and started to influence choices probabilities, retrospectively considering all previous outcomes.

In M2, we added the Go bias parameter *b*, which accounted for individual differences in participants' overall propensity to make Go responses, to the action values Q, resulting in action weights w:

$$w(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b & if\ a = Go \\ Q_t(a_t, s_t) & else \end{cases} \quad (3)$$

In M3, we added a Pavlovian response bias π, scaling how positive/ negative cue valence (Pavlovian values) increased/ decreased the weights of Go responses:

$$w(a_t, s_t) = \begin{cases} Q_t(a_t, s_t) + b + \pi V(s) & if\ a = Go \\ Q_t(a_t, s_t) & else \end{cases} \quad (4)$$

Participants were instructed that a cue was either a Win cue (affording rewards or neutral outcomes) or an Avoid cue (affording neutral outcomes or punishments). Hence, they could infer the cue valence (Win/ Avoid) as soon as they experienced a non-neutral outcome. Until that moment, cue valence $V(s)$ was set to zero. Afterwards, $V(s)$ was arbitrarily set to +0.5 for Win cues and -0.5 for Avoid cues. Note that choosing different values than 0.5 would merely rescale the bias parameter π (e.g., halving π with cue valences of +1 and -1) without any changes in the model's predictions. The Pavlovian response bias affected left-hand and right-hand Go responses similarly and thus reflected generalized activation/ inactivation by the cue valence.

In M4, we added a learning bias $\varkappa$, increasing the learning rate for rewards after Go responses and decreasing it for punishments after NoGo responses:

$$\varepsilon = \begin{cases} \varepsilon_0 + \kappa & if\ r_t = 1\ and\ a = go \\ \varepsilon_0 - \kappa & if r_t = -1\ and\ a = nogo \\ \varepsilon_0 & else \end{cases} \quad (5)$$

The learning bias was specific to the response shown, thus reflecting a specific enhancement in action learning/ impairment in unlearning for that particular response.

In the model M5, we included both the Pavlovian response bias and the learning bias.

The hyperpriors were $X_\rho \sim \mathcal{N}(2,3)$, $X_\varepsilon \sim \mathcal{N}(0,2)$, $X_{b,\pi,\kappa} \sim \mathcal{N}(0,3)$, chosen in line with previous implementations of this model(Swart et al. 2017, 2018). Alternative hyperpriors did not change the results. For computing the participant-level parameters, ϱ was exponentiated to constrain it to positive values, and the inverse-logit transformation was applied to ε to constraint it to the range [0 1] (Swart et al. 2017, 2018). We made sure that the effect of $\varkappa$ on ε was symmetrical by computing it as:

$$\varepsilon = \begin{cases} \varepsilon_0 = inv.logit(\varepsilon) & \\ \varepsilon_{punished\ NoGo} = inv.logit(\varepsilon - \kappa) & if\ \varepsilon_0 < .5 \\ \varepsilon_{rewarded\ Go} = \varepsilon_0 + (\varepsilon_0 - \varepsilon_{punished\ NoGo}) & if\ \varepsilon_0 < .5 \end{cases} \quad (6)$$

$$\varepsilon = \begin{cases} \varepsilon_{rewarded\ Go} = inv.logit(\varepsilon + \kappa) & if\ \varepsilon_0 > .5 \\ \varepsilon_{punished\ NoGo} = \varepsilon_0 - (\varepsilon_{rewarded\ Go} - \varepsilon_0) & if\ \varepsilon_0 > .5 \end{cases}$$

3.5.4.2 Model fitting and comparison

For model fitting and comparison, we used hierarchical Bayesian inference as implemented in the CBM toolbox in Matlab (Piray et al. 2019). This approach combines hierarchical Bayesian parameter estimation with random-effects model comparison (Stephan et al. 2009). The fitting procedure involves two steps, starting with the Laplace approximation of the model evidence to compute the group evidence, which quantifies how well each model fits the data while penalizing for model complexity. Both group-level and individual-level parameters are estimated using an iterative algorithm. We used wide Gaussian priors (see hyperpriors above) and exponential and sigmoid transforms to constrain parameter spaces. Subsequent random-effects model selection allows for the possibility that different models generated the data for different participants. Participants contribute to the group-level parameter estimation in proportion to how well a given model fits their data, quantified via a responsibility measure (i.e., the probability that the model at hand is responsible for generating data of the respective participant). This model-comparison approach has been shown to be less susceptible to the influence of outliers (Piray et al. 2019). We selected the "winning" model based on the protected exceedance probability.

3.5.4.3 Model validation

We assured that the winning model was able to reproduce the data, using the sampled combinations of participant-level parameter estimates to create 3600 agents that "played" the task. We employed two approaches to simulate the task: *posterior predictive model simulations* and *one-step-ahead model predictions*. In the posterior predictive model simulations, agents' choices were sampled probabilistically based on their action values, and outcomes probabilistically sampled based on their choices. This method ignores participant-specific choice histories and can thus yield choice/outcome sequences that diverge considerably from participants' actual experiences. In contrast, one-step-ahead predictions use participants' actual choices and experienced outcomes in each trial to update action values. We simulated choices for each participant using both methods, which confirmed that the winning model M5 ("asymmetric pathways model") was able to qualitatively reproduce the data, while an alternative implementation of biased learning ("action priming model") failed to do so (see S3.5).

3.5.5 fMRI data acquisition

fMRI data were collected on a 3T Siemens Magnetom Prisma fit MRI scanner with a 64-channel head coil. During scanning, participants' heads were restricted using foam pillows and strips of adhesive tape were applied to participants' forehead to provide active motion feedback and minimize head movement (Krause et al. 2019). After two localizer scans to position slices, we collected functional scans with a whole-brain T2*-weighted sequence (68 axial-oblique slices, TR = 1400 ms, TE = 32 ms, voxel size 2.0 mm isotropic, interslice gap 0 mm, interleaved multiband slice acquisition with acceleration factor 4, FOV 210 mm, flip angle 75°, A/ P phase encoding direction). The first seven volumes of each run were automatically discarded. This sequence was chosen because of its balance between a short TR and relatively high spatial resolution, which was required to disentangle cue and outcome-related neural activity. Pilots using different sequences yielded that this sequence performed best in reducing signal loss in striatum.

Furthermore, after task completion, we removed the EEG cap and collected a high-resolution anatomical image using a T1-weighted MP-RAGE sequence (192 sagittal slices per slab, GRAPPA acceleration factor = 2, TI = 1100 ms, TR = 2300 ms, TE = 3.03 ms, FOV 256 mm, voxel size 1.0 mm isotropic, flip angle 8°) which was used to aid image registration, and a gradient fieldmap (GRE; TR = 614 ms, TE1 = 4.92 ms, voxel size 2.4 mm isotropic, flip angle 60°) for distortion correction. For one participant, no fieldmap was collected due to time constraints. At the end of each session, an additional DTI data collection took place; results will be reported elsewhere.

3.5.6 fMRI preprocessing

All fMRI pre-processing was performed in FSL 6.0.0. After cleaning images from non-brain tissue (brain-extraction with BET), we performed motion correction (MC-FLIRT), spatial smoothing (FWHM 3 mm), and used fieldmaps for B0 unwarping and distortion correction in orbitofrontal areas. We used ICA-AROMA (Pruim et al. 2015) to automatically detect and reject independent components associated with head motion. Finally, images were high-pass filtered at 100 s and pre-whitened. After the first-level GLM analyses, we computed and applied co-registration of EPI images to high-resolution images (linearly with FLIRT using boundary-based registration) and to MNI152 2mm isotropic standard space (non-linearly with FNIRT using 12 DOF and 10 mm warp resolution).

3.5.7 ROI selection

For fMRI-informed EEG analyses, we first created a functional mask as the conjunction of the PE_{STD} and PE_{DIF} contrasts by thresholding both z-maps at $z > 3.1$, binarizing, and multiplying them (see S3.7). After visual inspection of the respective clusters, we created seven anatomical masks based on the probabilistic Harvard-Oxford Atlas (thresholded at 10%): striatum and ACC (see above), vmPFC (combined frontal pole, frontal medial cortex, and paracingulate gyrus), motor cortex (combined precentral and postcentral gyrus), PCC (Cingulate Gyrus, posterior division), ITG (Inferior Temporal Gyrus, posterior division, and Inferior Temporal Gyrus, temporooccipital part) and primary visual cortex (Lingual Gyrus, Occipital Fusiform Gyrus, Occipital Pole). We then multiplied this functional mask with each of the seven anatomical masks, returning seven masks focused on the respective significant clusters, which were then used for signal extraction. For the dACC mask, we manually excluded voxels in pgACC belonging to a distinct cluster. Masks were back-transformed to each participant's native space.

For bar plots in Fig. 3.3A, we multiplied the anatomical masks of vmPFC and striatum specified above with the binarized outcome valence contrast.

3.5.8 fMRI analyses

For each participant, data were modelled using two event-related GLMs. First, we performed a model-based GLM in which we used trial-by-trial estimates of biased PEs as regressors. Second, we used another model-free GLM in which we modeled all possible action x outcome combinations via outcome-locked categorical regressors while at the same time modeling response-locked left- and right-hand response regressors. This model free GLM also contained the outcome valence contrast reported as an initial manipulation check.

In the model-based GLM, we used two model-based regressors that reflected the trial-by-trial prediction error (PE) update term. The update term was computed by multiplying the prediction-error with the condition-specific learning rate. As described above, in the winning model M5, the learning bias term $\varkappa$ leads to altered learning from "congruent" action-outcome pairs, with faster learning of Go actions for rewards, but slower unlearning of NoGo actions to avoid punishments. To compute trial-by-trial updates, we extracted the group-level parameters of the best fitting computational model M5 (asymmetric pathways model) and used those parameters to compute the prediction error on every trial for every participant. Using the same parameter for each participant is warranted when testing for the same qualitative learning pattern across participants (Wilson and Niv 2015). Given that both standard (base model M1) and biased (winning model M5) PEs were highly correlated (mean correlation of 0.921 across participants, range 0.884–0.952), it appeared difficult to distinguish standard learning from biased learning. As a remedy, we decomposed the biased PE into the standard PE plus a difference term as $PE_{BIAS} = PE_{STD} + PE_{DIF}$ (Wittmann et al. 2008; Daw et al. 2011). Any region displaying truly biased learning should significantly encode *both* the standard PE term and the difference term. The standard PE and difference term were much less correlated (mean correlation of -0.020, range -0.326–0.237). To control for cue-related activation, we furthermore added four regressors spanned by crossing cue valence and performed action (Go response to Win cue, Go response to Avoid cue, NoGo response to Win cue, NoGo response to Avoid cue).

The model-free GLM included a separate regressor for each of the eight conditions obtained when crossing performed action (Go/ NoGo) and obtained outcome (reward/ no reward/ no punishment/ punishment). We fitted four contrasts: 1) one contrast comparing conditions with positive (reward/ no punishment) and negative (no reward/ punishment) outcomes, used as a quality check to identify regions that encoded outcome valence; 2) one contrast comparing Go vs. NoGo responses at the time of the outcome; 3) one contrast summing of left- and right-hand responses, reflecting Go vs. NoGo responses at the time of the response; and 4) one contrast subtracting right- from left-handed responses, reflecting lateralized motor activation. As this GLM resulted in empty regressors for several participants when fitted on a block level, making it impossible to use the data of the respective blocks on a higher level, we instead concatenated blocks and performed a single GLM per participant. We therefore registered the data from all blocks to the middle image of the first block (default reference volume in FSL) using MCFLIRT. The first and last 20 seconds of each block did not feature any task-related events, such that carry-over effects of task events in the design matrix from one block to another were not possible.

In both GLMs, we added four regressors of no interest: one for the motor response (left = +1, right = -1, NoGo = 0), one for error trials, one for outcome onset, and one for trials with invalid motor response (and no outcome respectively). We also added nine or more nuisance regressors: the six realignment parameters from motion correction, mean cerebrospinal fluid (CSF) signal, mean out-of-brain (OBO) signal, and a separate spike regressor for each volume with a relative displacement of more than 2 mm (occurred in 10 participants; in those participants: M = 7.40, range 1–29). For the model-free GLM, nuisance regressors were added separately for each block as well as an overall intercept per block. We convolved task regressors with double-gamma haemodynamic response function (HRF) and high-pass filtered the design matrix at 100 s.

First-level contrasts were fit in native space. Afterwards, co-registration and reslicing was applied to participants' contrast maps, which were then combined on a (participant and) group level using FSL's mixed effects models tool FLAME with a cluster-forming threshold of $z > 3.1$ and cluster-level error control at $\alpha < .05$ (i.e., two one-sided tests with $\alpha < .025$).

3.5.9 EEG data acquisition

We recorded EEG data with 64 channels (BrainCap-MR-3-0 64Ch-Standard; Easycap GmbH; Herrsching, Germany; international 10-20 layout, reference electrode at FCz) plus channels for electrocardiogram, heart rate, and respiration (used for MR artifact correction) at a sampling rate of 1000 Hz. We placed MRI-compatible EEG amplifiers (BrainAmp MR plus; Brain Products GmbH, Gilching, Germany) behind the MR scanner and attached cables to the participants once they were located in final position in the scanner. Furthermore, we fixated cables using sand-filled pillows to reduce artifacts induced through cable movement in the magnetic field. During functional scans, the MR helium pump was switched off to reduce EEG artifacts. After the scanning, we recorded the exact EEG electrode locations on participants' heads relative to three fiducial points using a Polhemus FASTRAK device. For four participants, no such data were available due to time constraints/ technical errors, in which case we used the average electrode locations of the remaining 32 participants.

3.5.10 EEG pre-processing

First, raw EEG data were cleaned from MR scanner and cardioballistic artifacts using BrainVisionAnalyzer (Allen et al. 2000). The rest of the pre-processing was performed in Fieldtrip (Oostenveld et al. 2011). After rejecting channels with high residual MR noise (mean 4.8 channels per participant, range 1–13), we epoched trials into time windows of -1,400–2,000 ms relative to the onset of outcomes. Timing of this epochs was determined by the minimal inter-stimulus interval beforehand until the minimal inter-trial interval afterwards. Data was re-referenced to the grand average, which allowed us to recover the reference as channel FCz, and then band-pass filtered using a two-pass 4th order Butterworth IIR filter (Fieldtrip default) in the range of 0.5–35 Hz. These filter settings allowed us to distinguish the delta, theta, alpha, and beta band, while filtering out residual high-frequency MR noise. This low-pass filter cut-off was different from a previous analysis of this data in which we set it at 15 Hz (Algermissen et al. 2022) because, in this analysis, we had a hypothesis on outcome valence encoding in the beta range. We then applied linear baseline correction based on the 200 ms prior to cue onset and used ICA to detect and reject independent components related to eye-blinks, saccades, head motion, and residual MR artifacts (mean number of rejected components per participant: 32.694, range 24–45). Afterwards, we

manually rejected trials with residual motion (for all 36 participants: *M* = 117.722, range 11–499). Based on trial rejection, four participants for which more than 211 (33%) of trials were rejected were excluded from any further analyses (rejected trials after excluding those participants: *M* = 81.875, range 11–194). Finally, we computed a Laplacian filter with the spherical spline method to remove global noise (using the exact electrode positions recorded with Polhemus FASTRAK), which we also used to interpolate previously rejected channels. This filter attenuates more global signals (e.g., signal from deep sources or global noise) and noise (heart-beat and muscle artifacts) while accentuating more local effects (e.g., superficial sources).

3.5.11 EEG time-frequency decomposition

We decomposed the trial-by-trial EEG time series into their time-frequency representations using 33 Hanning tapers between 1 and 33 Hz in steps of 1 Hz, every 25 ms from -1000 until 1,300 ms relative to outcome onset. We first zero-padded trials to a length of 8 sec. and then performed time-frequency decomposition in steps of 1 Hz by multiplying the Fourier transform of the trial with the Fourier transform of a Hanning taper of 400 ms width, centered around the time point of interest. This procedure results in an effective resolution of 2.5 Hz (Rayleigh frequency), interpolated in 1 Hz steps, which was more robust to the choice of exact frequency bins. To exclude the possibility of slow drifts in power over the time course of the experiment, we performed baseline correction across participants and trials by fitting a linear model for each channel/ frequency combination with trial number as predictor and the average power 250–50 ms before outcome onset as outcome, and subtracting the power predicted by this model from the data. This procedure is able to remove slow linear drifts in power over time from the data. In absence of such drifts, it is equivalent to correcting all trials by the grand mean across trials per frequency in the selected baseline time window. Afterwards, we averaged power over trials within each condition spanned by performed action (Go/ NoGo) and outcome (reward/ no reward/ no punishment/ punishment). We finally converted the average time-frequency data per condition to decibel to ensure that data across frequencies, time points, electrodes, and participants were on same scale.

3.5.12 EEG analyses

All analyses were performed on the average signal of a-priori selected channels Fz, FCz, and Cz based on (Swart et al. 2018; Algermissen et al. 2022). We again performed model-free and model-based analyses. For the model-free analyses, we sorted trials based on the performed action (Go/ NoGo) and obtained outcome (reward/ no reward/ no punishment/ punishment) and computed the mean TF power across trials for each of the resultant eight conditions for each participant. We tested whether theta power (average power 4–8 Hz) and beta power (average power 13–30 Hz) encoded outcome valence by contrasting positive (reward/ no punishment) and negative (no reward/ punishment) conditions (irrespective of the performed action). We also tested for differences between Go and NoGo responses in the lower alpha band (6–10 Hz). For all contrasts, we employed two-sided cluster-based permutation tests in a window from 0–1,000 ms relative to outcome onset. For beta power, results were driven by a cluster that was at the edge of 1,000 ms; to more accurately report the time span during which this cluster exceeded the threshold, we extended the time window to 1,300 ms in this particular analysis. Such tests are able to reject the null hypothesis of exchangeability of two experimental conditions, but they are not

suited to precisely locate clusters in time-frequency space. Hence, interpretations were mostly based on the visual inspection of plots of the signal time courses.

For model-based analyses, similar to fMRI analyses, we used the group-level parameters from the best fitting computational model M5 to compute the trial-by-trial biased PE term and decomposed it into the standard PE term and the difference to the biased PE term. We used both terms as predictors in a multiple linear regression for each channel-time-frequency bin for each participant, and then performed one-sample cluster-based permutation-tests across the resultant *b*-maps of all participants (Hunt et al. 2013). For further details on this procedure, see fMRI-inspired EEG analyses.

3.5.13 fMRI-informed EEG analyses

The BOLD signal is sluggish. It is thus hard to determine when different brain regions become active. In contrast, EEG provides much higher temporal resolution. A fruitful approach can be to identify distinct EEG correlates of the BOLD signal in different regions, allowing to test hypotheses about the temporal order in which regions might become active and modulated EEG power (Hauser et al. 2015; Algermissen et al. 2022). Furthermore, by using the BOLD signal from different regions in a multiple linear regression, one can control for variance shared among regions (e.g., changes in global signal; variance due to task regressors) and test which region is the best unique predictor of a certain EEG signal. In such an analysis, any correlation between EEG and BOLD signal from a certain region reflects an association above and beyond those induced by task conditions.

We used the trial-by-trial BOLD signal in selected regions in a multiple linear regression to predict EEG signal over the scalp (Hauser et al. 2015; Algermissen et al. 2022) (building on existing code from https://github.com/tuhauser/TAfT; see S3.13 for a graphical illustration). As a first step, we extracted the volume-by-volume signal (first eigenvariate) from each of the seven regions identified to encode biased PEs (conjunction of PE_{STD} and PE_{DIF}: striatum, dACC, pgACC, left motor cortex, PCC, left ITG, and primary visual cortex). We applied a highpass-filter at 128 s and regressed out nuisance regressors (6 realignment parameters, CSF, OOB, single volumes with strong motion, same as in the fMRI GLM). We then upsampled the signal by a factor 10, epoched it into trials of 8 s duration, and fitted a separate HRF (based on the SPM template) to each trial (58 upsampled data points), resulting in trial-by-trial regression weights reflecting the respective BOLD response. We then combined the regression weights of all trials and regions of a certain participant into a design matrix with trials as rows and the seven ROIs as columns, which we then used to predict power at each time-frequency-channel bin. As further control variables, we added the behavioral PE_{STD} and PE_{DIF} regressors to the design matrix. All results were identical with and without including PEs into the model, suggesting that EEG-fMRI correlations did not merely arise from both modalities encoded PEs as a "common cause" that induced correlations. Instead, these correlations reflected the incremental variance explained in EEG power that was afforded by the BOLD signal even beyond the PEs. All predictors and outcomes were demeaned such that the intercept became zero. Such a multiple linear regression was performed for each participant, resulting in a time-frequency-channel-ROI *b*-map reflecting the association between trial-by-trial BOLD signal and TF power at each time-frequency-channel bin. *B*-maps were Fisher-z transformed, which makes the sampling distribution of correlation coefficients approximately normal and allows for combining them across participants. Finally, we tested for fMRI-EEG

associations with a cluster-based one-sample permutation *t*-test (Hunt et al. 2013) on the mean regression weights over channels Fz, FCz, and Cz across participants in the range of 0–1000 ms, 1–33 Hz. We first obtained a null distribution of maximal cluster mass statistics from 10000 permutations. For each permutation, we flipped the sign of the *b*-map of a random subset of participants, computed a separate *t*-test at each time-frequency bin (bins of 25 ms, 1 Hz) across participants (results in *t*-map), thresholded these maps at $|t| > 2$, and finally computed the maximal cluster mask statistic (sum of all *t*-values) for any cluster (adjacent voxels above threshold). Afterwards, we computed the same *t*-map for the real data, identified the cluster with the biggest cluster-mass statistic, and computed the corresponding *p*-value as number of permutations in the null distribution that were larger than the maximal cluster mass statistic in the real data.

3.5.14 EEG-informed fMRI analyses

For the EEG-informed fMRI analyses, we fit three additional GLMs for which we entered the trial-by-trial theta/ delta power (1–8 Hz), beta power (13–30 Hz), and lower alpha band power (6–10 Hz) as parametric regressors on top of the task regressors of the model-free GLM. These measures were created by using the 3-D (time-frequency-channel) *t*-map obtained when contrasting positive vs. negative outcomes (theta/ delta and beta; Fig. 3.4 A, B) and Go vs. NoGo conditions (lower alpha band) as a linear filter (Fig. 3.4; see S3.13 for a graphical illustration of this approach). Note that these signals were selected based on the EEG-only results and not informed by the fMRI-informed EEG analyses. We enforced strict frequency cut-offs. For lower alpha band and beta, we used midfrontal channels (Fz/ FCz/ Cz). For theta/ delta power, given the topography that reached far beyond midfrontal channels and over the entire frontal scalp, we used a much wider ROI (AF3/ AF4/ AF7/ AF8/ F1/ F2/ F3/ F4/ F5/ F6/ F7/ F8/ FC1/ FC2/ FC3/ FC4/ FC5/ FC6/ FCz/ Fp1/ Fp2/ Fpz/ Fz). We extracted those maps and retained all voxels with $t > 2$. These masks were applied to the trial-by-trial time-frequency data to create weighted summary measures of the average power in the identified clusters in each trial. For trials for which EEG data was rejected, we imputed the participant mean value of the respective action (Go/ NoGo) x outcome (reward/ no reward/ no punishment/ punishment) condition. Note that this approach accentuates differences between conditions, which were already captured by the task regressors in the GLM, but decreases trial-by-trial variability within each condition, which is of interest in this analysis. This imputation approach is thus conservative. While trial-by-trial beta and theta power were largely uncorrelated, mean $r = 0.104$, range -0.118–0.283 across participants, and so were beta and alpha, mean $r = 0.097$, range -0.162–0.284 across participants, theta and alpha power moderately correlate, mean $r = 0.412$, range 0.121–0.836 across participants, warranting the use of a separate channel ROI for theta and using separate GLMs for each frequency band.

3.5.15 Analyses of behavior as a function of BOLD signal and EEG power

We used mixed-effects logistic regression to analyze "stay behavior", i.e., whether participants repeated an action on the next encounter of the same cue, as a function of BOLD signal and EEG power in selected regions. For analyses featuring BOLD signal, we used the trial-by-trial HRF amplitude also used for fMRI-informed EEG analyses. For analyses featuring EEG, we used the trial-by-trial EEG power also used in the EEG-informed fMRI analyses.

3.6 SUPPLEMENTARY MATERIALS FOR CHAPTER 3

3.6.1 S3.1: Behavioral, fMRI, and EEG analyses with only the 29 participants included in EEG-fMRI analyses

We repeated the behavioral, fMRI, and EEG analyses reported in the main text while excluding the seven participants that were also not included in the fMRI-inspired EEG analyses in the main text: (a) two participants due to fMRI co-registration failure, which were also not included in the fMRI-only analyses; (b) four further participants who exhibited excessive residual noise in their EEG data (> 33% rejected trials) and were thus also not included in the EEG-only analyses, and finally (c) one more participant who (together with four other participants already excluded) exhibited regression weights for every regressor about ten times larger than for other participants.

Participants in this subgroup learned the task, reflected in a significant main effect of required action on responses, $b = 0.896$, $SE = 0.129$, $\chi^2(1) = 28.398$, $p < .001$, and exhibited motivational biases, reflected in a significant main effect of cue valence on responses, $b = 0.439$, $SE = 0.084$, $\chi^2(1) = 19.308$, $p < .001$. The interaction between required action and cue valence was not significant, $b = 0.025$, $SE = 0.085$, $\chi^2(1) = 0.111$, $p = .739$.

Participants in this subgroup also showed biased learning: They were more likely to repeat an action after a positive outcome (main effect of outcome valence: $b = .0553$, $SE = 0.059$, $\chi^2(1) = 40.920$, $p < .001$. After salient outcomes, they adjusted their responses more strongly after feedback on Go than on NoGo responses, in line with our model of biased learning and as reflected in a significant three-way interaction between action, salience, and valence, $b = 0.266$, $SE = 0.055$, $\chi^2(1) = 16.862$, $p < .001$. When only analyzing trials with salient outcomes, outcome valence was more likely to affect response repetition following Go relative to NoGo responses, $b = 0.324$, $SE = 0.079$, $\chi^2(1) = 13.266$, $p < .001$, with a stronger effect of outcome valence after Go responses, $b = 1.342$, $SE = 0.120$, $\chi^2(1) = 49.003$, $p = .001$, than NoGo responses, $b = 0.693$, $SE = 0.129$, $\chi^2(1) = 18.988$, $p < .001$.

In this subgroup of participants, Bayesian model selection clearly favored the full asymmetric pathways models featuring response and learning biases (M5, model frequency: 81.81%, protected exceedance probability: 100%). In sum, behavioral results were qualitatively identical when analyzing only this subgroup of only 29 participants.

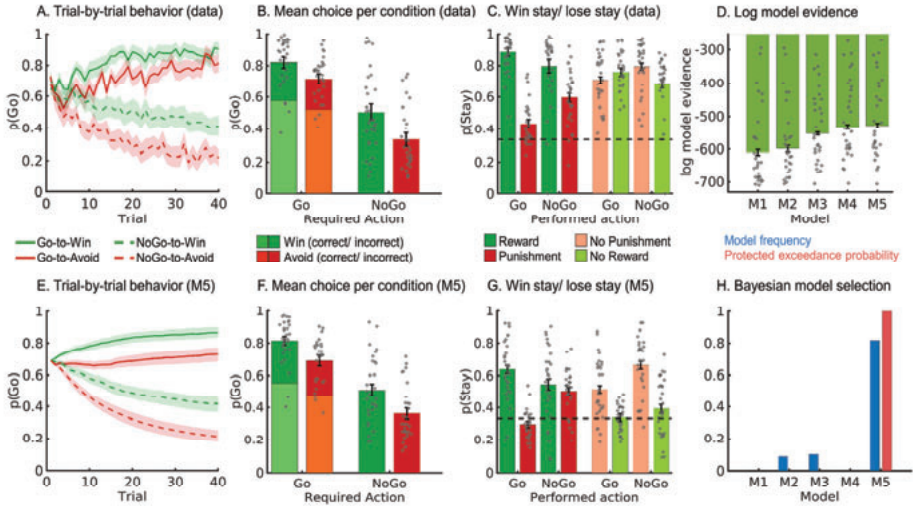


Figure 3.7. S3.1A. Behavioral performance in the subgroup of 29 participants included in the fMRI-inspired EEG analyses. **A.** Trial-by-trial proportion of Go responses (±SEM across participants) for Go cues (solid lines) and NoGo cues (dashed lines). The motivational bias was already present from very early trials onwards, as participants made more Go responses to Win than Avoid cues (i.e., green lines are above red lines). Additionally, participants clearly learn whether to make a Go response or not (proportion of Go responses increases for Go cues and decreases for NoGo cues). **B.** Mean (±SEM across participants) proportion Go responses per cue condition (points are individual participants' means). **C.** Probability of repeating a response ("stay") on the next encounter of the same cue as a function of action and outcome. Learning was reflected in higher probability of staying after positive outcomes than after negative outcomes (main effect of outcome valence). Biased learning was evident in learning from salient outcomes, where this valence effect was stronger after Go responses than NoGo responses. Dashed line indicates chance level choice (p_{Stay} = 0.33). **D.** Log-model evidence favors the asymmetric pathways model (M5 over simpler models (M1-M4). **E-G.** Trial-by-trial proportion of Go responses, mean proportion Go responses, and probability of staying based on one-step-ahead predictions using parameters (hierarchical Bayesian inference) of the winning model (asymmetric pathways model, M5). **H.** Model frequency and protected exceedance probability indicate best fit for model M5 (asymmetric pathways model), in line with log model evidence.

Regarding fMRI findings, we first repeated the model-free GLM just contrasting positive and negative outcomes. BOLD signal was higher for positive than negative outcomes in five clusters, namely in vmPFC, striatum, amygdala, and hippocampus ($\tilde{\chi}_{max}$ = 5.65, p = 2.24e-25, 6110 voxels, MNI coordinates xyz = [6 30 -12]), left superior lateral occipital cortex ($\tilde{\chi}_{max}$ = 4.40, p = .00144, 367 voxels, xyz = [-46 -68 46]), right occipital pole ($\tilde{\chi}_{max}$ = 4.45, p = .00154, 363 voxels, xyz = [12 -92 -12]), posterior cingulate cortex ($\tilde{\chi}_{max}$ = 4.36, p = .00181, 353 voxels, xyz = [-2 -48 28]), and left middle temporal gyrus ($\tilde{\chi}_{max}$ = 4.63, p = .00548, 289 voxels, xyz = [-60 -10 -16]). The clusters in left slOCC, PCC, and left MTG emerged anew compared to the original analysis comprising 34 participants. Also, compared to the original analysis, clusters in left orbitofrontal cortex and left superior frontal gyrus were merged with the cluster in vmPFC. In sum, all clusters from the original analysis were found back, plus some additional clusters.

There was also one cluster in right orbitofrontal cortex ($\tilde{\chi}_{max}$ = 4.37, p = .0209, 217 voxels, xyz = [30 62 -2]) in which BOLD signal was higher for negative than positive outcomes. Compared to the original analysis comprising 34 participants, clusters in precuneous and right superior frontal gyrus were not significant.

In the model-based GLM featuring regressors for standard PEs and the difference term towards biased PEs, BOLD signal correlated with standard PEs in ten clusters, namely in vmPFC, striatum, bilateral amygdala and hippocampus (z_{max} = 6.04, p = .4.78e-44, 8848 voxels, xyz = [12 14 -6]), left superior frontal gyrus (z_{max} = 5.58, p = 3.5e-10, 1043 voxels, xyz = [-18 34 52]), left occipital pole and lingual gyrus (z_{max} = 6.23, p = 7.18e-10, 998 voxels, xyz = [10 -92 -10]), posterior cingulate cortex (z_{max} = 5.12, p = 8.57e-10, 987 voxels, xyz = [4 -36 48]), left inferior temporal gyrus (z_{max} = 5.03, p = 7.07e-09, 859 voxels, xyz = [-52 -46 -10]), right anterior middle temporal gyrus (z_{max} = 5.32, p = .000292, 314 voxels, xyz = [62 -4 -16]), right cerebellum (z_{max} = 5.32, p = .002228, 231 voxels, xyz = [44 -72 -40]), left superior lateral occipital cortex (z_{max} = 4.69, p = .00322, 218 voxels, xyz = [-46 -74 -38]), right caudate (z_{max} = 4.33, p = .00538, 199 voxels, xyz = [20 12 22]), and right middle temporal gyrus (z_{max} = 4.09, p = .0129, 189 voxels, xyz = [54 -38 -12]). The clusters in left superior lateral occipital cortex, right caudate, and right posterior middle temporal gyrus emerged anew by splitting from larger clusters visible in the original analysis based on 34 participants. Vice versa, the cluster in left middle temporal gyrus reported for the original analysis was merged with a bigger cluster in the analysis of only 29 participants. The clusters in postcentral gyrus and ACC observed in the original analysis based on 34 participants were not significant anymore; however, they were still visible at a level of $z > 3.1$ uncorrected.

BOLD signal correlated significantly negatively with standard PEs in a single cluster in right superior frontal gyrus (z_{max} = 5.04, p = .00771, 186 voxels, xyz = [6 26 64]), similar to the respective cluster reported in the original analysis. In contrast, the clusters in right occipital pole, intracalcarine cortex, and left inferior lateral occipital cortex were not significant any more, though visible at a level of $z > 3.1$ uncorrected.

BOLD signal in six clusters correlated significantly positively with the difference term towards biased PEs, namely in large parts of cortex and subcortex including striatum (z_{max} = 6-54, p = 0, 29428 voxels, xyz = [34 -84 20]), dorsomedial prefrontal cortex (z_{max} = 5.94, p = 2.69e-40, 7001 voxels, xyz = [6 22 34]), right insula (z_{max} = 5.76, p = 7.84e-27, 3847 voxels, xyz = [34 20 -8]), thalamus and brainstem (z_{max} = 5.10, p = 4.06e-18, 2169 voxels, xyz = [4 -30 0]), left caudate (z_{max} = 4.71, p = .000188, 305 voxels, xyz = [-12 8 6]) and another cluster in brainstem (z_{max} = 4.05, p = .0151, 160 voxels, xyz = [4 -30 -30]). Clusters in dmPFC, right insula, and left caudate split from larger clusters reported in the original analysis. Vice versa, the cluster in left insula reported in the original analysis merged with the largest cluster. The clusters in right middle temporal gyrus and right insula were missing in the analysis of only 29 participants, but visible at a level of $z > 3.1$ uncorrected.

BOLD signal in three clusters correlated significantly negatively with the difference term towards biased PEs, namely in vmPFC (z_{max} = 4.23, p = .0051, 185 voxels, xyz = [-12 48 -6]), left hippocampus (z_{max} = 4.58, p = .00857, 168 voxels, xyz = [-26 -14 -22]), and left medial temporal gyrus (z_{max} = 4.30, p = .0172, 146 voxels, xyz = [-62 -4 -16]). Compared to the original analysis, the cluster in vmPFC emerged anew.

When computing the conjunction between both (positive) contrasts, BOLD signal encoded both the standard and the difference in four clusters, namely in vmPFC, bilateral striatum, bilateral ITG, and V1. Clusters in ACC, left motor cortex, and PCC were not significant any more (because they were $z > 3.1$, but not significant after cluster correction in the standard PE contrast).

However, new (though rather small) clusters of biased PE encoding emerged in right insula, left amygdala, and left OFC. In sum, results when analyzing only this subgroup of only 29 participants were largely similar to results based on the full sample; however, clusters of biased PE encoding in left motor cortex, ACC, and PCC were small and thus did not survive cluster correction in this subgroup.

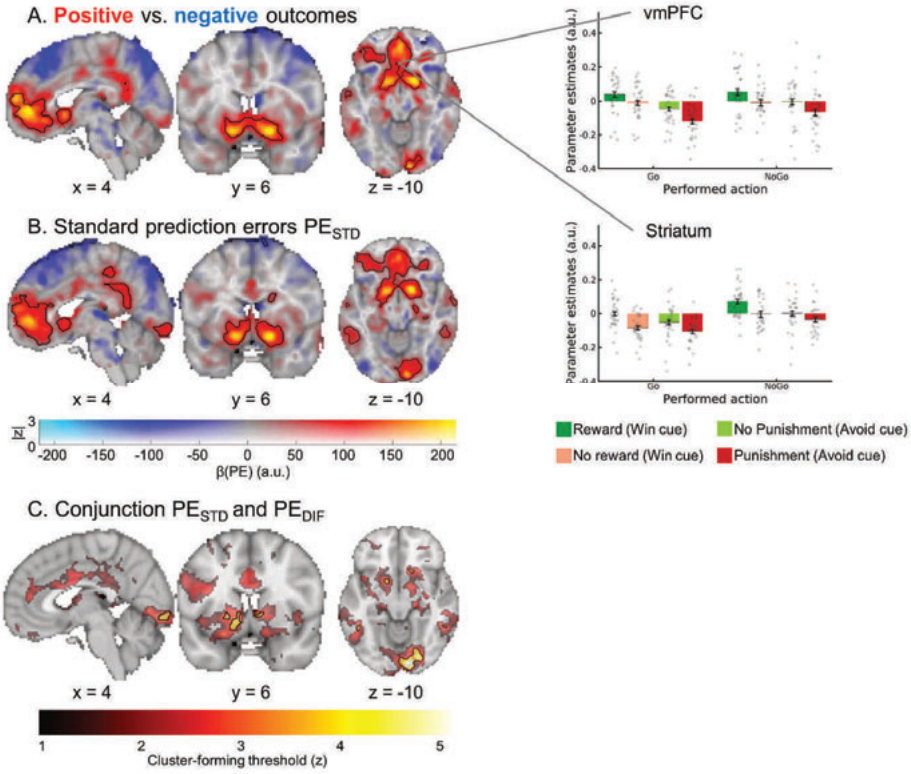


Figure 3.8. S3.1B. BOLD signal reflecting biased outcome processing in the subgroup of 29 participants included in the fMRI-inspired EEG analyses.

A. BOLD signal was higher for positive outcomes (rewards, no punishments) compared with negative outcomes (no rewards, punishments) in a range of regions including bilateral ventral striatum and vmPFC. BOLD effects displayed using a dual-coding data visualization approach with color indicating the parameter estimates and opacity the associated z-statistics. Significant clusters are surrounded by black edges. Bar plots show parameter estimates per action x outcome condition (±SEM across participants) **B.** When using the trial-by-trial PEs participants experienced as model-based regressors in our GLM, positive PE correlations occurred in several regions including importantly the ventral striatum, vmPFC, dACC, and PCC. **C.** Left panel: Regions encoding both the standard PE term and the difference term to biased PEs (conjunction) at different cluster-forming thresholds (color). Clusters significant at a threshold of $z > 3.1$ are surrounded by black edges. In bilateral striatum, pgACC, bilateral ITG, and primary visual cortex, BOLD was significantly better explained by biased learning than by standard learning. Clusters in dACC, left motor cortex, and PCC were not significant any more.

Regarding EEG findings in this subgroup, both midfrontal theta and beta power reflected outcome valence: Theta power was higher for negative than positive outcomes (driven by a cluster around 225–500 ms, p = .002), while beta power was higher for positive than negative outcomes (driven by a cluster around 325–1000 ms, p = .002). When using PE terms as regressor for midfrontal EEG power while controlling for PE valence, delta power did not encode PE_{STD} positively, though not significant (p = .056), and also the positive encoding of PE_{DIF} was non-significant (p = .053). The positive correlation of beta power with PE_{STD} was not significant anymore (p = .059), while the negative correlation with PE_{DIF} remained (p = .001, 450–950 ms). When adding PE_{STD} and PE_{DIF} together to achieve PE_{BIAS} , theta/delta power indeed significantly encoded PE_{BIAS}, first positively (p = .032, 224–475 ms) and then negatively (p = .019, 600 – 1,000 ms; around 8 Hz and thus rather in the alpha band). Also, beta power was significantly negatively correlated with PE_{BIAS} (p = .008, 450 – 975 ms).

In sum, all findings reported in the main text also held when analyzing only this subgroup of only 29 participants. In addition, also late beta power and theta/alpha power appeared to negatively encode the PE_{BIAS} term.

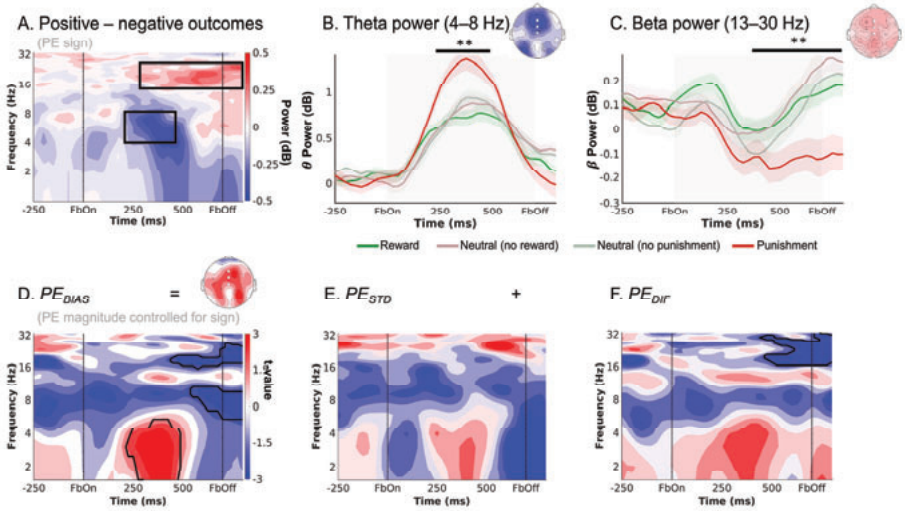


Figure 3.9. S3.1C. EEG time-frequency power midfrontal electrodes (Fz/ FCz/ Cz) reflecting outcomes processing in the subgroup of 29 participants included in the fMRI-inspired EEG analyses.

A. Time-frequency plot (logarithmic y-axis) displaying high theta (4–8 Hz) power for negative outcomes and higher beta power (16–32 Hz) for positive outcomes. **B.** Theta power transiently increases for any outcome, but more so for negative outcomes (especially punishments) around 225–475 ms after feedback onset. **C.** Beta was higher for positive than negative outcomes (especially punishments) over a long time period around 300–1,250 ms after feedback onset. **D-F.** Correlations between midfrontal EEG power and trial-by-trial PEs. Solid black lines indicate clusters above threshold. Biased PEs were significantly positively correlated with midfrontal theta power, but also negatively correlated with later alpha and beta power (**D**). The correlations of theta with the standard PEs (**E**) and the difference term to biased PEs (**F**) were also positive, though not significant. Beta power only encoded the difference term to biased PEs (F). ** p < 0.01.** p < 0.01.

Regarding fMRI correlates of the past action, similar to the original analysis comprising 34 participants, there were no clusters with higher BOLD after Go than NoGo actions at the time of outcomes, but vice versa, large parts of cortex and subcortex showed higher BOLD after NoGo than Go actions, highly similar to the original analysis (z_{max} = 7.65, p = 0, 124629 voxels, xyz = [-58 18 22]).

Furthermore, there were four clusters with higher BOLD for Go than NoGo actions at the time of the response, namely one large cluster across lateral prefrontal cortex, anterior cingulate cortex, striatum, thalamus, angular gyrus, cerebellum, left operculum and motor cortex, intracalcarine cortex, and occipital pole (z_{max} = 7.45, p = 0, 61057 voxels, xyz = [32 -4 -4]), one in right middle temporal gyrus (z_{max} = 4.90, p = 8.66e-05, 493 voxels, xyz = [66 -32 -12]), one in left inferior temporal gyrus (z_{max} = 4.43, p = .00294, 293 voxels, xyz = [-60 -44 -18]), and one in precuneous (z_{max} = 2.39, p = .0041, 276 voxels, xyz = [-8 -70 38]). All these regions were also found in the original analysis comprising 34 participants. Vice versa, BOLD signal was higher NoGo than Go actions at the time of the response in two clusters in vmPFC and subcallosal cortex (z_{max} = 4.23, p = .00864, 239 voxels, xyz = [-2 18 -6]) and right anterior temporal gyrus/ temporal pole (z_{max} = .4.14, p = .0193, 201 voxels, xyz = [48 -6 -8]), identical to the original analysis comprising 34 participants.

Finally, there was higher BOLD signal for left hand compared to right hand responses at the time of response in two clusters in right precentral and postcentral gyrus, superior parietal lobule, and operculum (z_{max} = 6.66, p = 0, 11597 voxels, xyz = [46 -24 64]) and left cerebellum (z_{max} = 6.76, p = 1.05e-18, 2672 voxels, xyz = [-18 -54 -16]), identical to the original analysis comprising 34 participants. Vice versa, there was higher BOLD signal for right hand than left hand responses at the time of responses in five clusters in left precentral and postcentral gyrus, superior parietal lobule, operculum, and thalamus (z_{max} = 6.4, p = 0, 12372 voxels, xyz = [-36 -20 66]), right cerebellum (z_{max} = 7.17, p = 3.41e-21, 3206 voxels, xyz = [20 -54 -20]), right superior lateral occipital cortex (z_{max} = 4.84, p = 2.28e-09, 988 voxels, xyz = [48 -86 -4]), right angular gyrus (z_{max} = 4.11, p = 7.68e-05, 396 voxels, xyz = [66 -50 28]), and left superior lateral occipital cortex (z_{max} = 5.03, p = .019, 164 voxels, xyz = [-18 -82 48]). The clusters in right occipital pole/ intracalcarine cortex and in right posterior cerebellum observed in the original analysis comprising 34 participants were not observed in this analysis. In sum, all major findings also held when analyzing only this subgroup of only 29 participants.

Regarding EEG time-frequency correlates of the past action, when testing for differences in broadband after outcome onset, there was no significant difference after Go and NoGo responses, p = .283. When restricting analyses to the low alpha range, the permutation test was marginally significant, p = .056, driven by a cluster around 0–100 ms around 7–10 Hz). When repeating the permutation test for the broadband signal including the last second before outcome onset, there was a significant difference after Go and NoGo responses, driven by clusters in the beta band. p = 0.002, -1000 – -275 ms, 13–32 Hz, and in the theta/ low alpha band, p = 0.020, -1000 – -525 ms, 4–10 Hz.

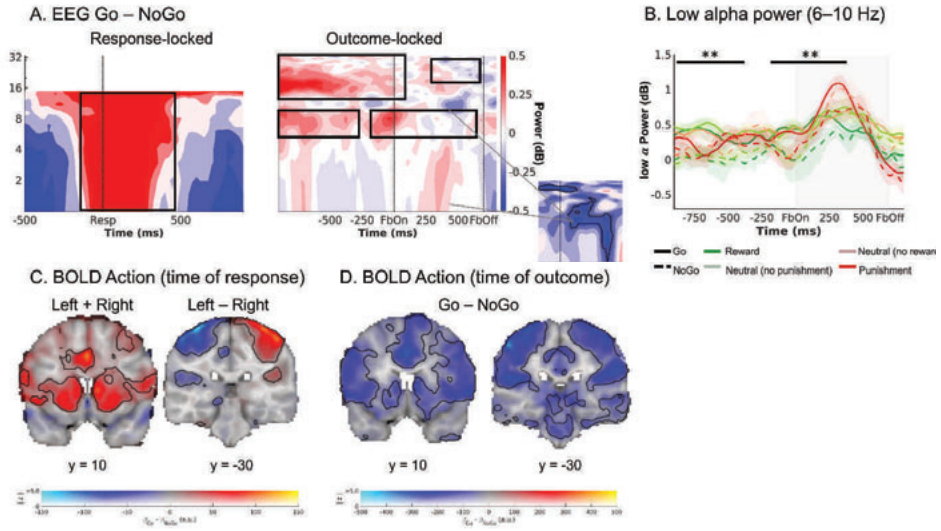


Figure 3.10. S3.1D. Exploratory follow-up analyses on dACC BOLD signal and midfrontal low-alpha power in the subgroup of 29 participants included in the fMRI-inspired EEG analyses.
A. Midfrontal time-frequency response-locked (left panel) and outcome-locked (right panel). Before and shortly after outcome onset, power in the lower alpha band was higher on trials with Go actions than on trials with NoGo actions. The shape of this difference resembles the shape of dACC BOLD-EEG TF correlations (small plot; note that this plot depicts BOLD-EEG correlations, which were negative). Note that differences between Go and NoGo trials occurred already before outcome onset in the alpha and beta range, reminiscent of delay activity; but were not fully sustained since the actual response. **B.** Midfrontal power in the lower alpha band per action x outcome condition. Lower alpha band power was consistently higher on trials with Go actions than on trials with NoGo actions, starting already before outcome onset. **C.** BOLD signal differences between Go and NoGo actions (left panel) and left vs. right hand responses (right panel) at the time or responses. Response-locked dACC BOLD was significantly higher for Go than NoGo actions. **D.** BOLD signal differences between Go and NoGo actions at the time of outcomes. Outcome-locked dACC BOLD signal (and BOLD signal in other parts of cortex) was significantly lower on trials with Go than on trials with NoGo actions.

When linking trial-by-trial BOLD signal in selected ROIs as well as midfrontal EEG TF power to response repetition on the next trial with the same cue, dACC BOLD signal did not significantly predict the response repetition, b = -0.013, SE = 0.018, $\chi^2(1)$ = 0.524, p = .469, and neither did PCC BOLD signal, b = -0.037, SE = 0.018, $\chi^2(1)$ = 2.079, p = .149. However, participants in this subgroup were significantly more likely to repeat the sample action when striatal BOLD signal was high, b = 0.097, SE = 0.025, $\chi^2(1)$ = 12.043, p < .001, but more likely to switch when vmPFC BOLD was high, b = -0.075, SE = 0.019, $\chi^2(1)$ = 13.170, p < .001.

When linking trial-by-trial midfrontal EEG TF power to response repetition on the next trial with the same cue, participants in this subgroup were more likely to repeat the same response when beta power was high, b = 0.124, SE = 0.036, $\chi^2(1)$ = 3.502, p < .001, or when low alpha power was high, b = 0.135, SE = 0.044, $\chi^2(1)$ = 8.789, p = .003, but more likely to switch to another response when theta power was high, b = -0.090, SE = 0.040, $\chi^2(1)$ = 4.812, p = .028.

3.6.2 S3.2: Stay behavior as a function of action, salience, and valence

Effect	χ^2	Df	*p*-value
Action	0.01	1	.924
Salience	5.15	1	.021
Valence	45.59	1	< .001
Action x Salience	0.12	1	.728
Action x Valence	3.24	1	.067
Salience x Valence	30.95	1	< .001
Action x Valence x Salience	19.73	1	< .001
Salient outcomes only:			
Action	0.01	1	.960
Valence	46.36	1	< .001
Action x Valence	17.80	1	< .001
Neutral outcomes only:			
Action	.102	1	.750
Valence	.830	1	.362
Action x Valence	12.32	1	< .001
Go with salient outcomes only:			
Valence	53.93	1	< .001
NoGo with salient outcomes only:			
Valence	18.23	1	< .001
Go with neutral outcomes only:			
Valence	0.13	1	.050
NoGo with neutral outcomes only:			
Valence	7.21	1	.007

Table S3.2. Full report of model of stay behavior. Mixed-effects logistic regression of stay vs. switch behavior (i.e., repeating vs. changing an action on the next occurrence of the same cue) as a function of performed action (Go vs. NoGo), outcome salience (salient: reward or punishment vs. neutral: no reward or no punishment), and outcome valence (positive: reward or no punishment vs. negative: no reward or punishment). Follow-up analyses were performed on trials with salient vs. neutral outcomes separately, and then separately based on Go vs. NoGo actions and salient vs. neutral outcomes. *P*-values were computed using likelihood ratio tests using the *mixed*-function (option "LRT") from package *afex*.

3.6.3 S3.3: Model parameters and fit indices for models M1-M6

	M1	M2	M3	M4	M5 (Asymmetric pathways)	M6 (Action priming)
Mean log model evidence	-609.30	-597.95	-554.46	-532.40	-528.13	-540.84
Model frequency	0	0.0278	0	0.0488	0.6815	0.2419
Protected exceedance probability	0	0	0	0	.9970	.0030
ϱ	7.75 [0.53 – 38.68]	6.81 [0.48 – 37.74]	6.38 [0.49 – 35.71]	10.05 [1.26 – 40.60]	9.41 [0.98 – 31.22]	6.64 [0.71 – 22.83]
ε_0	0.17 [0.002 – 0.77]	0.20 [0.003 – 0.82]	0.21 [0.003 – 0.85]	0.09 [0.003 – 0.38]	0.08 [0.003 – 0.41]	0.039 [0.003 – 0.11]
b		-0.05 [-1.23 – 0.82]	-0.01 [-1.23 – 1.09]	0.13 [-1.16 – 1.03]	0.14 [-1.18 – 1.10]	0.16 [-1.22 – 1.40]
π			0.77 [-0.78 – 3.73]		0.17 [-1.25 – 2.70]	-1.11 [-3.29 – 1.23]
$\varepsilon_{\text{rewarded Go}}$ ($\varepsilon_0+\varkappa$)				0.749 [0.29 – 0.99]	0.833 [0.43 – 0.99]	
$\varepsilon_{\text{punished NoGo}}$ ($\varepsilon_0-\varkappa$)				0.001 [0.001 – 0.02]	0.003 [0.001 – 0.09]	
$\varepsilon_{\text{salient Go}}$						0.49 [0.05 – 0.90]

Table S3.3. Model parameters for fitted models. Mean [minimum – maximum] of participant-level parameter estimates in model space, fitted with hierarchical Bayesian inference (only the respective model included in the fitting process). Model frequency and protected exceedance probability were based on a model comparison that involves models M1-M6. Note that Fig. 3.2 in the main text does not include M6.

3.6.4 S3.4: Parameter recovery analyses for model M5

We performed parameter recovery analyses to assess the identifiability of the model parameters in the winning "asymmetric pathways" model M5. We simulated 100 new data sets based on the best fitting parameters of each participant, fitted a separate model to each simulated data set (using first Laplace approximation and then hierarchical Bayesian inference), and finally averaged parameters across the 100 fitted models.

Parameter recovery was excellent for the feedback sensitivity ϱ ($r = .90$), the baseline learning rate ε_0 ($r = .98$), the Go bias b ($r > .99$), and the Pavlovian response bias π ($r > .99$), with between-participant differences in ground-truth parameters correlating at high levels (all $r > .90$) with between-participant differences in the recovered parameters. Note that, due to shrinkage to the mean as a consequence of hierarchical Bayesian inference, extreme parameter values tended to be shrunk to the overall group-level mean in the recovered parameters. Correlations for the learning bias parameter $\varkappa$ were considerably lower, though still strongly positive ($r = 0.50$; $r = 0.51$ when removing one outlier participant). Note however that the effect of $\varkappa$ on learning depended on participants' baseline learning rate ε_0. When computing increased learning rates for rewarded Go actions and decreased learning rates for punished NoGo actions—the parameters actually used for learning in the model—these learning rates tended to again be highly correlated with the ground truth parameters ($\varepsilon_{rewarded\ Go}$: $r = 0.96$; $\varepsilon_{punished\ NoGo}$: $r = 0.85$ resp. $r = 0.86$ when removing one outlier participant).

Further parameter recovery analyses on the models explored in S3.6 yielded that the recovery of $\varkappa$ was improved ($r = 0.78$) when adding perseveration parameters (which themselves had recovery performances of r's > 0.99). This observation suggested that models featuring such perseveration parameters might be better suited for quantifying individual differences in the learning bias.

In sum, parameter recovery was excellent for all parameters but the learning bias $\varkappa$. However, when combining the baseline learning rate ε_0 and the learning bias $\varkappa$, recovery was at similarly high levels. Note that, in this study, we did not investigate individual differences in biases. For our model-based fMRI and EEG analyses, we used a single set of parameters (i.e., the group-level parameters from hierarchical Bayesian inference) to compute trial-by-trial PE updates for the regressors. The exact parameter values for such analyses do not matter much (Wilson and Niv 2015) and indeed, using a different set of parameter values, we obtained essentially identical fMRI results (see S3.6).

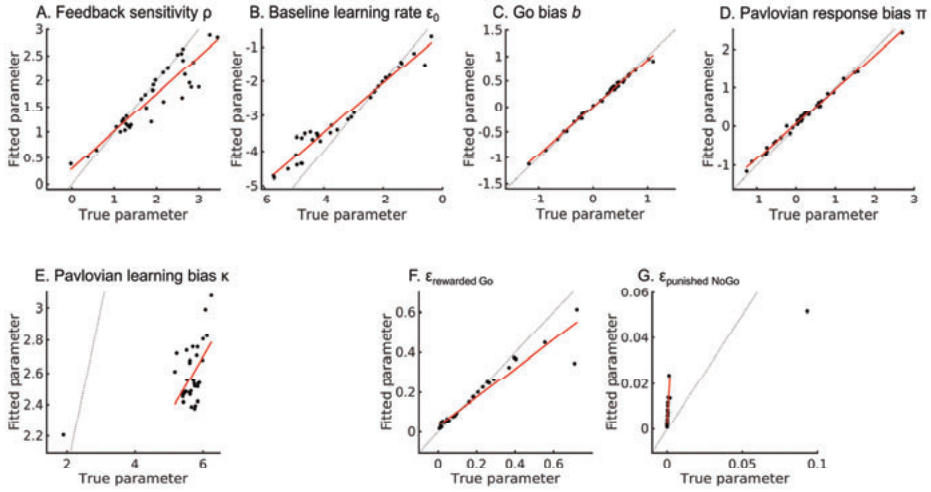


Figure 3.11. S3.4. Parameter recovery results for the asymmetric pathways (M5) model.
The feedback sensitivity parameter ϱ **(A)**, the baseline learning rate ε_0 (**B**), the Go bias *b* (**C**), and the Pavlovian response bias π (**D**) all showed excellent parameter recovery, i.e., between-participants correlations of ground-truth and fitted parameters all exceeded $r > 0.90$. Parameters ϱ and ε_0 are still in sampling space and thus untransformed (which means they can be negative). Dashed lines represent the identity line; red solid lines represent a linear regression line of fitted parameters regressed onto true parameters. Only recovery of the learning bias parameter ϰ (**E**) was not quite as good, though the correlation between ground-truth and fitted parameters was still strongly positive ($r > 0.50$). Note an outlier at the bottom left of ϰ values; the regression line was fitted without this data point. When combining the baseline learning rate ε_0 with the learning bias ϰ to compute the biased learning rates for rewarded Go actions $\varepsilon_{\text{rewarded Go}}$ (**F**) and punished NoGo actions $\varepsilon_{\text{punished NoGo}}$ (**G**), correlations between ground-truth and fitted parameter values were considerably higher (r's > 0.86). Note again an outlier at the top right of for $\varepsilon_{\text{punished NoGo}}$ values; the regression line was fitted without this data point.

3.6.5 S3.5: Simulations for asymmetric pathways and action priming model

Motivational learning biases are predicted by the *asymmetric pathways model* (Frank 2005; Collins and Frank 2014): Positive PEs, elicited by rewards, lead to long-term potentiation in the striatal direct "Go" pathway (and long term depression in the indirect pathway), allowing for a particularly effective acquisition of Go actions to obtain rewards. Conversely, negative PEs, elicited by punishments, lead to long term potentiation in the NoGo pathway, impairing the unlearning of NoGo actions in face of punishments.

An alternative account has recently suggested that self-generated (Go) actions lead to preferential learning (relative to non-self-generated actions, including inaction), more generally (henceforth called "action priming model")(Cockburn et al. 2014). A self-generated action could "prime" basal ganglia circuits and lead to subsequently larger PEs and thus faster learning. The main differential prediction between these two models is how they account for the failure to learn "Go" actions to avoid punishment: In the first model, this is due to a failure to unlearn punished "NoGo" actions, while in the second model, this is due to increased unlearning of punished "Go" actions.

Here, we directly tested both models against each other. We specified an alternative model M6 (Cockburn et al. 2014) with two separate learning rates, one learning rate for trials where self-generated (Go) action selection should prime the processing of any following salient outcome (i.e., Go actions followed by rewards/ punishments), and one learning rate for any other action-outcome combination. In this model, equation (6) was substituted by equation (7):

$$\varepsilon = \begin{cases} \varepsilon_{salGo} & for\ any\ Go\ action\ with\ salient\ outcomes \\ \varepsilon_0 & else \end{cases} \quad (7)$$

When comparing all models M1–M6 using Bayesian model selection, M5 (the asymmetric pathways model) received highest support (model frequency: 68.15%; protected exceedance probability: 99.70%), also compared to M6 (the action priming model; model frequency: 24.19%; protected exceedance probability: 0.30%). In fact, as visible in Fig. S3.4E-H, the action priming did not reproduce the motivational biases in learning curves and bar plots, which constitutes a case of qualitative model falsification (Nassar and Frank 2016; Palminteri et al. 2017). If anything, it seemed that the action priming model traded off both biases, leading to negative response biases for a majority of participants. In contrast, the asymmetric pathways model (M5) was well able to capture the qualitative patterns observed in the data (Fig. S3.4A-D). We conclude that only the asymmetric pathways model is able to qualitatively reproduce core characteristics of our data.

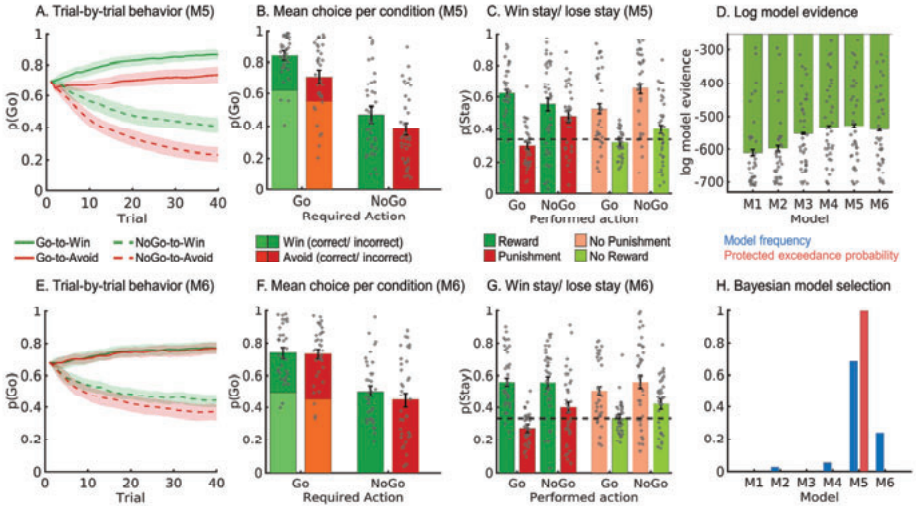


Figure 3.12. S3.5. Model comparison and validation of asymmetric pathways (M5) and action priming (M6) model.
(A-C) One-step-ahead predictions using parameters (hierarchical Bayesian inference) of the winning model asymmetric pathways model (M5). **A.** Trial-by-trial proportion of Go responses (±SEM across participants) for Go cues (solid lines) and NoGo cues (dashed lines); **B.** Mean (±SEM across participants) proportion Go responses per cue condition (points are individual participants' means); **C.** Probability of repeating a response ("stay") on the next encounter of the same cue as a function of action and outcome. The asymmetric pathways model was well able to capture core characteristics of the empirical data (see Fig. 3.2 in the main text). **D.** Log-model evidence favors the asymmetric pathways model (M5), even over the action priming model (M6). **E-G.** Trial-by-trial proportion of Go responses, mean proportion Go responses, and probability of for the action priming model (M6). This model did not reproduce motivational biases (i.e., the difference between green and red lines and bars) well. **H.** Model frequency and protected exceedance probability indicate best fit for model M5 (asymmetric pathways model), in line with log model evidence.

3.6.6 S3.6: Behavioral and neural results from the perseveration model M7, cue valence-modulated perseveration model M8, and neutral-outcome reinterpretation model M9

While the winning model M5 captured learning curves and the proportion of (correct/ incorrect) Go and NoGo responses well, it did not fully capture the propensity of staying (i.e., repeating the same response) in different action-outcome conditions (see Fig. 3.2G). Specifically, M5 underestimated the overall propensity of staying and did not capture the fact that the propensity of staying was (numerically, but not significantly) higher after non-rewarded Go actions than non-punished Go actions. We thus explored three extensions of M5 that had the potential to better capture this behavioral pattern. Specifically, we considered mechanisms that would make the model more likely to repeat a given response. Furthermore, any such mechanism should boost repetition of Go responses that were not rewarded in particular.

We hypothesized that two potential mechanisms could account for these data features, and present three new models to test these mechanisms. As a first mechanism, we considered overall "response stickiness" or "perseveration" (Rutledge et al. 2009), a process that leads participants to repeat a previously shown response independent of the obtained outcome. This mechanism could explain participants' overall higher propensity of staying. However, to account for the fact that staying tended to be higher after non-rewarded Go actions (i.e., Go actions to Win cues that were not rewarded) compared to non-punished Go actions (i.e., Go actions to Avoid cues that were not punished), we tested whether separate perseveration parameters for Win and Avoid cues could capture this behavioral difference.

As a second mechanism, we considered the possibility that participants might "re-interpret" neutral outcomes in line with the cue valence: although a non-reward after a Win cue constitutes negative feedback, the positive cue valence might "overshadow" this feedback and give participants the impression that they received a reward. Similarly, a non-punishment after an Avoid cue constitutes positive feedback, but the negative cue valence might overshadow this feedback and give participants the impression that they received a punishment.

Based on these mechanisms, we fitted three new models to the response data:

Model M7, called "***single perseveration model***", featured the same parameters as M5 plus a perseveration parameter φ that was added as a "bonus" to the action weight $w(a_i, s_t)$ of the specific action shown on the last occurrence of the respective cue (Rutledge et al. 2009):

$$w(a_i, s_t) = \begin{cases} w(a_i, s_t) + \varphi & if\ last\ action\ to\ same\ cue\ was\ a_i \\ w(a_i, s_t) & else \end{cases} \quad (8)$$

M7 featured equations 1–6 and 8. Bayesian model selection suggested that this model quantitively fitted behavior much better than base models M1–M5 (model frequency: 100%; protected exceedance probability: 100%; see Fig. S3.6A panels D, H). Model simulations indeed displayed a higher overall propensity of staying (at a level similar to the empirical data; Fig. S3.6A panel G). However, model M7 systematically overestimated the proportion of incorrect Go responses to both Win and Avoid cues (Fig. S3.6A panel F). Also, it underestimated the difference

in staying after rewarded vs. punished Go and (particularly) NoGo responses, respectively (Fig. S3.6A panel G). Finally, it underestimated the propensity of staying after non-rewarded Go responses, similar to M5. In sum, although this model showed a quantitatively superior fit to the data compared to the base models M1–M5, it did not completely capture the qualitative patterns of incorrect Go responses and the propensity of staying.

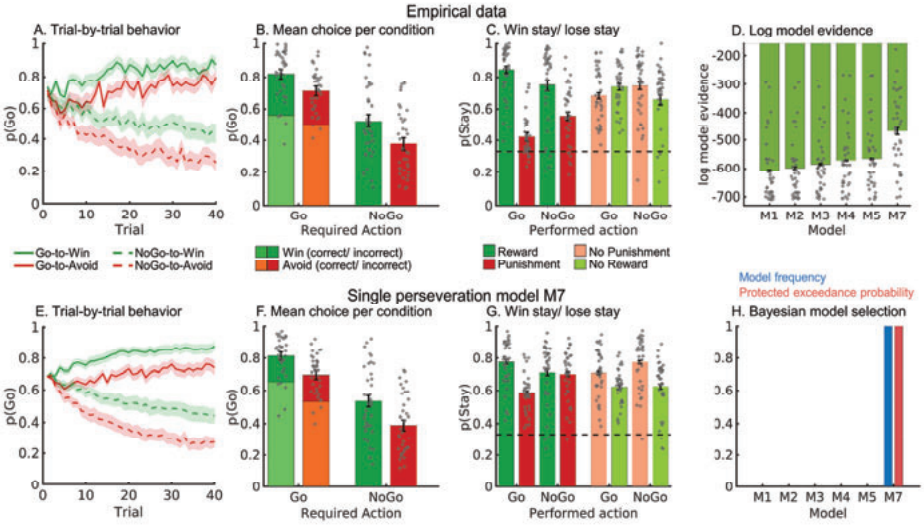


Figure 3.13. S3.6A. Model comparison and validation of the single perseveration model (M7).
A-C. Characteristic patterns of biased action selection and learning in the empirical data. **A.** Trial-by-trial proportion of Go responses (±SEM across participants) for Go cues (solid lines) and NoGo cues (dashed lines). **B.** Mean (±SEM across participants) proportion Go responses per cue condition (points are individual participants' means). **C.** Probability to repeat a response ("stay") on the next encounter of the same cue as a function of action and outcome. **D.** Log-model evidence favored the single perseveration model (M7) over the five base models M1–M5 described in the main text. While this model (**E)** predicted the trial-by-trial proportion of Go responses well, it (**F**) overestimated the proportion of incorrect Go responses to both Win and Avoid cues (light green and red bars). (**G**) The model predicted an overall higher propensity of staying after any action-outcome condition. However, it underestimated the difference in p(Stay) after rewarded and punished Go and NoGo responses (dark green and red bars) as well as the overall p(Stay) pattern after non-rewarded Go responses (light green bar for Go responses). (**H**) Model frequency and protected exceedance probability indicated a better fit for model M7 compared to the base models M1–M5, in line with the log model evidence.

As a second alternative, we considered **M8, the "cue valence-dependent perseveration model"**, which contained two separate perseveration parameters, one for Win cues φ_{WIN}, and one for Avoid cues φ_{AVOID} . The respective perseveration parameter was added to the action weight $w(a_i, s_t)$ of the specific action shown on the last occurrence of respective the cue:

$$w(a_i, s_t) = \begin{cases} w(a_i, s_t) + \varphi_{WIN} & if\ Win\ cue\ and\ last\ action\ to\ same\ cue\ was\ a_i \\ w(a_i, s_t) + \varphi_{AVOID} & if\ Avoid\ cue\ and\ last\ action\ to\ same\ cue\ was\ a_i \\ w(a_i, s_t) & else \end{cases} \quad (9)$$

M8 featured equations 1–6 and 9. Bayesian model selection yielded that M8 fitted the behavior quantitatively better than either the base models M1–M5 or the single perseveration model M7

(model frequency: 92.94%; protected exceedance probability: 100%; see Fig. S3.6B panels D, H). Furthermore, although it matched the propensity of staying after the different action-outcome pairings quite well, it overestimated the propensity of staying after rewarded and non-rewarded NoGo responses (see Fig. S3.6B panel G). Furthermore, it overestimated the proportion of incorrect Go responses to Win and Avoid cues (see Fig. S3.6B panel F). In sum, although this model showed a quantitatively superior fit to the data compared to both the base models M1–M5 and the single perseveration model M7 and furthermore fitted the overall pattern of staying quite well, it did not completely capture the pattern of incorrect responses and still mis-predicted the propensity of staying.

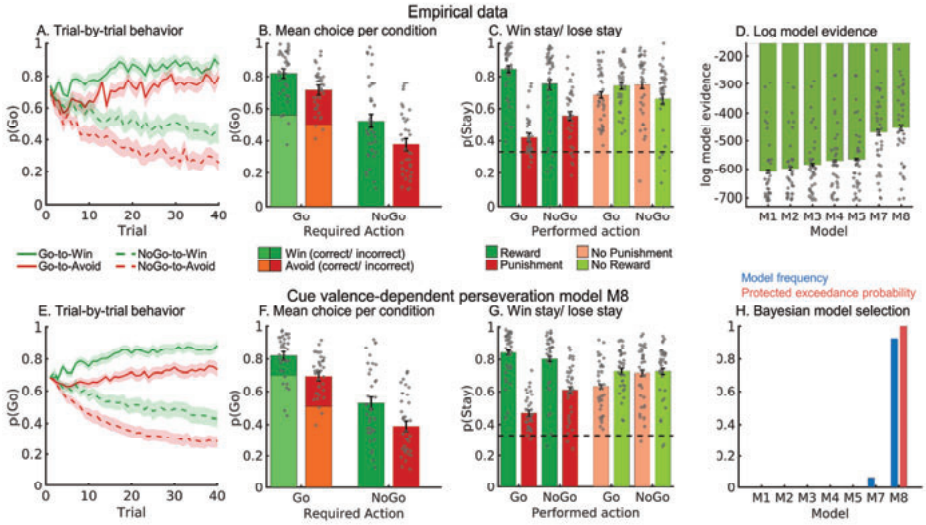


Figure 3.14. S3.6B. Model comparison and validation of the cue valence-dependent perseveration model (M8).
A-C. Characteristic patterns of biased action selection and learning in the empirical data. **A.** Trial-by-trial proportion of Go responses (±SEM across participants) for Go cues (solid lines) and NoGo cues (dashed lines). **B.** Mean (±SEM across participants) proportion Go responses per cue condition (points are individual participants' means). **C.** Probability to repeat a response ("stay") on the next encounter of the same cue as a function of action and outcome. **D.** Log-model evidence favored the cue valence-dependent perseveration model (M8) over the five base models M1–M5 described in the main text and the single perseveration model (M7). While this model (**E**) predicted the trial-by-trial proportion of Go responses well, it (**F**) overestimated the proportion of incorrect Go responses to both Win and Avoid cues (light green and red bars). **G**. The model predicted an overall higher propensity of staying after any action-outcome condition, although overestimating p(Stay) after rewarded (dark green bar) and non-rewarded (light green bar) NoGo responses. **H.** Model frequency and protected exceedance probability indicated a better fit for model M8 over the base models M1–M5 and model M7, in line with the log model evidence.

Lastly, we considered **M9,** called the "**neutral outcomes reinterpretation model**", which featured a single perseveration parameter φ as in equation (8), but in addition replaced neutral outcomes (coded as zero) with the cue valence $V(s)$ scaled by the parameter η as in equation (10):

$$r_{EFF} = \begin{cases} \mathrm{V(s)} * \eta & if\ r = 0 \\ r & else \end{cases} \tag{10}$$

We subsequently used r_{EFF} for computing prediction errors. **M9** featured equations 1–6, 8, and 10. The "reinterpretation" of neutral outcomes captures the intuition that neutral outcomes

for Win cues might still "feel" like a partial reward and lead to a higher likelihood of showing the same response again. Similarly, neutral outcomes after Avoid cues might still "feel" like a partial punishment and lead to a higher likelihood of switching to a different response. $V(s)$ was the cue valence, which was set to +0.5 for Win cues and -0.5 for Avoid cues. Note that cue valence was set to 0 until participants receive the first non-neutral (i.e., reward or punishment) outcome for a given cue, in line with how the Pavlovian response bias was "muted" until the first non-neutral outcome. Also note that for $\eta = 0$, neutral outcomes stay at zero and M9 becomes equivalent to M7.

Bayesian model selection suggested that M9 was inferior to M8 in quantitative fit (model frequency: M8: 62.93%, M9: 37.07%; protected exceedance probability: M8: 93.78%, M9: 6.22%), though still better than the base models M1–M5 (see Fig. S3.6C panel D, H). Simulations showed that M9 captured learning curves accurately, but overestimated the proportion of incorrect Go responses to both Win and Avoid cues (see Fig. S3.6C panel F). Furthermore, the model predicted the propensity of staying quite well, but overestimated staying after punished Go and NoGo responses, while it underestimated staying after punished NoGo and non-rewarded Go responses. In sum, M8 provided an inferior quantitative fit to the data compared to model M8. It did not accurately capture the pattern of incorrect responses and still mis-predicted the propensity of staying.

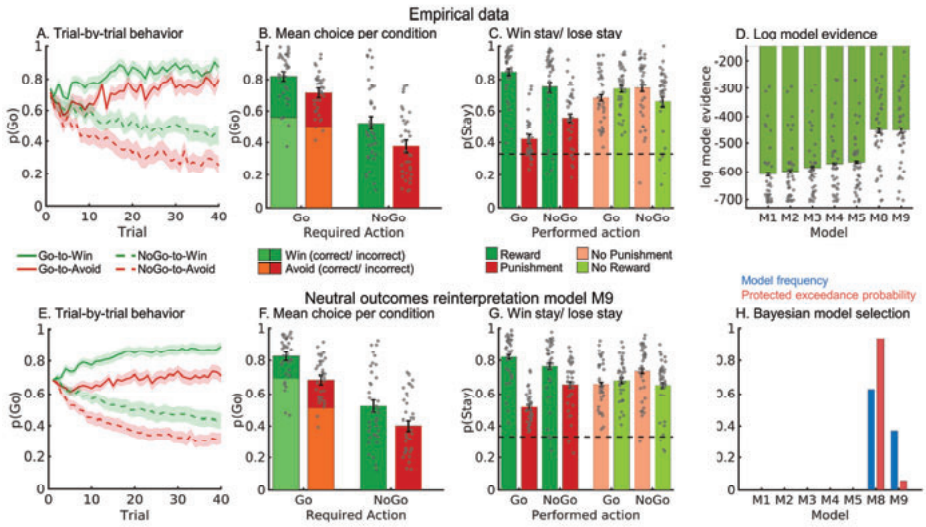


Figure 3.15. S3.6C. Model comparison and validation of the neutral outcomes reinterpretation model (M9).
A-C. Characteristic patterns of biased action selection and learning in the empirical data. **A.** Trial-by-trial proportion of Go responses (±SEM across participants) for Go cues (solid lines) and NoGo cues (dashed lines). **B.** Mean (±SEM across participants) proportion Go responses per cue condition (points are individual participants' means). **C.** Probability to repeat a response ("stay") on the next encounter of the same cue as a function of action and outcome. **D.** Log-model evidence favors the cue valence-dependent perseveration model (M8) over the five base models M1–M5 described in the main text and the neural outcomes reinterpretation model (M9). While this model (**E**) predicted the trial-by-trial proportion of Go responses well, it (**F**) overestimated the proportion of incorrect Go responses to both Win and Avoid cues (light green and red bars). (**G**) The model predicted an overall higher propensity of staying after any action-outcome condition, although overestimating p(Stay) after punished Go and NoGo (dark red bars) and underestimating p(Stay) after non-rewarded Go (light green bar) responses. (**H**) Model frequency and protected exceedance probability indicate best fit for model M8 over the base models M1–M5 and model M9, in line with log model evidence.

In sum, the three additional models provided a quantitatively superior fit to the data compared to the winning model M5 reported in the main text. Also, these additional models predicted the propensity of staying more accurately than the base models did. However, they systematically overestimated the proportion of incorrect Go responses. Furthermore, although the predicted patterns of the propensity of staying mimicked the data more closely than M5, these predicted patterns still mis-matched some aspects of the data. Taken together, these models could capture certain qualitative patterns in the data, but not others, which was expectable given that computational modeling constitutes a data reduction procedures that necessarily loses some details of the data (Nassar and Frank 2016; Palminteri et al. 2017). Given that these additional models combined several mechanisms that could explain the differential propensities of staying after Go/NoGo actions, we decided to report results for model M5 in the main text. Only M5 featured a single mechanism and thus allowed for an unambiguous characterization of neural correlates of this mechanism. In contrast, BOLD signal better captured by the prediction error updates from models M7–M9 could be due to either biased updates after rewarded Go and punished NoGo actions, due to (cue valence-specific) perseveration, or due to a reinterpretation of neutral outcomes, leaving ambiguity about the targeted mechanism.

To show that neural correlates of biased prediction-error updating did not disappear under these alternative model specifications, we repeated the model-based fMRI analyses for both the cue valence-dependent perseveration model M8 and the neutral outcomes interpretation model M9. Notably, M8 does not make different predictions about trial-by-trial learning updates; the only difference to M5 consisted in slightly different best fitting parameter estimates for ε and $\varkappa$. Neural correlates of learning typically reflect the qualitative learning pattern, which is the same for M5 and M8, but are hardly sensitive to the exact parameter values (Wilson and Niv 2015). Indeed, when we repeated our fMRI analyses with those different parameter values, we found almost identical results, with significant encoding of both PE_{STD} and PE_{DIF} in striatum, dACC, pgACC, PCC, left motor cortex, left ITG, and V1 (see Fig. S3.6D panel A). The only exception was the cluster in dACC, which was not significant at a whole-brain level, but significant when using small-volume correction with an anatomical ACC mask (from the Harvard-Oxford Atlas), warranted by our a-priori hypotheses based on previous literature (Behrens et al. 2007).

When we repeated our fMRI analyses with learning updates predicted by M9, we again found significant encoding of both PE_{STD} and PE_{DIF} in striatum, dACC, pgACC, PCC, left motor cortex, left ITG, and V1 (see Fig. S3.6D panel B). However, the pgACC cluster was much larger and extended into the vmPFC. Similarly, the PCC cluster was much larger. In addition, BOLD signal in left inferior frontal gyrus and in multiple clusters in superior and inferior lateral occipital cortex encoded both PE_{STD} and PE_{DIF} significantly. Using trial-by-trial BOLD signal from the extended vmPFC and PCC clusters identified with M9 regressors to predict midfrontal EEG power, we obtained results that were highly similar to the results for the pgACC and PCC clusters identified with M5 regressors.

In sum, model-based fMRI analyses based on PEs derived from M8 and M9 replicated the findings based on M5 reported in the main text. In addition, M9 led to larger clusters in vmPFC and PCC, suggesting that these regions might contribute to reinterpreting neutral outcomes based on cue valence (see also Fig. 3.2 in the main text).

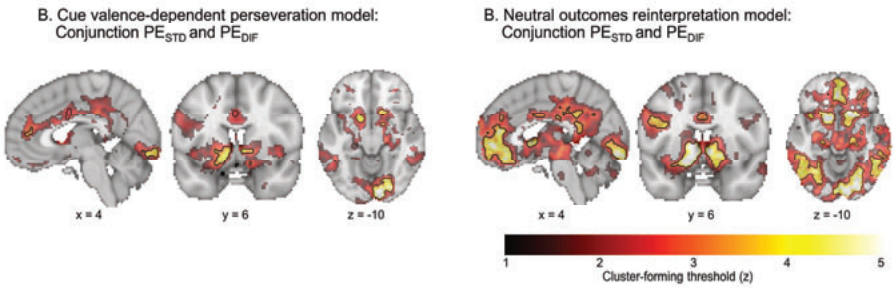


Figure 3.16. S3.6D. BOLD correlates of biased prediction errors as predicted by the cue valence-dependent perseveration model (M8) and the neutral outcomes reinterpretation model (M9).

A. Regions encoding both the standard PE term and the difference term to biased PEs (conjunction) as predicted from the cue valence-dependent perseveration model (M8) at different cluster-forming thresholds ($1 < z < 5$, color coding; opacity constant). Clusters significant at a threshold of $z > 3.1$ are surrounded by black edges. In line with correlates of biased PEs as predicted by M5, BOLD signal in bilateral striatum, dACC (small-volume corrected), pgACC, PCC, left motor cortex, left inferior temporal gyrus, and primary visual cortex was significantly better explained by biased learning than by standard learning. This finding was not surprising given that adding perseveration to the model did not change the learning mechanism, but only led to slightly different best fitting parameter values. **B.** Regions encoding both the standard PE term and the difference term to biased PEs (conjunction) as predicted from the neutral outcomes reinterpretation model (M9). In addition to the regions in which BOLD signal was significantly better explained by biased than standard PEs as derived from M5 and M8, biased PEs derived from M9 also explained BOLD signal in vmPFC (larger cluster than M5), PCC (larger cluster than M5), left inferior frontal gyrus and multiple clusters in superior and inferior lateral occipital cortex significantly better than standard PEs. These results tentatively suggested that vmPFC, PCC, and these other occipital regions might implement an additional mechanism besides biased learning which encodes the cue valence also at the time of the outcome, biasing the processing of neutral outcomes.

3.6.7 S3.7: Anatomical masks and conjunctions of anatomical and functional masks

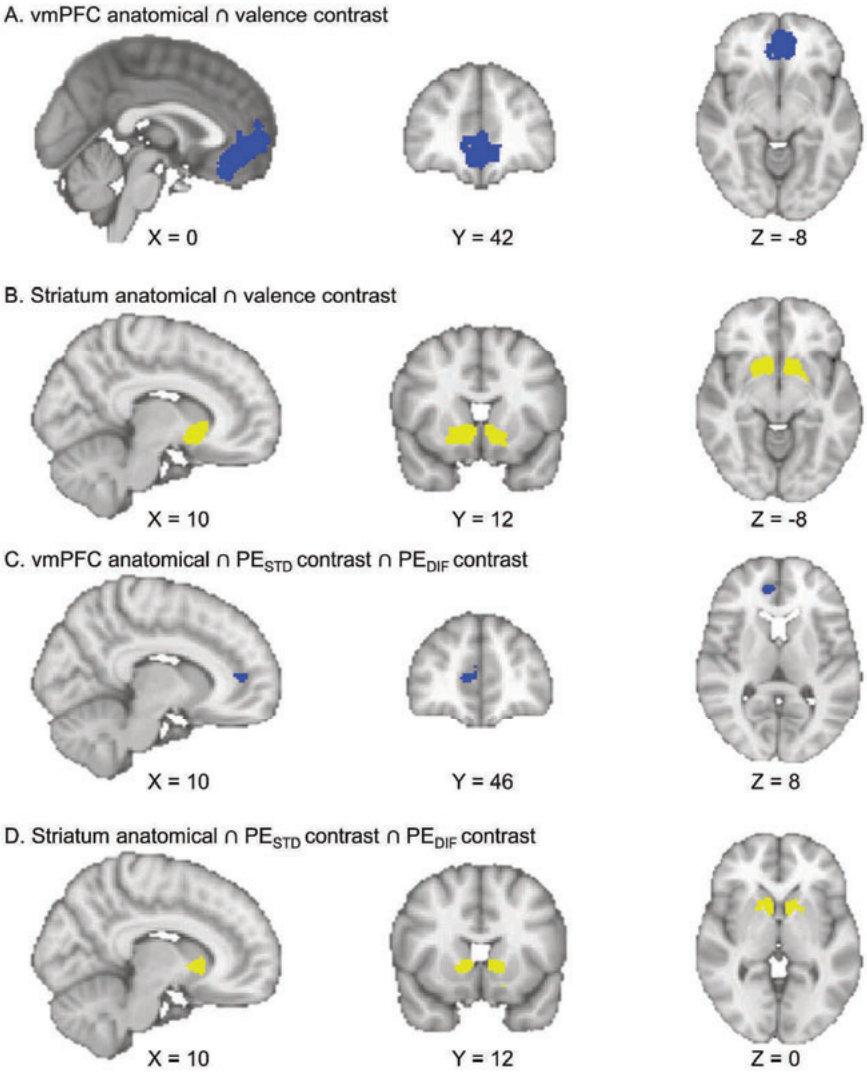


Figure 3.17. S3.7A. Conjunctions of anatomical masks with functional contrasts from fMRI GLM analyses used for fMRI-informed EEG analyses.
Anatomical masks were based on the Harvard-Oxford Atlas. Functional contrasts involve outcome valence and conjunction of PE_{STD} and PE_{DIF}. **A**. vmPFC outcome valence contrast (dark blue, conjunction of frontal pole, frontal medial cortex, and paracingulate gyrus). **B.** striatum outcome valence contrast (yellow, conjunction of bilateral nucleus accumbens, caudate, and putamen). **C.** vmPFC PE_{STD} ∩ PE_{DIF} contrast (dark blue, results in a cluster in pgACC). **D.** striatum PE_{STD} ∩ PE_{DIF} contrast (yellow). All anatomical masks were extracted from the probabilistic Harvard-Oxford Atlas, thresholded at 10%. Note that images are in radiological orientation (i.e., left brain hemisphere presented on the right and vice versa).

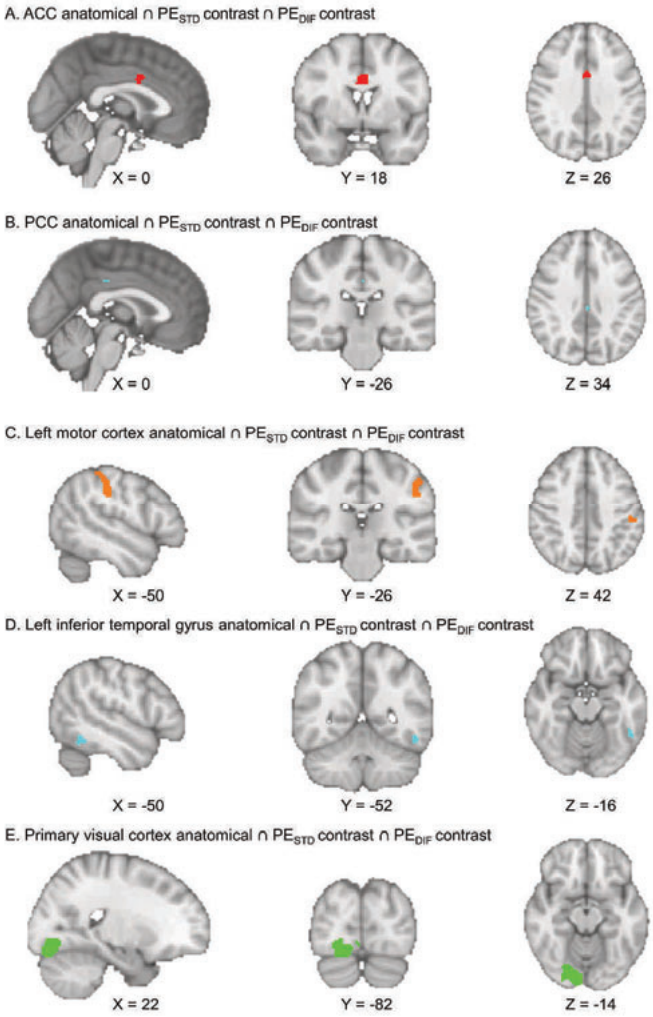


Figure 3.18. S3.7B. Conjunctions of anatomical masks with functional contrasts from fMRI GLM analyses used for fMRI-informed EEG analyses.

A. AAC PE_{STD} ∩ PE_{DIF} contrast (red, cingulate gyrus, anterior division, resulting in a cluster in dACC; **B.** PCC PE_{STD} ∩ PE_{DIF} contrast (light blue, cingulate gyrus, posterior division); **C.** Left motor cortex PE_{STD} ∩ PE_{DIF} contrast (orange, conjunction of precentral and postcentral gyrus). **D.** Left inferior temporal gyrus PE_{STD} ∩ PE_{DIF} contrast (turquoise, conjunction of inferior temporal gyrus, posterior division, and inferior temporal gyrus, temporooccipital part). **E.** Primary visual cortex PE_{STD} ∩ PE_{DIF} contrast (green, conjunction of lingual gyrus, occipital fusiform gyrus, occipital pole). All anatomical masks were extracted from the probabilistic Harvard-Oxford Atlas, thresholded at 10%. Note that images are in radiological orientation (i.e., left brain hemisphere presented on the right and vice versa).

3.6.8 S3.8: Regressors and contrast in fMRI analyses

Model-based GLM with PE_{STD} and PE_{DIF} regressor:

- WinGoOnset: for every trial with Win cue and Go action, at cue onset, duration 1, value +1
- AvoidGoOnset: for every trial with Avoid cue and Go action, at cue onset, duration 1, value +1
- WinNoGoOnset: for every trial with Win cue and NoGo action, at cue onset, duration 1, value +1
- AvoidNoGoOnset: for every trial with Avoid cue and NoGo action, at cue onset, duration 1, value +1
- Handedness: for every trial, at cue onset, duration 1, value +1 for left hand response, 0 for NoGo 10 response, -1 for right hand response 11
- Error: for every trial, at cue onset, duration 1, value +1 for incorrect response, 0 for correct response
- OutcomeOnset: for every trial, at outcome onset, duration 1, value +1 for every trial
- PE_{STD}: for every trial, at outcome onset, duration 1, value is the demeaned PE times learning rate for model M1
- PE_{DIF}: for every trial, at outcome onset, duration 1, value is the demeaned difference between (PE times learning rate) for model M1 and (PE times learning rate) for model M5
- Invalid: for trials where uninstructed button was pressed, at outcome onset, duration 1, value 1

	Regressor	1	2	3	4	5	6	7	8	9	10
	Contrast	WinGoOnset	AvoidGoOnset	WinNoGoOnset	AvoidNoGoOnset	Handedness	Error	Outcome Onset	PE_{STD}	PE_{DIF}	Invalid
1	PE_{STD}								1		
2	PE_{DIF}									1	

Model-free GLM using response-locked and outcome-locked response regressors:

- GoReward: for every trial with Go action and reward obtained, at outcome onset, duration 1, value +1
- GoNoReward: for every trial with Go action and no reward obtained, at outcome onset, duration 1, value +1
- GoNoPunishment: for every trial with Go action and no punishment obtained, at outcome onset, duration 1, value +1
- GoPunishment: for every trial with Go action and punishment obtained, at outcome onset, duration 1, value +1
- NoGoReward: for every trial with NoGo action and reward obtained, at outcome onset, duration 1, value +1
- NoGoNoReward: for every trial with NoGo action and no reward obtained, at outcome onset, duration 1, value +1
- NoGoNoPunishment: for every trial with NoGo action and no punishment obtained, at outcome onset, duration 1, value +1
- NoGoPunishment: for every trial with NoGo action and punishment obtained, at outcome onset, duration 1, value +1
- LeftHand: for every trial with left hand response, at response onset, duration 1, value + 1
- RightHand: for every trial with right hand response, at response onset, duration 1, value +1
- Error: for every trial, at cue onset, duration 1, value +1 for incorrect response, 0 for correct response
- OutcomeOnset: for every trial, at outcome onset, duration 1, value +1 for every trial
- Invalid: for trials where uninstructed button was pressed, at outcome onset, duration 1, value 1

Regressors	1	2	3	4	5	6	7	8	9	10	11	12	13
Contrast	GoReward	GoNoReward	GoNoPunishment	GoPunishment	NoGoReward	NoGoNoReward	NoGoNoPunishment	NoGoPunishment	LeftHand	RightHand	Error	OutcomeOnset	Invalid
1 Valence	1	-1	1	-1	1	-1	1	-1					
2 Action	1	1	1	1	-1	-1	-1	-1					
3 Hand Sum									1	1			
4 Hand Dif									1	-1			

3.6.9 S3.9: Significant clusters in BOLD-GLMs with behavioral regressors only

Model-based GLM with PE_{STD} and PE_{DIF} regressor:

	Contrast				Peak coordinates		
No	Brain region	Maximal Z-value	Cluster size (voxels)	Corrected p	x	y	z
	PE_{STD} Positive						
1	Ventromedial prefrontal cortex, Nucleus accumbens, caudate, putamen, bilateral amygdala, bilateral hippocampus	6.47	8762	1.02e-43	12	14	-6
2	Occipital pole, lingual gyrus, occipital fusiform gyrus	6.64	1012	6.10e-10	10	-92	-10
3	Posterior cingulate cortex	4.72	985	9.40e-10	4	-50	18
4	Left superior frontal gyrus	5.56	910	3.19e-09	-18	34	50
5	Right middle temporal gyrus, anterior division	5.48	381	6.47e-05	62	-4	-18
6	Left inferior temporal gyrus, temporooccipital part	5.16	360	.000103	-52	-46	-10
7	Left middle temporal gyrus, anterior division	4.70	329	.000209	-60	-10	-14
8	Left postcentral gyrus	4.33	271	.000838	-52	-28	48
9	Right cerebellum	4.89	147	.0239	44	-72	-40
10	Anterior cingulate cortex	4.27	146	.0247	2	6	34
	PE_{STD} Negative						
1	Right superior frontal gyrus	5.20	351	.000127	6	26	62
2	Right occipital pole, right inferior lateral occipital cortex	4.76	211	.00391	30	-94	4
3	Left lingual gyrus	4.21	186	.00776	-22	-64	2
4	Left inferior lateral occipital cortex	4.28	147	.0239	-44	-86	-10
	PE_{DIF} Positive						
1	Bilateral superior frontal gyrus, paracingulate gyrus, anterior cingulate cortex, posterior cingulate cortex, ventromedial frontal cortex, bilateral frontal orbital cortex, bilateral frontal pole, bilateral supramarginal gyrus, bilateral middle temporal gyrus, bilateral inferior temporal gyrus, bilateral fusiform gyrus, bilateral inferior occipital cortex, bilateral superior occipital cortex, precuneous, bilateral cerebellum	7.11	35109	0	34	-84	20
2	Right insula, right frontal operculum, right inferior frontal gyrus, right middle frontal gyrus,	6.36	10364	0	34	20	-8

	right frontal orbital cortex, bilateral caudate, bilateral Nucleus accumbens, bilateral thalamus, brainstem						
3	Left insula, left frontal operculum, left inferior frontal gyrus, left middle frontal gyrus, left frontal orbital cortex	6.51	10132	0	-36	20	-6
4	Right middle temporal gyrus, posterior division	4.66	307	.0003	56	-32	-4
5	Right insula, right planum polare	4.72	143	.0248	40	-8	-12
	PE_{DIF} Negative						
1	Left middle temporal gyrus, anterior division	4.22	191	.00607	-64	-6	-14
2	Left hippocampus	4.49	158	.0158	-26	-14	-22

Model-free GLM using response-locked and outcome-locked response regressors:

	Contrast				Peak coordinates		
No	Brain region	Maximal Z-value	Cluster size (voxels)	Corrected p	x	y	z
	Positive > Negative						
1	Ventromedial prefrontal cortex, left lateral orbitofrontal cortex, Nucleus accumbens, caudate, putamen, bilateral amygdala, bilateral hippocampus	5.65	3999	2.86e-19	8	12	-4
2	Left superior frontal gyrus	4.03	331	0.00239	-18	28	60
3	Left lateral orbitofrontal cortex	4.31	288	0.00512	-34	40	-8
4	Right occipital pole	4.59	213	0.0212	18	-92	-16
	Negative > Positive						
1	Right lateral orbitofrontal cortex	4.59	367	0.00142	30	62	-2
2	Precuneous	4.58	356	0.00170	8	-66	58
3	Right superior frontal gyrus	4.32	340	0.00223	12	14	72
	Go > NoGo outcome-locked						
	No significant clusters						
	NoGo > Go outcome-locked						
1	Bilateral lateral orbitofrontal cortex, Bilateral superior frontal gyrus, anterior cingulate cortex, posterior cingulate cortex, pre-SMA, bilateral precentral gyrus, bilateral postcentral gyus, bilateral supramarginal gyrus, bilateral operculum, bilateral planum temporale, bilateral superior temporal gyrus, bilateral middle temporal gyrus, bilateral inferior temporal gyrus, bilateral superior lateral occipital cortex, bilateral inferior lateral occipital cortex, bilateral thalamus	7.32	114090	0	-42	-6	12

	Go (left + right hand response) > NoGo response-locked						
1	Cerebellum, bilateral thalamus, bilateral putamen, bilateral caudate, bilateral Nucleus Accumbens, posterior cingulate cortex, right operculum, right angular gyrus, right superior parietal lobule. anterior cingulate cortex, paracingulate gyrus, bilateral ventrolateral frontal cortex, right middle frontal gyrus	7.08	46437	0	32	-4	-6
2	Left operculum, left angular gyrus, left superior parietal lobule	5.88	3936	3.13e-17	-46	-24	26
3	Intracalcarine cortex	3.79	374	0.00248	-12	-88	6
4	Right middle temporal gyrus	4.63	287	0.00956	68	-32	-12
	NoGo > Go (left + right hand response) response-locked						
1	Right medial temporal gyrus, right temporal pole	4.09	465	0.000636	50	-8	-16
2	vmPFC, subcallosal cortex	3.95	435	0.000973	0	40	-12
	Left Hand > Right Hand Response response-locked						
1	Right precentral gyrus, right postcentral gyrus, right superior parietal lobule, right operculum	7.05	9460	9.41e-39	46	-24	64
2	Left cerebellum	7.18	2208	2.1e-14	-18	-54	-18
	Right Hand > Left Hand Response response-locked						
1	left precentral gyrus, left postcentral gyrus, left superior parietal lobule, left operculum, left thalamus	7.06	14870	0	-36	-20	66
2	Right anterior cerebellum	7.90	3735	1.44e-20	18	-54	-20
3	Right inferior lateral occipital cortex, right superior lateral occipital cortex	4.96	1452	9.66e-11	48	-86	-4
4	Right angular gyrus	4.98	551	2.06e05	66	-50	28
5	Left occipital pole, right intracalcarine cortex	3.93	409	0.000236	-4	-96	26
6	Right posterior cerebellum	4.64	200	0.0157	48	-78	-32

3.6.10 S3.10: EEG time-frequency results after ERPs were removed

Given that differences in theta power between positive and negative outcomes as well as differences in lower alpha band power after Go and NoGo responses occurred quite soon after cue onset, we aimed to test whether these effects reflected differences in evoked rather than induced activity. For this purpose, we removed evoked components from our data by computing the ERP for each of the eight conditions (action x outcome) for each participant and then subtracting the condition-specific ERP from the trial-by-trial data (Cohen and Donner 2013). Only afterwards, we performed time-frequency decomposition.

In line with the results reported in the main text, power was higher for negative compared to positive outcomes in the theta band (p =.018, driven by cluster at 225–475 ms; Fig. S3.8B), but higher for positive than negative outcomes in the beta band (p < .001, driven by cluster at 0–1250 ms; Fig. S3.8C). Notably, unlike the results reported in the main text (Fig. 3.4A), the cluster of high power for negative compared to positive outcomes was constrained to the theta range, and did not extend further into the delta range (Fig. S3.8A). When using the trial-by-trial PEs (both the standard PE and the difference term to a biased PE) as predictors in a multiple linear regression at each time-frequency-channel bin while controlling for PE valence, delta power encoded PE_{STD} positively, though not significantly (p = .198). However, at a later time point around outcome offset, delta (and theta) power in fact correlated negatively with PE_{STD} (575–800 ms, p = .002; Fig. S3.8E). The correlation between delta and the PE_{DIF} term was still positive, but not significant (p = .228, Fig. S3.8F). Similarly, the correlation of the PE_{BIAS} term with delta power was positive, but not significant (p = .084; Fig. S3.8D).

Regarding beta power, there was a positive, though non-significant correlation of beta power with PE_{STD} (p = .096). There was again a significantly negative correlation of beta power with PE_{DIF} (425–875 ms, p < .001, Fig. S3.8B). Likewise, beta power correlated significantly negatively with PE_{BIAS} (450–800 ms, p = .018), driven by the correlation with PE_{DIF}.

In sum, after subtracting the condition-wise ERP from each trial before time-frequency decomposition, supposedly removing the phase-locked aspect of power, both beta and theta still encoded PE valence. However, the encoding of PE magnitude by delta power was attenuated and not significant any more.

This reduction in magnitude encoding might occur of several reasons. Firstly, it might be that this correlation in the delta range was in fact (partly) reflecting correlations with phase-locked, i.e., evoked activity (ERPs), especially in the N2 (FPN)/ P3 (RewP) time range (see S3.9) (Yeung and Sanfey 2004; Sato et al. 2005; Cohen, Wilmes, et al. 2011; Kreussel et al. 2012; Talmi et al. 2013; Bernat et al. 2015; Cavanagh 2015; Proudfit 2015; Sambrook and Goslin 2016; Paul et al. 2020). Nonetheless, a positively correlation between delta power and biased PEs was still visible in Fig. S3.8D, suggesting that at least part of the signal encoding biased PEs was not phase-locked. Secondly, it might be that the removal of the condition-wise ERPs has introduced additional noise in the data, attenuating any true correlation. Thirdly, there was a negative correlation between PE_{STD} and theta/ delta power at later time points which was visible, though not significant in the results reported in the main text (Fig. 3.4D). Subtraction of an ERP-like template acts like a high-pass filter. High-pass filtering at relatively high cut-offs (> 0.5 Hz) can artificially postpone or

induce effects at later points (Tanner et al. 2015). It is possible that in this case, ERP subtraction attenuated a positive correlation in the theta/ delta range, but enhanced a later negative correlation.

Taken together, it is possible that part of the PE magnitude encoding in the theta/ delta range is due to correlations with the phase-locked (ERP) signal. However, this finding did not compromise the conclusion that overall, theta/delta power seemed to be more strongly associated with the PE_{BIAS} term than the PE_{STD} term. Our primary goal was not to pinpoint the precise nature of electrophysiological correlates of biased learning, but rather test the relative temporal order of when different regions exhibiting biased learning signals become active.

Finally, we tested whether after ERP subtraction, low alpha (and beta power) still encoded the previously performed action. When testing for differences in broadband power after Go and NoGo responses, power was indeed significantly different between conditions, driven by clusters in beta band (p = 0.002, 0.125 – 625 ms; p = 0.052, 700 - 1000 ms, 23 - 29 Hz) and theta/ low alpha band (p = 0.024, 575 – 1000 ms, 5–9 Hz; p = 0.056, 0 –225 ms, 6–11 Hz). For power before outcome onset, there were again broadband differences between Go and NoGo (p = 0.002, -1000 – +225 ms, 1 –33 Hz), but note that there was no ERP subtracted before outcome onset. We thus conclude that the differences between Go and NoGo responses were attributable to differences in induced rather than evoked activity.

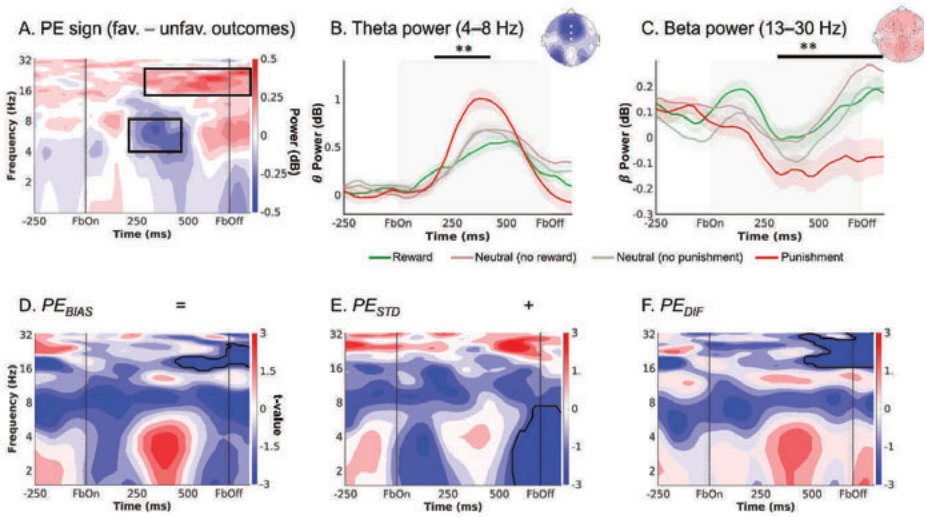


Figure 3.19. S3.10. EEG time-frequency power over midfrontal electrodes (Fz/ FCz/ Cz) after the (action x outcome) condition-wise ERPs has been removed.

A. Time-frequency plot (logarithmic y-axis) displaying high theta (4–8 Hz) power for negative outcomes and higher beta power (16–32 Hz) for positive outcomes. **B.** Theta power transiently increases for any outcome, but more so for negative outcomes (especially punishments) around 225–475 ms after feedback onset. **C**. Beta was higher for positive than negative outcomes (especially punishments) over a long time period around 300–1,250 ms after feedback onset. **D-F**. Correlations between midfrontal EEG power and trial-by-trial PEs. Solid black lines indicate clusters above threshold. There still was a visible positive correlation between biased PEs and midfrontal delta power, but this correlation was not significant (**D**). The correlation of delta with the standard PEs (**E**) was also positive, though not significant; in fact, at a later time point around stimulus offset, delta power correlated significantly negatively with standard PEs. The difference term to biased PEs (F) also correlated positively, though not significantly with delta power. Beta power encoded the difference term and biased PEs themselves (F). ** p < 0.01.

3.6.11 S3.11: ERPs as a function of action and outcome

In addition to the induced activity in time-frequency power reported in the main text, we also analyzed the data in the time domain to test for differences in evoked activity. These analyses were particularly motivated given that differences in time-frequency power between positive and negative outcomes (theta/delta range) and after Go and NoGo responses (lower alpha/ theta range) occurred soon after outcome onset, warranting the assumption that differences might also occur in evoked activity. A large range of previous research has reported a modulation of evoked potentials by outcome valence in form of the feedback-reduced negativity (Yeung and Sanfey 2004; Sato et al. 2005; Foti et al. 2011; Kreussel et al. 2012; Talmi et al. 2013; Proudfit 2015; Sambrook and Goslin 2016; Paul et al. 2020), i.e., a stronger N2 component for negative compared to positive outcomes around ~ 250 post-cue over midfrontal electrodes, recently also characterized as rather constituting a reward positivity (RewP) (Proudfit 2015). Also, some studies have reported a modulation of the P3 by outcome valence, which has been attributed to outcome magnitude or salience rather than valence (Yeung and Sanfey 2004; Sato et al. 2005; Wu and Zhou 2009; Kreussel et al. 2012).

Similar to the analysis of time frequency power, we sorted trials into the eight conditions spanned by the performed action (Go/ NoGo) and the obtained outcome (reward/ no reward/ no punishment/ punishment), computed the average ERP for each condition per participant, and tested for differences between positive (reward/ no punishment) and negative (no reward/ punishment) outcomes as well as conditions of relative stronger (rewarded Go and punished Go) vs. relatively weaker learning (rewarded NoGo and punished NoGo). We used cluster-based permutation tests on the average signal over midfrontal electrodes (Fz/ FCz/ Cz) in the time range of 0–700 ms after outcome onset (where evoked potentials visible in condition-averaged plot).

First, midfrontal ERPs were significantly different between positive and negative outcomes, driven by two separate clusters of differences above threshold (Cluster 1: around 246 – 294 ms, p = .034; Cluster 2: around 344 – 414 ms, p =.004, Fig. S3.9A panel A, C). The first cluster the classical feedback-related negativity, i.e., a stronger N2 component for negative compared to positive outcomes. The second cluster reflected weaker P3 component for negative compared to positive outcomes, similar the reward positivity reported before. In fact, the N3 was rather absent for negative outcomes (Fig. S3.9B). Both effects were clearly focused on midfrontal electrodes. These findings replicate previous findings of outcome valence modulating N2 (feedback-related negativity) and P3 components, and complement our time-frequency findings of theta and beta power reflecting outcome valence.

Second, when contrasting trials with Go vs. NoGo responses, no significant difference was observed (p = .358; Fig. S3.9A panel D). Visual inspection of the topoplot yielded that, if anything, differences emerged over right occipital electrodes. If one performed a test over those right occipital electrodes (O2, 04, PO4; Fig. S3.9A panel F; note that this procedure constitutes double-dipping because the test was informed by first looking at the data), this test would have yielded significant results (p = .016) driven by cluster around 423–466 ms, reflecting a slightly larger P3 after Go than NoGo responses (Fig. S3.9A panel E). This finding appears to be the strongest (if any) difference in amplitude after outcome onset between Go and NoGo actions. Given that this difference was not hypothesized and occurred far away from our a-priori selected channels of interest, we are careful not to over-interpret those differences.

Third, contrasting trials with positive and negative at the same right occipital electrodes yielded a significant difference, driven by clusters around 46–103 ms (p = 0.034), 141–255 ms (p = .002), and 519 – 580 ms (p = .034). Most notably, the P1 amplitude was much larger for positive than negative outcomes (Fig. S3.9A panel B). However, given that these differences were not hypothesized and occurred far away from our a-priori selected channels of interest, we are careful not to over-interpret those differences.

Taken together, we found a bigger midfrontal N2/ FRN for negative compared to positive outcomes, and a bigger midfrontal P3/ RewP for positive compared to negative outcomes, in line with a vast literature of previous findings (Yeung and Sanfey 2004; Sato et al. 2005; Wu and Zhou 2009; Foti et al. 2011; Kreussel et al. 2012; Talmi et al. 2013; Proudfit 2015; Sambrook and Goslin 2016; Paul et al. 2020). Midfrontal voltage did not significantly differ after Go or NoGo responses. If anything, differences after Go and NoGo responses were maximal over right occipital electrodes, with a larger P3 after Go than after NoGo responses. Signal at these channels also differed between positive and negative outcomes, most notably with a bigger P1 after positive than negative outcomes. In sum, we replicate classical reward learning ERP effects, which shows that the motivational Go/NoGo learning task taps into reward learning processes reported before, but these processes appeared to be unaffected by the previously performed action.

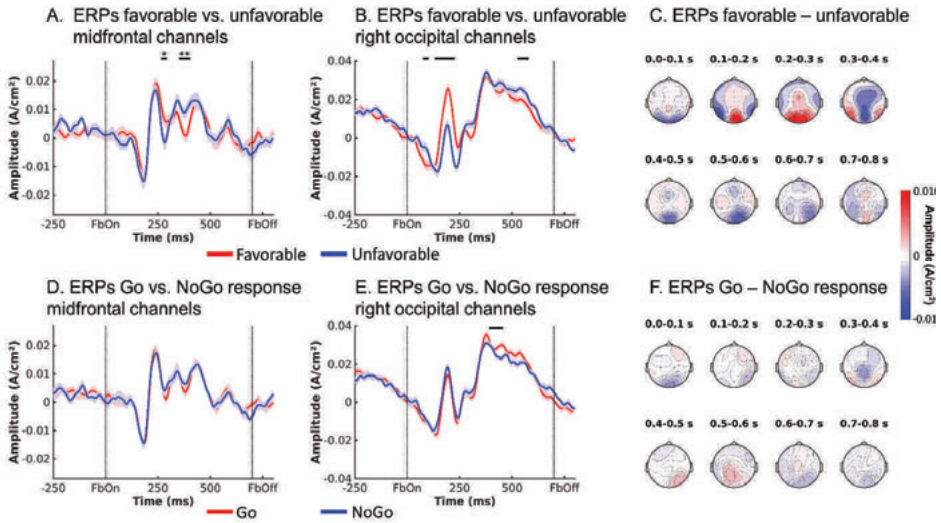


Figure 3.20. S3.11A. ERPs reflecting outcome valence and performed action.
A. Voltage (±SEM) over midfrontal electrodes (Fz/FCz/Cz) was lower for negative than positive outcomes around 246–294 ms (stronger N2, FRN) and higher for positive than negative outcomes around 344 – 414 ms (stronger P3/ RewP). **B.** Over right occipital electrodes, the P3 was slightly bigger for positive than negative outcomes. ** $p < 0.01$. * $p < .05$ **C.** Topoplots of difference in voltage between trials with positive and negative outcomes over selected time windows. **D.** There was no difference in voltage over midfrontal electrodes between trials with Go and NoGo responses. **E.** Over right occipital electrodes, the P3 was slightly stronger after Go than NoGo actions (no *p*-value because ROI selected based on visual inspection). **F.** Topoplots of difference in voltage between trials with Go and NoGo actions over selected time windows.

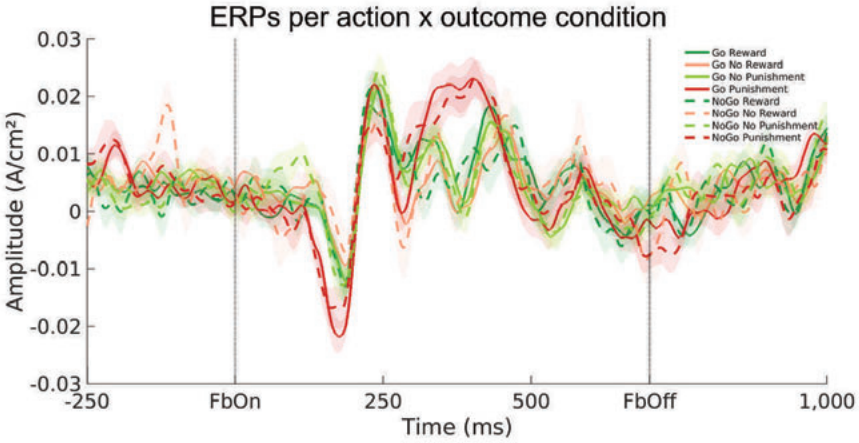


Figure 3.21. S3.11B. ERPs per action x outcome condition.
Biggest differences occurred around the time of the N2 (FRN) and P3 (RewP). N2 and P3 exhibited larger amplitudes on trials with punishments. There was no apparent modulation by the previous action (Go/ NoGo).

3.6.12 S3.12: Model-based EEG analyses in the time domain

In addition to testing whether midfrontal time-frequency power reflected signatures of biased learning (see main text), we also tested whether the midfrontal time domain signal reflected biased learning. Again, we used the standard PE term and the difference term to biased PEs as regressors in a multiple linear regression on each channel-time bin.

Focusing on midfrontal electrodes, and controlling for outcomes valence, first, the PE_{STD} term was negatively correlated with midfrontal voltage around 529–575 ms (p = .039; Fig. S3.10B). Note that so late after outcome onset, signal was not part of any "classical" ERP component any more. Second, the PE_{DIF} correlated negatively with midfrontal voltage around 123–166 ms (p = .029) in the time range of the N1 and later positively around 365–443 ms (p < .001; Fig S3.10C) in the time range of the P3/ RewP. Third, a similar pattern of correlations occurred for the PE_{BIAS} term (Cluster 1: negative, 111–184 ms, p = .004; Cluster 2: positive, 346–449 ms, p < .001; Fig. S3.10A). Fourth, around these same time windows, midfrontal voltage also encoded outcome valence itself, but with opposite sign (Cluster 1: positive, 99–184 ms, p < .001; Cluster 2: negative, 308–448 ms, p < .001; see S3.9).

In sum, similar to analyses of midfrontal power reported in the main text, PE sign and magnitude were encoded in midfrontal voltage around the same time, but with opposite polarity: Signal around the time of the N1 encoded PE sign positively, but PE magnitude negatively. Vice versa, signal around the time of the P3/ RewP encoded PE sign negatively, but PE magnitude positively. The same phenomenon of separate valence and magnitude encoding in midfrontal EEG signal has been reported before (Talmi et al. 2013; Bernat et al. 2015; Cavanagh 2015). Notably, magnitude encoding in midfrontal voltage emerged for the PE_{BIAS} term, but not the PE_{STD}, indicating that this correlation was driven by the PE_{DIF} term and that biased learning described midfrontal voltage better than standard learning. These results complement our findings of theta/delta power encoding outcome valence and magnitude with opposite polarities (see main text).

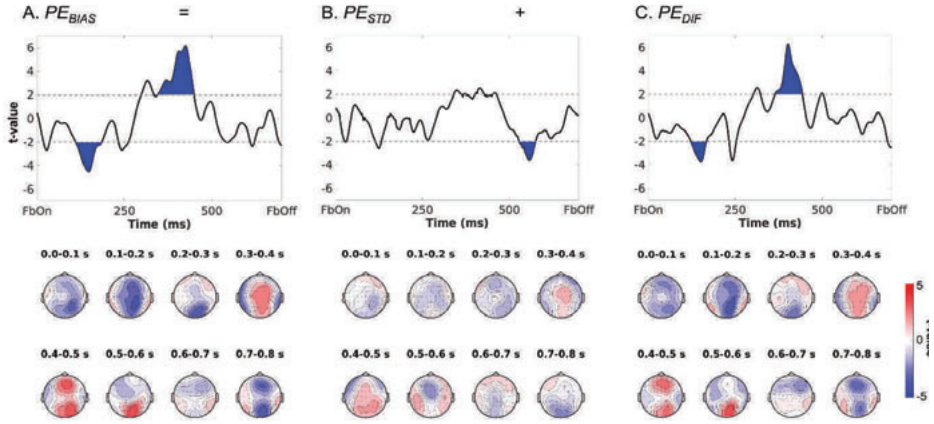


Figure 3.22. S3.12. Modulation of EEG voltage by biased PEs and decomposition into the standard PE term and the difference term to biased PEs.

A. Mean EEG voltage over midfrontal electrodes (Fz, FCz, Cz) was significantly modulated by biased PEs around 111–184 (negatively) and 353–414 ms (positively) after outcome onset. **B.** Correlations with the standard PE term only emerged around 529 – 575 ms (negatively). **C.** Correlations with the difference term to biased PEs were similar to correlations for the biased PE term itself, i.e., around 123–166 (negatively) and 365–443 ms (positively). Bottom row: Topoplots displaying *t*-values of beta-weights for the respective regressor over the entire scalp in steps of 100 ms from 0 to 800 ms.

3.6.13 S3.13: Graphical illustration of the fMRI-informed EEG and EEG-informed fMRI analysis approaches

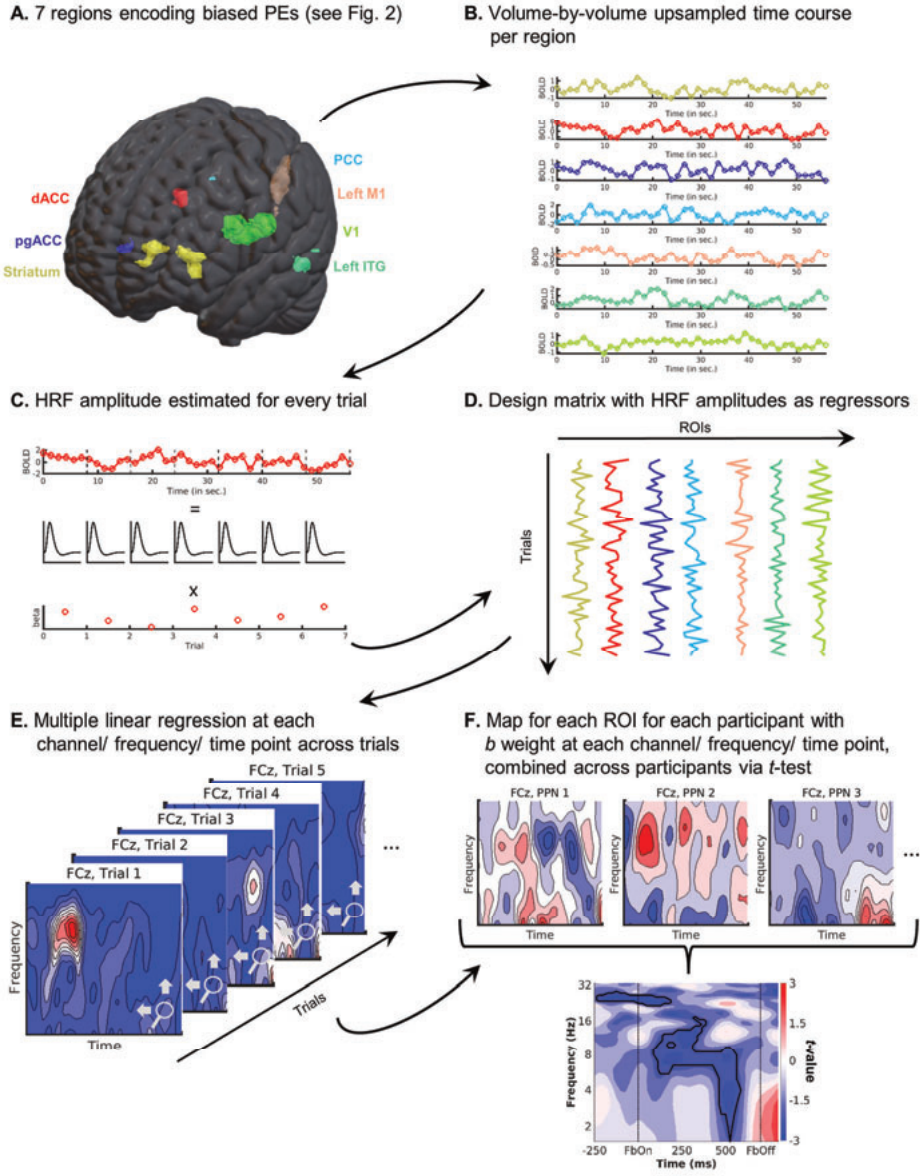
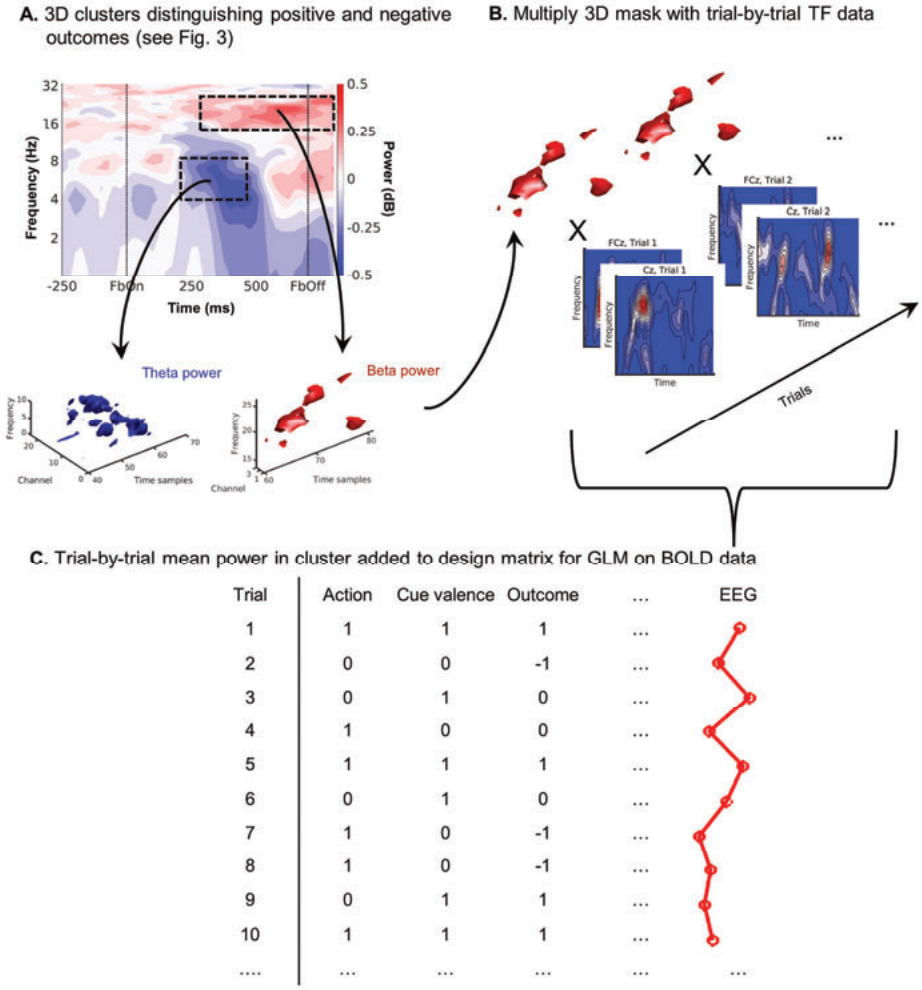


Figure 3.23. S3.13A. Graphical illustration of the fMRI-informed EEG analysis approach.
A. Regions are identified to encode biased PEs via a model-based GLM on BOLD data (see Fig. 3.2). **B.** The volume-by-volume time-series of the signal in each ROI is extracted and upsampled. **C.** Time series are epoched into trials and the HRF amplitude is estimated for every trial. **D.** HRF amplitudes in every ROI for every trial are combined into a design matrix. **E.** The design matrix is applied in a multiple linear regression for each participant at each channel, frequency, and time point across trials. **F.** Regressions yield a sensor-frequency-time map of *b* regression weights for each ROI for each participant. Maps are combined across participants using a one-sample *t*-test.



Trial	Action	Cue valence	Outcome	...
1	1	1	1	...
2	0	0	-1	...
3	0	1	0	...
4	1	0	0	...
5	1	1	1	...
6	0	1	0	...
7	1	0	-1	...
8	1	0	-1	...
9	0	1	1	...
10	1	1	1	...
....	...	...	...	...

Figure 3.24. S3.13B. Graphical illustration of the EEG-informed fMRI analysis approach.
A. 3D clusters of channel-frequency-time points where power significantly distinguishes trials with positive from trials with negative outcomes are identified via a cluster-based permutation test (see Fig. 3.3A). The *t*-values above a threshold |2| are retained, weights at all other grid points are set to zero. **B.** The 3D t-value cluster is multiplied with the trial-by-trial channel-frequency-time data, yielding a single average value of power in the cluster at each trial. **C.** Trial-by-trial average power in the cluster is added as a parametric regressor in the GLM on BOLD-data and fitted with FSL.

3.6.14 S3.14: Supplementary fMRI-inspired EEG results in time-frequency space

Besides the results for striatum, ACC, and PCC reported in the main text, there were also significant EEG correlates over midfrontal electrodes for trial-by-trial BOLD signal from left motor cortex (p = .002, around 0–625 ms, 16–27 Hz; Fig. S3.11A). There were however no significant EEG correlates over midfrontal electrodes for BOLD signal from pgACC (p = .174; Fig. S3.11B), left inferior temporal gyrus (p = .097; Fig. S3.11C), and primary visual cortex (p = .170; Fig. S3.11D).

As quality checks, we checked whether visual cortex BOLD correlated negatively with alpha over occipital electrodes (Scheeringa et al. 2011; Zumer et al. 2014) and whether motor cortex BOLD correlated negatively with beta power over central electrodes (Jurkiewicz et al. 2006; Ritter et al. 2009). Both was the case (see Fig. S3.11E and F), showing that our data was of sufficient quality to detect these well-established associations.

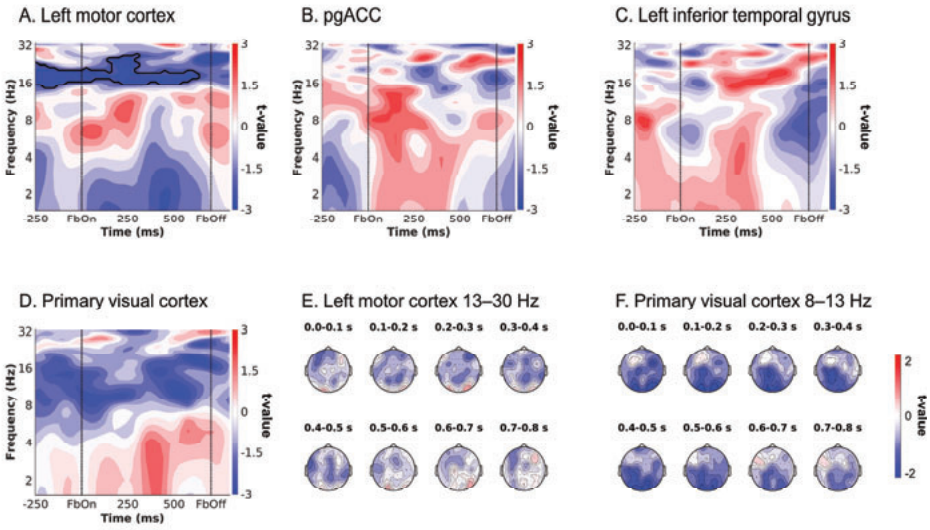


Figure 3.25. S3.14. Supplementary fMRI-informed EEG results in the time-frequency domain.
Unique temporal contributions of BOLD signal in (**A**) left motor cortex, (**B**) pgACC, (**C**) left ITG and (**D**) primary visual cortex to midfrontal EEG power. Group-level *t*-maps display the modulation of the EEG power over midfrontal electrodes (Fz/ FCz/ Cz) by trial-by-trial BOLD signal in the selected ROIs. There significant correlations between midfrontal EEG TF power in the beta range and left motor cortex BOLD signal (p = .002), but no significant midfrontal EEG correlates for BOLD signal from other ROIs. **E.** Topoplots displaying *t*-values of left motor cortex BOLD over the entire scalp between 13 and 30 Hz (beta band) in steps of 100 ms from 0 to 800 ms. There were significant negatively correlates over central electrodes, especially round 300–500 ms. **F.** Topoplot displaying *t*-values of primary visual cortex BOLD over the entire scalp between 8 and 13 Hz (alpha band) in steps of 100 ms from 0 to 800 ms. There were significantly negatively correlations over occipital electrodes throughout outcome presentation.

3.6.15 S3.15: Supplementary fMRI-inspired EEG results in the time domain

For fMRI-inspired analysis of the EEG signal in the time domain (voltage), we applied the same approach as reported in main text, but with voltage signal (time-domain) instead of time-frequency power as dependent variable. As independent variables, we entered the trial-by-trial BOLD signal from all seven regions encoding biased PEs plus the trial-by-trial standard PE and the different term towards the biased PE (exact same procedure as for EEG TF analyses), all in one single multiple linear regression. On a group-level, we again focused on the mean signal over midfrontal electrodes (Fz/ FCz/ Cz) in a time range of 0–700 ms, for which ERPs had been visible in the condition-averaged plots (see S3.9).

First, trial-by-trial striatal BOLD correlated significantly with midfrontal voltage at two time points, namely positively around 152–196 ms (p = .017) in the time range of the N1 and again negatively around 316–383 ms (p < .001, see Fig. S3.12A) in the time range of the N2/ FRN and P3/RewP. Second, trial-by-trial pgACC BOLD correlated significantly positively with midfrontal voltage around 347–412 ms (p = .006, see Fig. S3.12A) in the time range of the N2/ FRN and P3/RewP. Third, trial-by-trial BOLD from primary visual cortex correlated significantly positively with midfrontal voltage around 307–367 ms (p = .011, see Fig. S3.12B), overlapping with (but slightly earlier than) correlations from pgACC BOLD, i.e., in the time range of the N2/ FRN and P3/RewP. For midfrontal voltage split up per high vs. low BOLD signal (revealing which ERP components were respectively modulated), see Fig. S3.12C-E. There were no significantly correlations between midfrontal voltage and trial-by-trial BOLD from dACC (p = .927, see Fig. S3.12A), left motor cortex (p = .649, see Fig. S3.12B), PCC (p = .796, see Fig. S3.12A), or left inferior temporal gyrus (p = .649, see Fig. S3.12B). For further details on BOLD-EEG voltage correlations in the time domain, see Fig. S3.12F–L.

Taken together, trial-by-trial BOLD signal in striatum, pgACC, and V1 all correlated with FRN/ RewP amplitude, which was the dominant phenomenon over midfrontal electrodes reflecting outcome valence (see S3.9 and S3.10). Notably, correlations with striatal and pgACC BOLD were of opposite signs, which aligns with the finding that striatal and pgACC BOLD predicted opposite behavioral tendencies on future trials (see main text; see S3.15). However, crucially, the time domain signal did not allow for a temporal dissociation of these different regions. Possibly, the midfrontal evoked signal (i.e., the part of the signal that was phase-locked to outcome onset) was so stereotyped that only the FRN/ RewP complex showed enough variation across trials to allow for substantial correlations with trial-by-trial BOLD signal. This finding demonstrates that the time-frequency domain signal (i.e., the part of the signal that is not necessarily phase-locked to outcome onset) might be more suited for dissociating the activity of different regions in time.

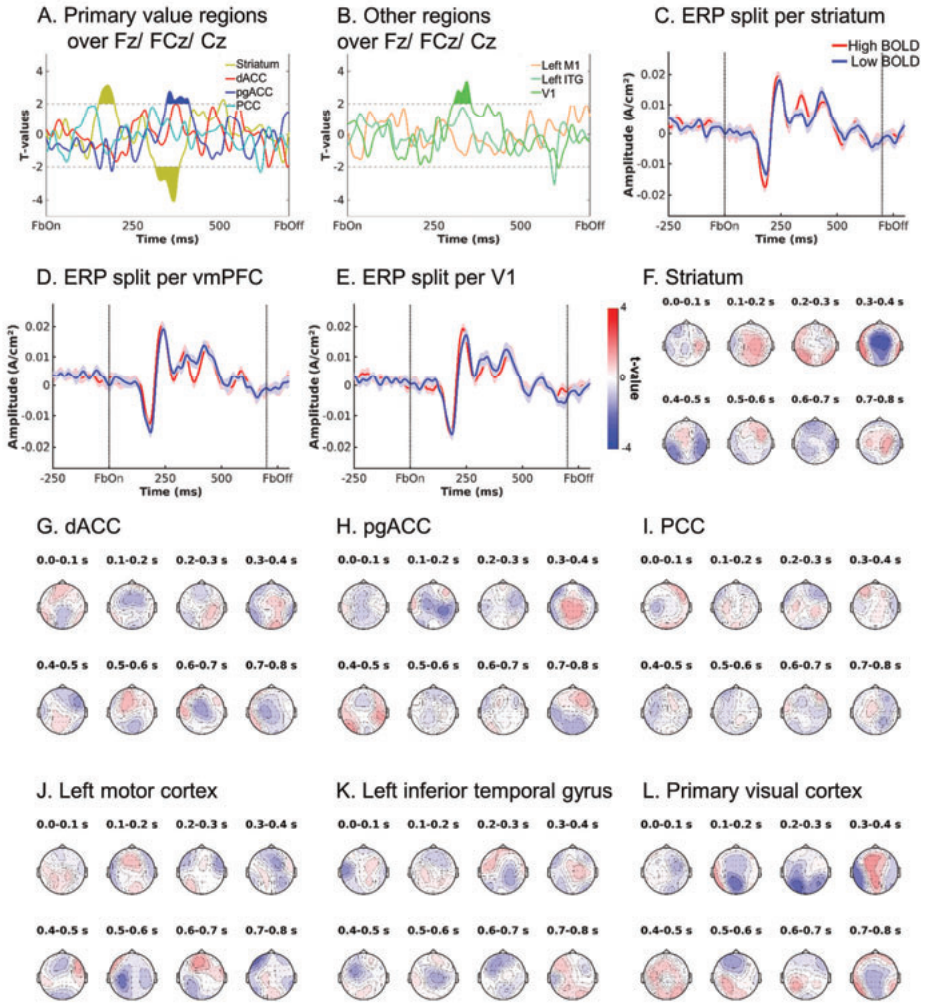


Figure 3.26. S3.15. fMRI-informed EEG analyses in the time-domain.
Group-level *t*-value time courses display the modulation of the EEG voltage over midfrontal electrodes (Fz/ FCz/ Cz) by trial-by-trial BOLD signal in the selected ROIs. **A.** Correlations between midfrontal voltage and trial-by-trial BOLD signal from core value regions, i.e., striatum, dACC, pgACC, and PCC. Striatal BOLD modulates the amplitude of the N1 and P3, while the P3 amplitude was also modulated by pgACC BOLD. **B.** Correlations between midfrontal voltage and trial-by-trial BOLD signal from other regions, i.e., left motor cortex, left inferior temporal gyrus, and primary visual cortex. Visual cortex BOLD modulates the amplitude of the P3, as well. **C-E.** Midfrontal voltage split up for high vs. low BOLD signal (median split) from regions significantly modulating voltage. Striatal BOLD modulated N1 and P2 amplitude, while pgACC BOLD and visual cortex BOLD modulated N2 (FRN) amplitude. **F-L.** Topoplots displaying *t*-values of correlations between midfrontal voltage and trial-by-trial BOLD for all regions in steps of 100 ms from 0 to 800 ms.

3.6.16 S3.16: Full list of significant clusters with EEG regressors in fMRI GLMs

	Contrast				Peak coordinates		
No	Brain region	Maximal Z-value	Cluster size (voxels)	Corrected p	x	y	z
	Central Lower Alpha Band Positive						
	No significant clusters						
	Central Lower Alpha Band Negative						
1	Precuneous, cuneal cortex, right superior lateral occipital cortex	5.78	8346	2.50e-33	6	-60	66
2	Anterior cingulate gyrus, right superior frontal gyrus	4.77	2449	1.75e-14	24	12	66
3	Left middle frontal gyrus,	5.59	1828	7.63e-12	-38	8	34
4	Right insula, right central opercular cortex	4.71	1794	1.08e-11	42	2	28
5	Right frontal pole, right middle frontal gyrus, right inferior frontal gyrus, pars triangularis	5.43	1300	2.37e-09	30	40	20
6	Left supramarginal gyrus, anterior division	4.61	959	1.19e-07	-64	-36	42
7	Left angular gyrus	5.83	916	2.38e-07	-48	-52	18
8	Right cerebellum, anterior	4.79	480	.000131	42	-38	-38
9	Posterior cingulate cortex, parahippocampal gyrus, right thalamus	4.41	424	.000328	14	-38	-2
10	Left temporal pole, left inferior frontal gyrus, pars opercularis left insula	4.08	413	.000394	-56	16	-6
11	Left cerebellum, anterior	5.44	263	.00598	-30	-40	-42
12	Right lingual gyrus	3.43	235	.0104	10	-74	-10
13	Left cerebellum, posterior	5.74	215	.0158	-14	-76	-42
14	Brainstem	4.35	207	.0186	8	-34	-20
	Frontal Theta Band Positive						
1	Right bilateral precentral gyrus	4.82	394	.000577	12	-16	80
2	Left bilateral precentral gyrus	5.25	357	.0011	-20	-28	78
	Frontal Theta Band Negative						
1	Right supramarginal gyrus, posterior division, right superior lateral occipital cortex	3.94	1002	1.10e-07	-54	-50	44
2	Left supramarginal gyrus, posterior division, Left superior lateral occipital cortex	4.39	508	8.96e-05	56	-50	20

3	Posterior cingulate cortex	4.58	419	.000378	-6	-30	38
4	Ventromedial prefrontal cortex	4.03	342	.00143	0	42	4
	Central Beta Band Positive						
1	Right caudate	4.19	258	.00481	16	30	6
2	Left parahippocampal gyrus, posterior divison	4.86	221	.0106	-38	-36	-8
	Central Beta Band Negative						
1	Right frontal pole, right middle frontal gyrus, right superior frontal gyrus	5.49	6599	7.06e-30	-32	8	28
2	Left frontal pole, left middle frontal gyrus, Left superior frontal gyrus	5.51	6144	1.82e-28	40	38	36
3	Left supramarginal gyrus, posterior division, left superior parietal lobule, left superior lateral occipital cortex, Left middle temporal gyrus, temporooccipital part	5.51	5175	2.43e-25	-66	-44	28
4	Right supramarginal gyrus, posterior division, Right superior parietal lobule, right superior lateral occipital cortex	5.13	3264	1.62e-18	30	-74	54
5	Left superior frontal gyrus, paracingulate gyrus, precuneous	4.54	1235	1.80e-09	-4	12	52
6	Right superior temporal gyrus, posterior division	4.59	1076	1.33e-08	48	-14	-10
7	Left temporal pole, left planum temporale	4.96	320	.00139	-46	4	-18

3.6.17 S3.17: Go/NoGo differences in BOLD signal, alpha, and beta power over time

We observed differences between trials with Go responses and trials with NoGo responses in the low alpha power before and shortly after outcome onset (Fig. 3.6A, B main text). Alpha typically increases over the time course of an experiment, potentially related to fatigue and decreasing arousal (Klimesch 1999). If the ratio of Go and NoGo responses changed over time, as well, such an increase over time could spuriously lead to a difference between Go and NoGo responses (though note that this ratio did not noticeably change over time; Fig. S3.14D). To exclude this possibility, we extracted trial-by-trial time-frequency power from the three significant clusters report in the main text in which power differed between Go and NoGo responses: i) lower alpha band power after outcome onset, ii) lower alpha band power before and after outcome onset, iii) beta band power before outcome onset. We log10-transformed this data to decibel and analyzed it as a function of the performed response (factor), block number (1–6; z-standardized), and the interaction between both. We reasoned that if power differences occurred merely due to fatigue effects, the main effect of performed response should not be significant when accounting for time on task (i.e., block number).

For lower alpha band power after outcome onset, there was a significant main effect of performed response, $b = 0.035$, $SE = 0.015$, $\chi^2(1) = 5.350$, $p = .021$, with higher power for Go than NoGo responses, a significant main effect of block number with lower alpha band power increasing over time, $b = 0.052$, $SE = 0.019$, $\chi^2(1) = 6.645$, $p = .010$, but no significant interaction, $b = 0.003$, $SE = 0.008$, $\chi^2(1) = 0.156$, $p = .693$. As Fig. S3.14A reveals, lower alpha band power was consistently higher after Go than after NoGo responses for every block of the task, suggesting that differences in lower alpha band power were not merely due to time on task.

For lower alpha band power before and after outcome onset, as well, there was a significant main effect of performed response, $b = 0.068$, $SE = 0.030$, $\chi^2(1) = 5.010$, $p = .025$, with higher power after Go than NoGo responses, a significant main effect of block number with lower alpha band power increasing over time, $b = 0.072$, $SE = 0.029$, $\chi^2(1) = 6.757$, $p = .016$, but no significant interaction, $b = 0.010$, $SE = 0.009$, $\chi^2(1) = 1.184$, $p = .277$ (Fig. S3.14B), leading to identical conclusions.

For beta band power before and after outcome onset, there was a significant main effect of performed response, $b = 0.083$, $SE = 0.032$, $\chi^2(1) = 6.301$, $p = .012$, with higher power after Go than NoGo responses, a significant main effect of block number with beta power decreasing over time, $b = -0.042$, $SE = 0.021$, $\chi^2(1) = 4.007$, $p = .045$, but no significant interaction, $b = 0.001$, $SE = 0.007$, $\chi^2(1) = 0.030$, $p = .864$ (Fig. S3.14C). In sum, even in presence of changes in power over the time course of the task, lower alpha band and beta band power were consistently higher after Go responses than after NoGo responses, suggesting that these effects were not due to time on task.

Furthermore, we asked whether differences in dACC BOLD between trials with Go and trials with NoGo response at the time of the outcome were due to outcome-related activity or might rather the reflect action on the next trial. We thus plotted the "raw" BOLD signal per action x outcome condition. We used the first eigenvariate of the BOLD in signal in the dACC cluster that reflected biased learning, upsampled the BOLD signal, epoched it into trials relative to outcome

onset (same procedure as for fMRI-informed EEG analyses), and averaged the signal across trials and participants separately per performed action (Go/NoGo) and outcome valence (positive/negative). This plot yielded higher dACC BOLD signal on trials with NoGo responses than on trials with Go responses at the time of outcomes (Fig. S3.14E). However, this difference could potentially be driven by the response on the following task. Hence, we further split the data according to whether the action on the following trial was a Go or a NoGo response. Irrespective of the action on the following trial, dACC BOLD signal was higher when the action on the current trial was a NoGo response compared to a Go response (Fig. S3.15F). In sum, these analyses corroborate that dACC BOLD signal was indeed higher after NoGo than Go responses at the time of outcomes.

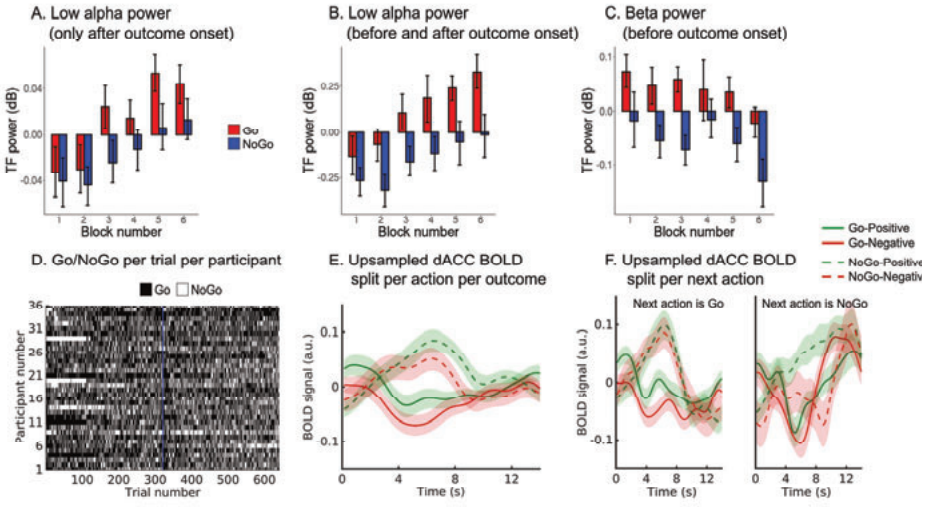


Figure 3.27. S3.17. Control analyses excluding temporal confounds in midfrontal lower alpha band power and dACC BOLD. **A.** Mean midfrontal low alpha power (±SEM across participants) after outcome onset, (**B**) before and after outcome onset, and (**C**) beta power before outcome onset as a function of the performed action and block number (i.e., time on task). While low alpha power increases and beta power decreases over the time course of the task, power was always consistently higher for trials with Go than trials with NoGo responses, suggesting that action effects were not reducible to time on task. **D.** Response for each participant (rows) on each trial (columns). There was no noticeable change in the overall ratio of Go to NoGo responses over time. The vertical blue line indicates the start of the second session featuring new stimuli. **E.** Mean upsampled dACC BOLD signal (±SEM across participants) at the time of the outcome, split per performed action (Go/NoGo) and outcome valence (positive/negative). BOLD signal was higher after NoGo than Go responses. **F.** Same plot as (E), but split based on whether the next action was a Go (left panel) or an NoGo (right panel) response. Even if the next response was NoGo, BOLD signal was higher for trials with NoGo responses (on the current trial) than trials Go responses.

3.6.18 S3.18: Stay behavior as a function of BOLD and EEG TF power

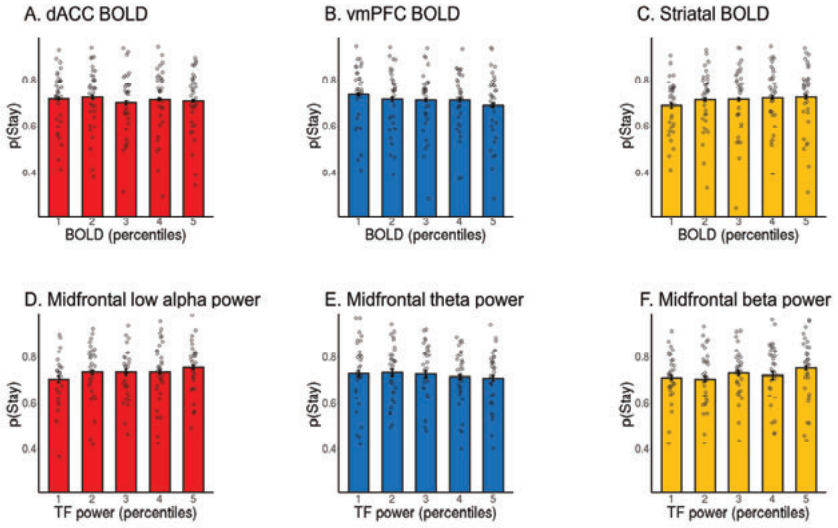


Figure 3.28. S3.18. Probability of repeating the same response ("stay") on the next cue encounter as a function of outcome-related BOLD and EEG signal.

A-C. Probability of repeating the same action ("staying") as a function of BOLD signal from (**A**) dACC, (**B**) vmPFC (cluster correlating with theta power in Fig. 3.5F), and (**C**) striatum (split into 5 bins). While dACC BOLD was not significantly linked to the probability to stay, high BOLD signal in vmPFC predicted a higher chance to switch to another action, while high BOLD signal in striatum predicted a higher probability of staying with the same action. **D-E.** Probability of staying as a function of midfrontal time-frequency power in the (**D**) low alpha, (**E**) theta/delta, and (**F**) beta range. Higher low alpha power and higher beta power predict a higher probability of staying with the same action, while higher theta power predicts a higher chance to switch to another action. Grey circles represent individual per condition-per-participant means. Error bars were very narrow (and thus hardly visible) and computed based on the Cousineau-Morey methods based on per-condition-per-participant means.

Chapter 4

Goal-directed recruitment of Pavlovian biases through selective visual attention

4 GOAL-DIRECTED RECRUITMENT OF PAVLOVIAN BIASES THROUGH SELECTIVE VISUAL ATTENTION

4.1 ABSTRACT

Prospective outcomes bias behavior in a "Pavlovian" manner: Reward prospect invigorates action, while punishment prospect suppresses it. Theories have posited Pavlovian biases as global action "priors" in unfamiliar or uncontrollable environments. However, this account fails to explain the strength of these biases—causing frequent action slips—even in well-known environments. We propose that Pavlovian control is additionally useful if flexibly recruited by instrumental control. Specifically, instrumental action plans might shape selective attention to reward/ punishment information and thus the input to Pavlovian control. In two eye-tracking samples (N = 35/ 64), we observed that Go/ NoGo action plans influenced when and for how long participants attended to reward/ punishment information, which in turn biased their responses in a Pavlovian manner. Participants with stronger attentional effects showed higher performance. Thus, humans appear to align Pavlovian control with their instrumental action plans, extending its role beyond action defaults to a powerful tool ensuring robust action execution.

4.2 INTRODUCTION

The valence of potential outcomes biases action selection: The prospect of rewards invigorates action ("Go"), while the prospect of punishment suppresses it ("NoGo"). These so-called motivational, or "Pavlovian", biases constitute a decision-making strategy that is particularly "fast-and-frugal" (Dayan et al. 2006; Boureau et al. 2015). Past theorizing has assumed that, while inflexible, these biases are fast, computationally cheap, and likely attuned to global environmental statistics (Dayan et al. 2006). They can thus act as sensible "default" response strategies in situations in which instrumental, goal-directed control fails to deliver rewards beyond chance levels, such as novel or uncontrollable environments (Daw et al. 2005; O'Doherty et al. 2017; Dorfman and Gershman 2019). These accounts assume that Pavlovian and instrumental control co-exist, largely segregated from another, and merely compete at the behavioral output level. In case of conflict, the former has to be actively suppressed—a requirement humans only imperfectly master (Breland and Breland 1961; Hershberger 1986; Cavanagh et al. 2013; Swart et al. 2018).

In contrast to such a parallel, strictly segregated architecture, we suggest that the instrumental system can adaptively recruit and steer the Pavlovian system by selecting its input via visual attention. Humans are not just passively exposed to reward and punishment cues that drive these biases. Instead, they can actively seek out or ignore these cues and thereby modulate their influence via selective visual attention ("active sensing") (Friston et al. 2010; Yang et al. 2016; Gottlieb and Oudeyer 2018). In a world full of distractions, where actions unfold over time and are prone to interference, instrumental control could harness the power of cue-driven, "automatic" behavioral tendencies by directing visual attention to cues that activate them and then automatically trigger the intended action. Such a recruitment or "training" of an inflexible decision system by a more flexible one has previously been shown in retrospective reward revaluation (Robinson and Berridge 2013; Gershman et al. 2014), credit assignment (Moran et al. 2019), and memory replay (Mattar and Daw 2018). Previous task designs measuring Pavlovian biases do not match such scenarios in which agents actively seek out information that helps them achieve their goals. We developed a new paradigm that temporally separates action selection, attention to reward and punishment information, and action execution. We then tested whether humans seek out reward and punishment information—and allow Pavlovian biases to shape responding—in a way that is aligned with their action goals. Note that, in the following, we will use the term "goal-directed" to denote such a synchronization between action goals and information search—remaining tacit about whether the underlying cognitive process involves prospective planning or devaluation sensitivity, features typically taken as indicators of "goal-directedness" of behavior(Balleine and Dickinson 1998).

Research in the past decade supports the notion that overt attention (eye gaze) towards positive aspects of choice options predicts their eventual selection (Krajbich et al. 2010; Fiedler and Glöckner 2012; Cavanagh et al. 2014), while attention to negative aspects predicts their rejection (Armel et al. 2008; Pachur et al. 2018; Westbrook et al. 2020). In these studies, positive and negative information is required for making the correct choice. Theoretical perspectives have speculated that longer attention to an option facilitates memory retrieval of its features, which could accentuate its value (Shadlen and Shohamy 2016; Weilbächer et al. 2021). However, attention to task-irrelevant positive or negative cues—which have no apparent relationship to the choice options and thus cannot serve as anchors for memory retrieval—might have similar effects.

Indeed, in Pavlovian-to-Instrumental-Transfer (PIT) paradigms, incidental background cues associated with positive/ negative outcomes induce Go/ NoGo actions (Estes 1943, 1948; Rescorla and Soloman 1967; Huys et al. 2011; Geurts et al. 2013b, 2013a). Linking those PIT effects to the role of attention in value-based choice implies that directing attention to (task-irrelevant) reward or punishment cues should activate the Pavlovian system and, in this way, automatically invigorate or suppress choice.

Beyond effects of attention on action, there is also evidence that action plans themselves can direct attention (Heuer et al. 2020; Olivers and Roelfsema 2020; van Ede 2020). Task goals modulate which stimulus features we are sensitive to and distracted by (Folk et al. 1992; Eimer and Kiss 2008; Van der Stigchel and Hollingworth 2018). "Active sensing" perspectives frame attention as a tool to actively interrogate the environment while implementing action plans (Cisek and Pastor-Bernier 2014; Yang et al. 2016; Gottlieb and Oudeyer 2018). The premotor theory of attention goes as far as proposing that the primary purpose of attention is to monitor target features relevant for preparing an action towards the target (Rizzolatti et al. 1987; Sheliga et al. 1997). Studies have indeed found perceptual sensitivity to be selectively sharpened for features relevant for an ongoing action, e.g. object location for reaching movements or object size and orientation for grasping movements (Craighero et al. 1999; Bekkering and Neggers 2002; Fagioli et al. 2007). However, in the domain of value-based decision-making, similar evidence for task goals shaping attention is scarce. One relevant finding might be that humans tend to seek out a choice option one final time before selecting it ("last fixation" or "late onset " bias) (Krajbich et al. 2010; Westbrook et al. 2020), even if they already know this option to be superior to other options (Hunt et al. 2016; Kaanders et al. 2021). In this case, attention appears to be guided by choice rather than vice versa, extending of the premotor theory of attention to value-based decision-making.

Taken together, there appear to be mechanisms synchronizing agents' attention with their action plans, and there is tentative evidence for attention to reward and punishment information triggering automatic responses in the fashion of Pavlovian biases. Hence, it seems indeed possible that an instrumental system could "recruit" the Pavlovian system to "aid" the execution of action plans by strategically steering attention toward relevant information. We tested this idea in two samples (the second one a direct, pre-registered replication) using eye-tracking. For this purpose, we designed a novel Go/ NoGo learning task in which action planning and execution were separated by a phase in which participants could preview the positive or negative outcomes at stake. Notably, information about these outcomes was not informative for the selection of the correct action. We predicted that action plans would shape attention to reward and punishment stakes, i.e., that participants' first fixation (not confounded by bottom-up saliency effects due to a gaze-contingent design) would be more often on the reward information when participants planned a Go (compared to a NoGo) action. Vice versa, we predicted an effect of attention duration to rewards vs. punishments on the final response, i.e., that longer attention to reward compared to punishment information would lead to more Go responses and speed up reaction times (Fig. 4.1A, B). Such a goal-directed recruitment of Pavlovian biases would extend their role beyond mere "default" strategies in novel environments towards a powerful aiding robust action execution.

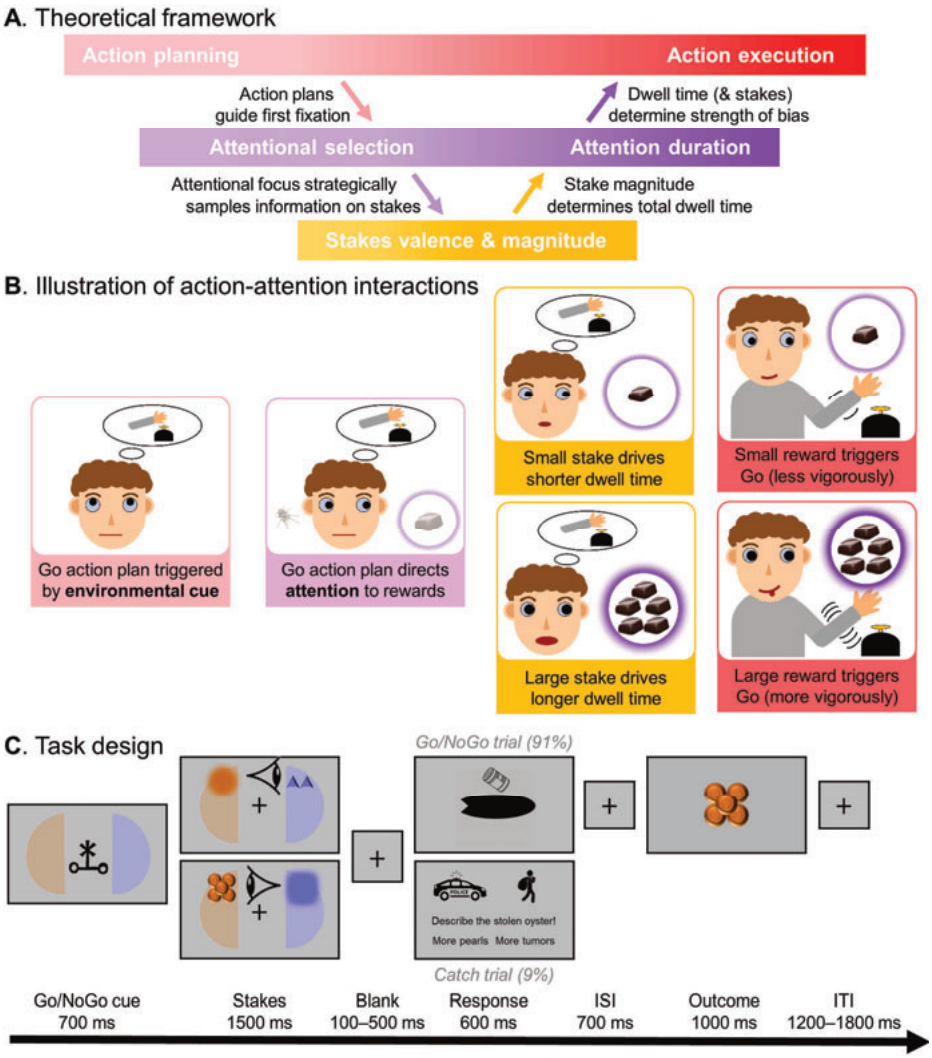


Figure 4.1. Theoretical framework and task design.
A. Theoretical framework of the interaction between action and attention. An environmental cue elicits an action plan, which directs top-down attention (first fixation) towards information about potential reward/ punishment outcomes (stakes). The first fixation anchors attention and (partly) determines which stakes will receive more attention, which is additionally modulated by bottom-up signals such as the magnitude of the stakes. The relative attention on reward versus punishment stakes (dwell time) biases the final Go/ NoGo action in a Pavlovian manner. **B.** Cartoon illustration of the proposed interaction of action planning and attention. **C.** Task design. Participants learned Go/ NoGo responses to various cues (cover story: feed/ not feed various oyster types to maximize pearls and minimize toxic tumors). Cue presentation (instructing the correct action) and action execution are separated by a phase in which rewards (pearls, here orange) and punishments (toxic tumors, here blue) at stake for correct/ incorrect responses are presented in a gaze-contingent manner. Afterwards, the oyster (black oval) can be fed, and for Go responses, participants have to press the button on the side where it is "still open". Outcomes are delivered in a probabilistic manner (75% feedback validity). On catch trials, participants have to indicate whether the oyster featured more pearls or tumors (cover story: The oyster is stolen by thieves and has to be retrieved back from the police, which requires identification).

4.3 Results

Participants (Sample 1: $N = 35$; Sample 2: $N = 64$) performed a newly developed motivational Go/ NoGo task, the Oyster Farming task" (Fig. 4.1C), while we measured their gaze using eye-tracking. In this task, participants learned whether to feed (Go) or not feed (NoGo) different oyster types (abstract stimuli) in order to maximize the chances to grow pearls (rewards) instead of tumors (punishments). Between the presentation of the action cue (oyster type) instructing the required response (Go/ NoGo) and the release cue instructing which button (left/ right) to press, participants could already inspect the potential numbers of pearls and tumors (reward and punishment stakes, ranging from 1 – 5 items) which the oyster might grow. Participants had to monitor these stakes to indicate on randomly interspersed catch trials whether the oyster featured more potential pearls or tumors. Stakes were revealed in a gaze-contingent manner, preventing bottom-up saliency effects on early peripheral vision and rendering the first fixation of each trial a measure of goal-directed attention. By separating action selection (action cue) and execution (release cue), we aimed to the test for i) an effect of the action plan (action required by the cue) on subsequent attention (first fixation, dwell time) to the reward and punishment stakes (tumors and pearls), and ii) an effect of attention to these stakes on the eventually executed response (Go/ NoGo upon the release cue).

4.3.1 Learning and Pavlovian biases

Overall, participants learned the Go/ NoGo task (% correct, Sample 1: $M = 70.0$, $SD = 10.4$, range 50.0–87.1; Sample 2: $M = 73.4$, $SD = 13.2$, range 36.3–91.7), performing significantly more Go responses to Go cues than NoGo cues (Sample 1: $b = 1.08$, $SE = 0.10$, $\chi^2(1) = 53.2$, $p < .001$; Sample 2: $b = 1.27$, $SE = 0.10$, $\chi^2(1) = 89.2$, $p < .001$; Fig. 4.2A). Participants also performed well in the catch trials (% correct: Sample 1: $M = 85.8$, $SD = 10.1$, range 56.5–100; Sample 2: $M = 86.2$, $SD = 15.5$, range 25.0–100; Fig. 4.2D). Five (seven) people in Sample 1 (2) did not perform significantly above chance (56% correct based on a 1-sided binomial test with 240 trials) in the Go/ NoGo task. In line with our pre-registration, we still included these subjects in all our analyses (for results without these participants, see S4.2). To account for variability in learning, we estimated action (Q) values for each trial based on a Rescorla-Wagner learning model.

Beyond outcome-based learning, responding was affected by the stake magnitudes in a way similar to previously observed Pavlovian biases. A more positive stake difference (reward minus punishment stake magnitude) increased the proportion of Go responses (Sample 1: $b = 0.12$, $SE = 0.03$, $\chi^2(1) = 15.3$ $p < .001$; Sample 2: $b = 0.09$, $SE = 0.03$, $\chi^2(1) = 7.9$, $p = .005$; Fig. 4.2B, C) and increased response speed (Sample 1: $b = -0.041$, $SE = 0.015$, $\chi^2(1) = 7.3$, $p = .007$; Sample 2: $b = -0.025$, $SE = 0.011$, $\chi^2(1) = 6.3$, $p = .012$). Separating these effects for the reward and punishment stakes showed that effects were driven by both valences: Higher (relative to lower) reward stake magnitude increased responding and speeded up responses, while higher (relative to lower) punishment stake magnitude decreased responding and slowed responses (see S4.3).

In sum, we found evidence that participants learned the task and that the reward and punishment stake magnitudes biased responding in opposite directions, reflecting Pavlovian biases. For reaction times, we found larger reward stake magnitudes to speed up responding and larger punishment stake magnitudes to slow down responding, again in line with Pavlovian biases as reported in previous literature (Guitart-Masip, Fuentemilla, et al. 2011; Swart et al. 2017).

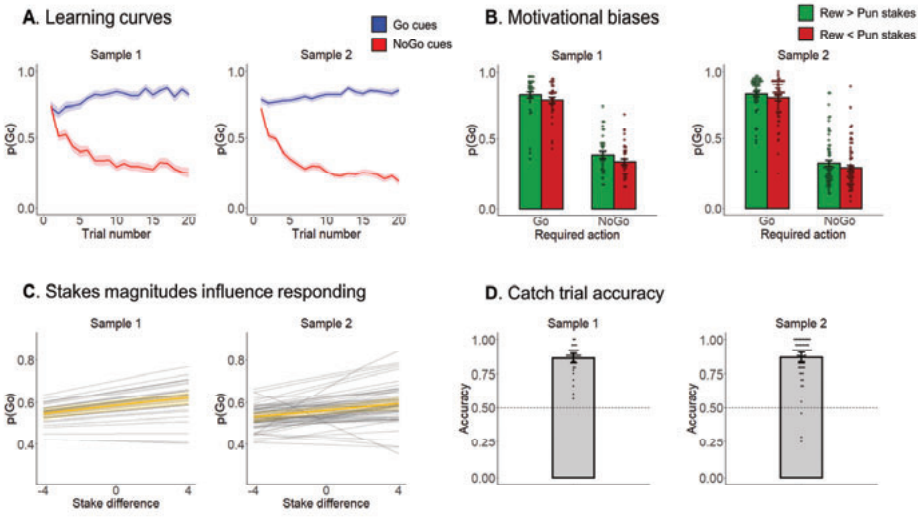


Figure 4.2. Task performance and Pavlovian biases.

A. Performance in the Pavlovian Go/ NoGo task. Trial-by-trial proportion of Go actions (±SEM) for Go cues (blue lines) and NoGo cues (red lines). Shadows indicate standard errors for per-condition-per-participant means. Participants clearly learn whether to make Go actions or not (blue lines go up; red lines go down). **B.** Pavlovian biases. Participants perform more Go responses on trials where the reward stake was higher than the punishment stake (green bars) than vice versa (red bars). Individual data points reflect response proportion per participant. **C.** Stake magnitudes biased responding in a continuous fashion. A higher stake difference (i.e., a reward stake minus punishment stake) resulted in a higher proportion of Go responses. Faint grey lines represent regression lines per participant as predicted by the mixed-effects regression model; the bronze line represents the group-level regression line; bronze shading represent mean and 95% confidence intervals. Note the two strong outliers in Sample 2; excluding these outliers did not change conclusions. **D.** Performance in the catch trials. Individual data points reflect accuracy per participant.

4.3.2 Action plans direct eye gaze

Next, we tested whether participants' attention was synchronized to their action plans. Such a link would allow Pavlovian biases to be elicited specifically by reward/ punishment cues that trigger an action in line with participants' intentions. As a measure of goal-directed attention, we used the first fixation on each trial (Konovalov and Krajbich 2016), which was unaffected by any bottom-up saliency effects of the (yet to be uncovered) stakes in our gaze-contingent design. On trials that required a Go response, participants were significantly more likely to first fixate rewards than on trials that required a NoGo response (Sample 1: $b = 0.11$, $SE = 0.04$, $\chi^2(1) = 13.9$, $p < .001$; Sample 2: $b = 0.09$, $SE = 0.03$, $\chi^2(1) = 7.9$, $p = .005$; Fig. 4.3A).

This analysis used the required response as a predictor on every trial, which is globally appropriate given that participants learnt the task. However, at the beginning of blocks, participants could not know the required response yet. Furthermore, some participants failed to learn the correct response for (some of) the cues. Thus, as a more proximate measure of participants' beliefs of what they should do, we fitted a simple Rescorla-Wagner model (Rescorla and Wagner 1972) to the Go/ NoGo response data of each participant, simulated the action (Q) values for Go and NoGo responses on each trial, and used the difference $Q_{Go} - Q_{NoGo}$ as a regressor to quantify the trial-by-trial relative value of making a Go relative to NoGo response. At the beginning of a block, this regressor will be zero, and it will stay (close to) zero in case

participants fail to learn the correct response. We found that the more Q-values favored a Go compared to a NoGo response, the more likely were participants to first fixate the reward (Sample 1: $b = 0.09$, $SE = 0.03$, $\chi^2(1) = 8.3$, $p = .004$; Sample 2: $b = 0.13$, $SE = 0.04$, $\chi^2(1) = 9.4$, $p = .002$; Fig. S4.4B).

We furthermore performed exploratory analyses to test whether action plans affect attention beyond the first fixation, i.e., also the overall difference in dwell time to the stakes (dwell time on the reward stake minus dwell time on the punishment stake). This difference was higher when the reward stake was fixated first (Sample 1: $b = 0.18$, $SE = 0.06$, $\chi^2(1) = 12.2$, $p < .001$; Sample 2: $b = 0.16$, $SE = 0.04$, $\chi^2(1) = 13.23$, $p < .001$), showing that the first fixation anchored which stakes would receive overall more attention. Over and above this effect, action value kept shaping dwell times, such that people dwelt longer on the reward (compared to the punishment) stake for Go relative to NoGo cues (Sample 1: $b = 0.03$, $SE = 0.01$, $\chi^2(1) = 4.7$, $p = .030$; Sample 2: $b = 0.03$, $SE = 0.01$, $\chi^2(1) = 13.8$, $p < .001$; Fig. S4.4C), corroborated when approximating action plans alternatively via Q-values (Sample 1: $b = 0.03$, $SE = 0.01$, $\chi^2(1) = 4.2$, $p = .040$; Sample 2: $b = 0.04$, $SE = 0.01$, $\chi^2(1) = 17.5$, $p < .001$; Fig. S4.4D). Furthermore, dwell time was influenced by the stake magnitudes, with significantly longer dwell time on the reward stake compared to the punishment stake for more positive stakes differences (Sample 1: $b = 0.09$, $SE = 0.02$, $\chi^2(1) = 16.5$, $p < .001$; Sample 2: $b = 0.12$, $SE = 0.02$, $\chi^2(1) = 41.6$, $p < .001$; see Fig. 4.3B). This latter effect shows that total dwell time was not completely determined by the first fixation, which was shaped by "top down" action values, but was additionally sensitive to bottom-up saliency effects of the stake magnitudes.

In sum, we find evidence that that participants' attention to valenced stakes information, in terms of both initial fixation and total dwell time, was synchronized to their initial action plans.

4.3.3 Eye gaze predicts responses

We next assessed whether and how attention shaped the ultimate response. We used the difference in dwell times (reward minus punishment stakes) as an integral measure of total attention (Konovalov and Krajbich 2016). We controlled for the required action to show that attention predicted the eventual response even beyond participants' likely intentions.

The longer participants attended to rewards compared to punishments, the more likely they were to make a Go response (Sample 1: $b = 0.13$, $SE = 0.03$, $\chi^2(1) = 12.2$, $p < .001$; Sample 2: $b = 0.19$, $SE = 0.03$, $\chi^2(1) = 28.4$, $p < .001$; Fig. 4.3C). Furthermore, in Sample 2 (but not Sample 1), longer attention to rewards compared to punishments lead to faster reaction times (Sample 1: $b = -0.036$, $SE = 0.026$, $\chi^2(1) = 1.9$, $p = .168$; Sample 2: $b = -0.030$, $SE = 0.012$, $\chi^2(1) = 4.5$, $p = .033$). When considered in isolation, higher dwell time on rewards increased responding, but did not significantly affect reaction times, while higher dwell time on punishment decreased responding and slowed responses (see S4.5). We did not observe any interaction effects between stakes and dwell time effects (see S4.5).

As action plans both affected attention as well the ultimate response, on might wonder if the link between attention and the ultimate response was induced by action plans as a "common cause". To exclude this possibility, all analyses using dwell times to predict responses included the required action as a regressor. Furthermore, we obtained causal evidence for an effect of attention on the ultimate response in a separate online study, in which we manipulated attention. In this

study, action cues were presented simultaneously with stakes, but located in close spatial proximity to either the reward or the punishment stakes. We reasoned that the stakes closer to the action cue would receive more attention. Indeed, we observed that action cues were located closer to reward (instead of punishment) stakes resulted in more and faster Go responses. This additional dataset corroborates a causal effect of attention on the ultimate choice. For details, see the Supplementary Materials (S4.6).

In sum, we found evidence in both samples that dwell time on rewards/ punishments drove responses towards Go/ NoGo and speeded/ slowed responses, respectively, such that attention determined the eventual strength of Pavlovian biases. Tentative evidence suggested that effects of stake magnitudes and dwell times were highly similar.

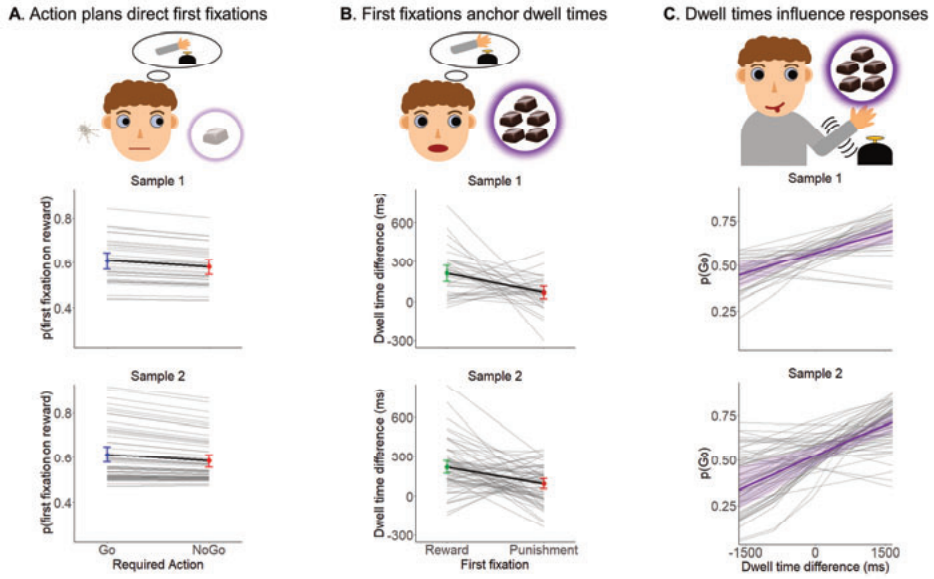


Figure 4.3. Mutual influences between action and attention.

A. Action plans direct first fixations. When required to make a Go action, participants are more likely to first fixate reward information than when a NoGo action was required. **B.** First fixation anchors attention. Dwell times are longer on reward stakes compared to punishment stakes when the first fixation was already on reward stakes. Dwell times are additionally shaped by other factors such as the stake magnitudes. **C.** Dwell time differences affect final responses. Longer attention to reward compared to punishment stakes resulted in a higher proportion of Go responses. Grey lines = regression lines per participant as predicted by the mixed-effects regression model; Black line = group-level regression line; Shading = the 95% confidence interval.

4.3.4 Stake magnitude and attentional effects differently relate to performance

Lastly, given that both stake magnitudes and dwell times affected responses and RTs in a highly similar way, we asked whether these effects also had similar consequences for participants' overall performance. Crucially, stakes were controlled by the experimental protocol and were therefore unrelated to the required response on each trial. In contrast, attention was under the control of the participant. If participants fixated reward or punishment cues in line with their action goals and then let attention guide their eventual response, strong attention effects could putatively improve their performance. We performed exploratory analyses testing whether effects of stake magnitudes and dwell times on responding were related to accuracy across participants.

The effect of stake difference on responses correlated significantly negatively with accuracy, $r(97) = -0.24$, $p = .017$ (Fig. S4.7B; after removing two outliers visible: $r(95) = -0.26$, $p = .010$; Fig. 4.4A), while the effect of dwell time difference correlated significantly positively with accuracy, $r(97) = 0.45$, $p < .001$ (Fig. 4.4B). Effects were not exclusively driven by reward or punishment stakes/ dwell times, but both (in opposite directions, respectively; see S4.7). We excluded two simpler explanations of the association between the attentional effect and task accuracy: First, this association was not driven by more accurate participants providing higher-quality eye-tracking data (see S4.7). Furthermore, accuracy was not linked to a stronger focus on reward information (i.e., more first fixation to rewards or longer attention to rewards); if anything, more accurate participants showed a more variable gaze pattern, which support the idea that these participants could rely in their responses on their gaze patterns (see S4.7).

In sum, although correlational, these results suggest that strong attentional effects might facilitate performance, while strong stake magnitudes effects impair it. Based on these analyses, stake magnitude and attentional effects appear to be dissociable.

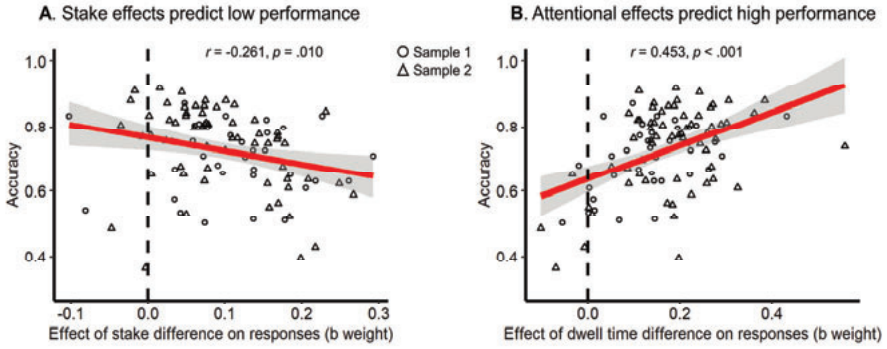


Figure 4.4. Between-subjects correlations between global Go/ NoGo task performance and stake magnitude and attentional effects. **A.** Participants with stronger effects of the stake difference on responding (i.e., steeper slopes in Fig. 4.2C) showed lower performance. **B.** Participants with stronger effects of the dwell time difference on responding (i.e., steeper slopes in Fig. 4.3C) showed higher performance. Individual data points reflect per-participants scores, the red line reflects the regression of accuracy on stake magnitude/ attentional effects (shades for ±1 SE). Points = individual participant effects, purple line = regression line, shading = ±1 SE.

4.4 Discussion

We report evidence from two independent samples showing that instrumental action plans steer attention towards rewards and punishments and in this way shape the input to the Pavlovian control system, triggering responses in line with those action plans. These results shed new light on the possible function of Pavlovian control. In contrast to current theories, we suggest that these biases have an important role beyond providing reasonable response defaults in novel or seemingly uncontrollable environments. Crucially, in addition, Pavlovian control can support instrumental control for efficient and robust action execution. In a novel task, participants successfully learned to perform Go and NoGo actions to various cues. Their responses and reaction times were biased by task-irrelevant information about potential reward/ punishment outcomes (stakes), similarly to previously reported Pavlovian biases. Most crucially, we found that participants aligned their attention to these stakes with their action plans: they paid more attention to reward stakes when they had to perform a Go action, and relatively more attention to punishment stakes when they

had to perform a NoGo action. In turn, attention to these stakes biased ultimate responses, such that more attention to rewards increased the frequency and speed of Go responding. Exploratory between-subjects analyses showed that stronger attentional effects on choice were associated with higher performance, hinting at the adaptive nature of using attention to elicit an automatic response. In sum, these results support the notion that humans can adaptively direct attention to reward and punishment information to selectively elicit Pavlovian biases in line with their action plans.

Current theories often emphasize the "hardwired" nature of Pavlovian biases (Dayan et al. 2006; Boureau et al. 2015) that allow for fast, but inflexible responding. Under the assumption that these biases embody environmental statistics on an evolutionary time scale, they should lead to the correct response in most situations. Normative models assign a dominant role to these biases in contexts that cannot be controlled (yet) by instrumental knowledge about action-outcome relationships (Dorfman and Gershman 2019). However, once an environment is controllable, biases should disappear. Frequent action slips reveal that Pavlovian biases continue to interfere with goal-directed behavior and require active suppression (Cavanagh et al. 2013; Swart et al. 2018). These cases of interference seem to question their putatively adaptive nature, warranting an update on previous theories.

Here, we suggest that a strong Pavlovian system can be adaptive, even in well-known environments, when it is actively brought into alignment with the goals of other (instrumental) systems. Pavlovian and instrumental control do not need to operate in a strict parallel fashion and merely interact at the output stage. Instead, we show that instrumental control can determine the input to Pavlovian control by selectively steering attention to (potentially unrelated) reward or punishment information. In this way, it sets the Pavlovian system on an "ballistic" track that will eventually lead to the intended response. Having such an auxiliary mechanism that will trigger the intended response might be particularly adaptive in real-life contexts in which the implementation of actions unfolds over time and is prone to interruption by distractors. By "aligning" Pavlovian with instrumental control, action selection becomes more robust against interference. Such an faciliatory effect of Pavlovian control is in line with our finding of better performance in participants with stronger attentional shaping of responses.

Beyond the context of Pavlovian biases, our results extend previous literature on the upstream determinants (rather than downstream consequences) of attention allocation. Previous studies have found that, at least early in the choice process, attention appears to be randomly allocated to choice options in a way that is independent of their value (Manohar and Husain 2013; Westbrook et al. 2020). In contrast, recent Bayesian accounts of "active sensing" have proposed that attention should be actively driven by the value and uncertainty of choice options in order to gather the maximal amount of information (Sepulveda et al. 2020; Callaway et al. 2021; Jang et al. 2021). We highlight yet another role of attention allocation: to stabilize (or even speed up) action implementation in face of delays and distraction. This role stipulates that (visual) attention is at least partly under control of ongoing motor processes—as proposed by the premotor-theory of attention (Rizzolatti et al. 1987; Sheliga et al. 1997; Olivers and Roelfsema 2020)—as well as recent accounts highlighting that vision and visual working memory primarily serve action (Heuer et al. 2020; van Ede 2020).

The idea of Pavlovian biases being recruited by instrumental action plans extends such accounts into the domain of value-based decision-making. It provides a potential explanation for why humans seek out a choice option right before selecting it, even when this will not reveal new information on what is the optimal choice (Hunt et al. 2016; Kaanders et al. 2021). Fixating an (appetitive) option might trigger Pavlovian biases that ensure its selection in face of distractors. Even more so, after participants have made the decision to select an option, its collection and consumption (potentially in face of competitors) might require further motor actions that can benefit from invigoration via these biases. Hence, the role of Pavlovian biases in invigorating motor programs might potentially explain phenomena of human (and animal) curiosity and information seeking (Vasconcelos et al. 2015; Cervera et al. 2020) even after the decision process is finished.

Our results also shed new light on the potential mechanisms by which attention to different choice options affects their eventual choices. Past research has not yet provided evidence on how fixating a choice option (e.g., a well-known food item like a Snickers) helps its evaluation or affords more information about it. Most accounts have proposed that attention towards a choice option facilitates memory retrieval of its features and in this way reduces uncertainty about its value (Shadlen and Shohamy 2016; Callaway et al. 2021). In contrast, our results suggest that attentional effects might be "Pavlovian" in nature in the sense that attending to (any) positive information disinhibits motor cortex and facilitates selection, while attending to (any) negative information inhibits motor cortex and leads to rejection. Crucially, in our paradigm, positive and negative information was unrelated (and orthogonal) to the action that needed to be selected, and thus should not be incorporated into the choice process. However, even this unrelated information did bias choice. To dissociate whether attentional effects are truly driven by increased knowledge about an option's features rather than a simple (dis-) inhibitory effect of its valence, future research should systematically manipulate the relevance of positive and negative option features to the eventual choice.

There are a few important considerations when generalizing our findings to real world situations. First, possible outcomes of a choice are often not explicitly presented to an agent. Rather, agents must make a selection among many potentially relevant pieces of information on what they deem important. Our task tried to mimic such situations by allowing agents to freely choose how much to attend to information about rewards and punishments at stake. Still, attention allocation differed from "naturalistic" free viewing settings in two important ways. Participants were not completely free to attend to the stakes, but were incentivized to do so by the secondary catch task. Furthermore, only two pieces of potential information—exemplary of positive and negative aspects of the situation—were presented, which is a drastic simplification of our information-dense environment. Future extensions of this research should provide participants with a larger set of information to select from, allowing them complete freedom to seek out any information during action preparation.

Second, in real life situations as well as in this task, people might initiate an action plan, but then change their mind. We only had access to the participants' ultimate response, which does not allow us to disentangle situations in which they maintained a determined action plan throughout the trial from situations in which actions plans were changed based on reward/ punishment information. Neuroimaging techniques with high temporal resolution such as EEG and MEG

could shed light on the dynamic interactions between motor processes and how these change as a function of attentional focus.

Third and finally, exploratory analyses suggested that participants whose ultimate response relied more strongly on attentional inputs showed higher performance. This result corroborates the postulated adaptive nature of a strong Pavlovian system that can be harnessed by instrumental systems. In contrast, the degree to which responses were shaped by the stakes magnitudes (i.e., larger magnitudes resulting in stronger Pavlovian biases) was associated with lower performance. This—at first perhaps surprising—dissociation likely arose from our task design in which stakes magnitudes were orthogonal to action requirements. When participants performed substantially above chance, stakes magnitudes had a greater potential to disturb action selection on "incongruent" trials (where the required action and the action triggered by the net stakes difference mismatched) than to facilitate it on "congruent" trials. In contrast, in many real-world contexts, it is adaptive to take into account the size of available rewards or punishments when choosing whether and how vigorously to respond.

Still, even if stakes magnitudes and attention to stakes are both meaningful contributors to choices in real-world settings, it is noteworthy that both had different consequences for performance in our task, suggestive of dissociable behavioral phenotypes. While relying on stake magnitudes might be linked to "sign-tracking" behavior previously observed in animals and humans (Flagel et al. 2009, 2010; Schad et al. 2020) and suggested to constitute a risk factor for addiction (Robinson and Berridge 1993; Garbusow et al. 2016; Chen et al. 2023), relying on attention might be a "novel" phenotype reflecting a strategic recruitment of Pavlovian biases. To conclusively demonstrate the adaptive nature of using attention to invigorate Pavlovian biases, future studies would need to causally manipulate participants' strategies. Such studies could for example train participants to strategically seek out reward or punishment information under a certain action plan. The ability to strategically up- or down-regulate Pavlovian biases could then be relevant for future interventions in psychopathologies characterized by aberrant biases, such as depression (Huys et al. 2016) or alcohol addiction (Garbusow et al. 2016; Sommer et al. 2017; Schad et al. 2020; Chen et al. 2023).

In sum, our results suggest a broadening of the current view of Pavlovian control: In addition to providing sensible "default" actions in novel or uncontrollable environments, a strong Pavlovian system can be adaptive even in well-known environments when its robust, almost "ballistic" nature is recruited to ensure that an action plan is implemented even in face of distraction.

4.5 Methods

4.5.1 Participants and exclusion criteria

In Sample 1, we recorded eye-tracking data from 35 participants (M_{age} = 23.7, SD_{age} = 4.1, range 18–35, one outlier at age 58; 27 women, 8 men; 30 right-handed; 21 with the right eye as dominant eye). In Sample 2 (replication sample), we recorded data from 64 participants (M_{age} = 21.5, SD_{age} = 3.0, range 18–34; 50 women, 13 men, 1 other; 62 right-handed; 41 with the right eye as dominant eye). In this replication sample, data collection and analyses were pre-registered (https://osf.io/nsy5x). The sample size for this sample was based on the effect size of the primary

effect of interest in Sample 1, i.e., action requirements affecting first fixations (z = 2.89, Cohen's d = 0.49), which yielded a required sample of N = 57 to detect such an effect with 95% power (two-sided one-sample t-test) (Murayama et al. 2022). We initially collected data from 57 participants, but given that seven participants did not perform significantly above chance level, we collected additional seven participants. Performance above 56% in 240 trials was significantly above chance (one-sided binomial test). Note that, in line with our pre-registration, all results in the main text are based on all participants (see S4.1 for an overview of all results); results for only those participants that performed significantly above chance are reported in S4.2 and led to identical conclusions.

Participants were recruited via the SONA Radboud Research Participation System of Radboud University. Exclusion criteria comprised glasses, color blindness, and prior treatment for neurological or psychiatric disorders. The study protocol was identical for both samples. Participants took part in a 1h session that comprised informed consent, eye-tracker calibration, a 10-minute practice phase including written instructions and practice trials, and finally the 30-minute eye-tracking experiment. Upon completion of the task, participants filled in a structured debriefing about their presumed hypothesis of the experiment, and any strategies they applied. None of the participants guessed the study hypotheses. Participants received a participation fee of €10 or 1h of course credit plus a performance dependent-bonus of €0–2 (Sample 1: M = €0.77, SD = €0.43, range €0.09–1.58; Sample 2: M = €0.91, SD = €0.47, range €0.10–1.67). Research was approved by the local ethics committee of the Faculty of Social Sciences at Radboud University (proposal no. ECSW-2018-171).

4.5.2 Apparatus

Reporting follows recently suggested guidelines for eye-tracking studies (Fiedler et al. 2020). The experiment was performed in a dimly lit, sound-attenuated room, with participants' head stabilized with a chin rest. The experimental task was coded in PsychoPy 2020.2.7 on Python 3.7.0, presented on a 24" BenQ XL2420Z screen of resolution (1920 x 1080 pixels resolution, refresh rate 144 Hz). Manual button presses were applied via a custom-made button box with two buttons (index and middle finger of the dominant hand). Participants' dominant eye was tracked with an EyeLink 1000 tracker (SR Research, Mississauga, Ontario, Canada; sampling rate of 1,000 Hz; spatial resolution of 0.01° of visual angle, monocular recording), controlled via Pylink for Python 3.7.0. The eye-tracker was placed 20 cm in front of the screen, and participants' chin rest 90 cm in front of the screen. Before the task, participants performed a 9-point calibration and validation procedure (software provided by SR Research). Calibration was repeated until an error < 1° was achieved for all points. The screen background grey tone (RGB 180, 180, 180) was constant across calibration and the experimental task.

4.5.3 Task

Participants performed a Go/ NoGo learning task with delayed response execution, called the Oyster Farming Task (Fig. 4.1C). On each trial, participants cultivated an oyster that could either grow 1–5 pearls or 1–5 hazardous tumors. Pearls gained money while tumors cost money for disposal. To maximize the probability that oysters grew pearls, participants needed to learn which oysters to "feed" (Go) and which ones not to feed ("NoGo"). Crucially, participants could choose to reveal the reward (number of pearls) and punishment (number of tumors) at stake prior to action execution in a gaze-contingent design. Participants' score of accumulated money was

turned into a bonus of 0–2€ at the end of the task. Participants performed 264 trials split into three blocks of 88 trials (80 trials of the Go/ NoGo task, 8 catch trials), each with a new set of four oyster types.

Each trial started with one (of four) abstract *action cues* (letters from the Agathodaimon alphabet; size 5.2° x 5.2°) presented for 700 ms in the center of the screen, representing an oyster type. For each oyster type, there was an optimal action (feed or not feed) that participants needed to learn by trial-and-error. Feeding was only possible when the oysters "opened" later in the trial. The optimal action led to rewards (pearls) in 75% of (valid) trials, otherwise to punishments (tumors; on "invalid trials"). Vice versa, suboptimal actions led to punishments on valid trials, but to rewards on invalid trials. During action cue presentation, participants were informed about the sides (left vs. right) on which upcoming stakes information (rewards vs. punishments) would appear via faintly colored semi-circles in the respective colors (blue and orange, counter-balanced across participants).

Directly after action cue off-set, participants were cued with the exact locations of the stakes and given 1,500 ms to unveil the tumors and pearls at stake on the respective trial. Stakes were revealed in a gaze-contingent fashion: fuzzy circular color patches appeared on the semi-circles, which changed into the number of pearls/ tumors at stake when participants fixated them. This eliminated any bottom-up saliency effects (e.g., of stake magnitude) on peripheral vision that could affect participants' first fixations. To prevent exact pre-programming of saccades, exact locations of stakes varied across trials. Stakes were located on an invisible circle with a radius of 5.2° visual angle around the screen center (i.e., distance of stakes from the center was kept constant), with a potential vertical displacement of -45 – +45 degrees from the horizontal midline. Vertical displacement was always identical for both pearls and tumors. Stakes were represented by circular areas of interest (AOI) of 150 pixels (2.7°), with a minimal distance between stakes (at maximal vertical displacement) of 514 pixels (9.4°) and a maximal distance (positioned on the horizontal midline) of 852 pixels (15.6°). Stakes were presented in orange (RGB 200, 100, 7) and blue (RGB 104, 104, 255) of equal luma. Stakes varied in magnitude (1–5 items; total display size 2.6° x 2.6°) and magnitude was balanced within action cues (i.e., each of the 20 possible combinations used once per cue, excluding the five possible combinations in which both magnitudes were identical). The mapping of pearls and tumors to the left/ right side varied across trials and was balanced within action cues (each side 10 times per cue) to control for possible participant-specific side biases in gaze.

Stakes offset was followed by a variable interval of 100–500 ms (uniform distribution in steps of 100 ms), after which a release cue (black oyster shape and a food can, 5.2° x 5.2°) appeared for 600 ms, indicating that the oyster was about to close and could be fed if necessary. The oyster remained open at either the left or right side, indicating the side where the oyster could be fed. If participants chose to feed the oyster, they had to press the respective button on the open side. The uncertainty about the response side (left/ right) at the time of the action cue, which was only resolved with the release cue, prevented pre-mature responding. In-time responses were confirmed by the food can (1.7° x 1.7°) tipping over to the respective side. 700 ms after release cue offset, the outcome (3.5° x 3.5°) was presented for 1,000 ms. Late responses during the release cue-outcome interval were recorded, but did not affect the outcome. Pressing the incorrect button (i.e., the oyster was open on the left/ right, but participants pressed the right/ left button) counted as

incorrect (i.e., yielded tumors on valid trials) and was confirmed by the can tipping over in the respective direction. Participants received a number of either pearls or tumors, depending on the stakes, their response, and trial validity. Trials finished with a variable inter-trial interval between 1,200 and 1,800 ms (uniform distribution in steps of 100 ms).

On 8 out of 88 trials per block, participants performed a catch task which incentivized attention to the stakes: Instead of the release cue, participants had to report whether the reward or punishments stakes were of greater magnitude (Fig. 4.1D). These catch trials encouraged participants to monitor both stakes and process their magnitude.

4.5.4 Preprocessing

4.5.4.1 Behavior

Catch trials were excluded from all analyses of responses and RTs. We further excluded trials with RTs below 200 ms (% trials with button presses per participant: Sample 1: $M = 0.1$, $SD = 0.3$, range 0–1.5; Sample 2: $M = 0.2$, $SD = 0.3$, range 0–1.1) because such fast responses could not be expected to incorporate processing of the cue. Likewise, we excluded trials RTs above 800 ms (% trials with button presses per participant: Sample 1: $M = 0.9$, $SD = 1.6$, range 0–6.8; Sample 2: $M = 0.5$, $SD = 1.8$, range 0–14.0). This deadline was 200 ms after release cue offset (i.e., closing of the response window) as we reasoned that any later responses could have been triggered by the release cue offset.

4.5.4.2 Eye-tracking preprocessing

Gaze data was processed in R with custom-code. Continuous data was epoched into trials of 1500 ms relative to stakes onset. Gaps of missing samples up to a duration of 75 ms (due to blinks or saccades) were interpolated using linear interpolation. Trials with more than 50% of missing samples were discarded altogether (% trials per participant: Sample 1: $M = 4.5$, $SD = 8.0$, range 0–34.1; Sample 2: $M = 3.5$, $SD = 7.9$, range 0–52.7). Gaze position was marked as being on the reward/ punishment stakes when gaze position was less than 150 pixels away from the center of the respective stakes image, which was also the criterion in our gaze-contingent design for rendering stakes visible. For each trial, we computed the first fixation on any stakes object (reward or punishment) as well as the total duration (in ms) with which rewards and punishments were fixated over the entire trial ("dwell time"). Absolute dwell times were converted into dwell time difference (reward time minus punishment time) and dwell time ratios (reward time divided by the sum of reward and punishment time) (Westbrook et al. 2020).

On some trials, participants only fixated one stake (% trials with at least one fixation per participant: Sample 1: $M = 11.0$, $SD = 14.6$, range 0–61.4; Sample 2: $M = 10.0$, $SD = 14.4$, range 0–50.4), leading to ratios of 0 or 1. We thus deviated from our pre-registration and report results for dwell time difference (reward minus punishment dwell time) in the main text, which avoids such an accumulation of values at the edges; analyses of dwell time ratio are reported in S4.1 and lead to identical conclusions.

4.5.5 Analyses

4.5.5.1 General strategy

We tested hypotheses using mixed-effects linear regression (function lmer) and logistic regression (function glmer) as implemented in the package lme4 in R (Bates et al. 2015). We used generalized linear models with a binomial link function (i.e., logistic regression) for binary dependent variables such as responses (Go vs NoGo) and first fixation, and linear models for continuous variables such as RTs or dwell time. We used zero-sum coding for categorical independent variables. All continuous dependent and independent variables were standardized such that regression weights can be interpreted as standardized regression coefficients. We added all possible random intercepts, slopes, and correlations to achieve a maximal random effects structure (Barr et al. 2013). *P*-values were computed using likelihood ratio tests with the package afex (Singmann et al. 2018). We considered *p*-values smaller than $\alpha = 0.05$ as statistically significant.

The main analyses were pre-registered for Sample 2 (replication sample; preregistration available under https://osf.io/nsy5x). We deviated from our pre-registration by reporting results based on dwell time differences (reward minus punishment dwell time) instead of dwell time ratios (reward dwell time divided by reward plus punishment dwell time) in the main text. When participants fixated only one stake, the dwell time ratios were either 0 or 1, regardless of the absolute dwell time on each single fixated option, leading to a loss of information and an accumulation of values at the edges, yielding a distribution with three modes. In contrast, dwell time differences are approximately normally distributed and statistically more comparable to stake differences. Nonetheless, analyses of dwell time ratio and dwell time differences lead to identical conclusions as reported in the Supplementary Materials (see S4.1).

4.5.5.2 Baseline learning and Pavlovian biases

First, following previously established motivational Go-NoGo learning tasks (Guitart-Masip, Fuentemilla, et al. 2011; Swart et al. 2017), we tested i) the degree to which participants learned the task, i.e., performed more Go responses to Go cues than NoGo cues, and ii) whether responses were influenced by the magnitude of the reward and punishment stakes, reflecting the presence of a Pavlovian bias. For this purpose, we fitted mixed-effects regressions with responses (Go/ NoGo) and (as secondary variable) reaction times as dependent variables and a) the required action (Go/ NoGo) as well as b) the difference in reward and punishment stake magnitude (ranging from -4 to +4) as independent variables. A significant effect of stake difference was followed up with post hoc analyses separating the effects of reward and punishment stake magnitudes, reported in the Supplementary Materials (see S4.3).

4.5.5.3 Analysis of gaze patterns

Our first key prediction was that action plans, elicited by the oyster cues, directed attention towards action-congruent stakes (reward stake for Go requirement, punishment stake for NoGo requirement). The crucial test of this prediction was whether the action requirement elicited by the cue affected the location of the first fixation (on the reward versus the punishment stake). This first fixation was not confounded by any bottom-up saliency effects since, in our gaze-contingent design, the magnitudes of the stakes was not visible yet. We used both required action (Go or NoGo) and the difference in the modeled Q-values for Go relative to NoGo responses as independent variables to predict the first fixation. These analyses also included catch trials since,

during the stakes phase, participants were unaware of whether the trial would be a Go/ NoGo or catch trial. All eye-tracking analyses contained a regressor capturing any participant-specific side biases (overall preference to fixate the left or right).

4.5.5.4Computational modelling of action values

We tested the impact of participants' action intentions on their attention towards the reward and punishment stakes using two operationalizations: Firstly, we approximated participants' intentions by the action required by the presented cue (oyster type). However, this operationalization assumes that participants (have learned and) know the required action. This assumption is violated i) at the beginning of blocks when participants cannot know the required action yet and still have to acquire it through trial-and-error, and ii) even more so in participants who fail to learn the correct response for (some of) the cues. Thus, secondly, as a more proximate measure of participants' beliefs of what they should do, we fitted a simple Rescorla-Wagner model to the Go/ NoGo response data of each participant. This model uses outcomes *r* (+1 for rewards, -1 for punishments; given that the exact outcome magnitude is irrelevant for learning) to update the action-value *Q* for the chosen action *a* towards cue *s*:

$$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \alpha * (r - Q_{t-1}(a_t, s_t)) \quad (1)$$

Action values were then translated to action probabilities using a Softmax choice rule:

$$p(Go, s_t) = \frac{\beta * e^{Q_t(Go, s_t)}}{\beta * e^{Q_t(Go, s_t)} + \beta * e^{Q_t(NoGo, s_t)}} \quad (2)$$

The model featured the free parameters α and β. The learning rate α determines the impact of prediction errors (i.e., higher α leads to stronger incorporation of recent outcomes and discounting of past outcomes). The inverse temperature β determines the stochasticity of choices (i.e., higher β leads to more deterministic choices in line with action values and lower β to more noisy, stochastic choices). Both parameters were estimated to each participants' data using a grid search, with α in the range [0, 1] in steps of 0.01 (Sample 1: $M = 0.07$, $SD = 0.08$, range 0.01–0.35; Sample 2: $M = 0.14$, $SD = 0.18$, range 0.001–0.84) and β in the range of [1, 20] in steps of 0.1 (Sample 1: $M = 7.11$, $SD = 5.87$, range 1.0–20.0; Sample 2: $M = 6.63$, $SD = 5.74$, range 1.0–19.5). Starting values for Q_{Go} and Q_{NoGo} were set to 0. Using each participants' best fitting parameters as well as their action and outcomes on each trial, we then simulated the action values for Go and NoGo responses on each trial using one-step-ahead predictions (Steingroever et al. 2014). We used the difference term $Q_{Go} - Q_{NoGo}$ as more proximate measure of participants' action intentions on each trial based on their past experience with each cue. On catch trials and trials with a response to the incorrect direction (i.e., press left when the oyster was open on the right, counting as incorrect and mostly leading to tumors), Q-values were not updated, but simply maintained from the last cue encounter.

4.5.5.5 Analysis of effects of attention on responses and reaction times

Our second key prediction was that attention to the reward and punishment stakes would shape action execution. To test this prediction, we tested whether the dwell time difference (milliseconds spent on reward stakes minus milliseconds spent attending to punishment stakes) predicted responses (Go vs. NoGo) and response speed (RT, for Go responses only). These

analyses excluded catch task trials (where responses did not relate to learning but to comparing stake magnitudes). All analyses involving responses or reaction times as dependent variable controlled for the required response as well as participant-specific side biases (overall preference to first fixate the left or right).

4.5.6 Between-subjects correlations of task accuracy

If humans synchronized their attention with their action plans such that Pavlovian biases would align with instrumental action requirements, one would expect this process to facilitate task performance and lead to higher accuracy. To test whether participants with stronger effects of attention on the final response indeed showed higher accuracy, we performed exploratory analyses by computing between-subjects correlations between overall task accuracy and i) the degree to which stake differences (reward minus punishment stake magnitude) affected responses as well as ii) the degree to which relative dwell time (reward minus punishment dwell time) affected responses. For this purpose, we refit the respective models on all participants, collapsing across both samples (total $N = 99$), and computed between-subjects correlations between participants' percent correct responses and their respective regression coefficient (fixed + random effect extracted).

4.6 SUPPLEMENTARY MATERIALS FOR CHAPTER 4

4.6.1 S4.1: Results overview full sample

Here, we report an overview over all major statistical results reported in the main text and the supplementary material. These results are based on all participants in both samples. For details on how mixed-effects regression were performed, see the Methods section of the main text.

	DV	IV	Sample	b	SE	$\chi^2(1)$	p
Task performance	Go/ NoGo	Required action	1	1.075	0.097	53.191	< .001
			2	1.265	0.091	89.190	< .001
Effect of stake valence and magnitude on action (i.e., Pavlovian bias)	Go/ NoGo	Stake difference	1	0.117	0.027	15.320	< .001
			2	0.092	0.031	7.916	.005
		Reward stake	1	0.135	0.028	20.791	< .001
			2	0.081	0.027	8.151	.004
		Punishment stake	1	-0.051	0.026	3.301	.069
			2	-0.063	0.028	4.707	.030
	RT	Stake difference	1	-0.041	0.015	7.323	.007
			2	-0.025	0.011	6.313	.012
		Reward stake	1	-0.028	0.014	3.983	.046
			2	-0.012	0.010	0.031	.861
		Punishment stake	1	0.034	0.017	4.012	.045
			2	0.029	0.011	7.311	.006
Effect of attention on action (Go/ NoGo and Go RTs)	Go/ NoGo	Dwell time difference	1	0.132	0.034	12.203	< .001
			2	0.192	0.032	28.443	< .001
		Dwell time ratio	1	0.140	0.031	15.331	< .001
			2	0.221	0.039	27.528	< .001
		Reward dwell time	1	0.035	0.034	0.945	.331
			2	0.069	0.031	4.617	.032
		Punishment dwell time	1	-0.185	0.037	18.042	< .001
			2	-0.278	0.041	35.080	< .001
		First fixation on rewards	1	-0.053	0.025	4.495	.034
			2	-0.059	0.022	7.164	.007
	RT	Dwell time difference	1	-0.036	0.026	1.900	.168
			2	-0.030	0.012	4.533	.033
		Dwell time ratio	1	-0.032	0.026	1.489	.222
			2	-0.030	0.014	4.429	.035
		Reward dwell time	1	-0.034	0.027	1.619	.203
			2	0.013	0.015	0.757	.384
		Punishment dwell time	1	0.027	0.028	0.939	.333
			2	0.039	0.013	7.668	.006
		First fixation on rewards	1	-0.010	0.016	0.255	.613
			2	0.008	0.011	0.461	.497
Effect of required action on attention (first fixation and dwell time)	First fixation	Required action	1	0.113	0.035	13.915	< .001
			2	0.090	0.028	7.882	.005
		Q-value difference	1	0.093	0.033	8.398	.004
			2	0.132	0.039	9.445	.002
	Dwell time diff.	Required action[1]	1	0.030	0.010	4.711	.030
			2	0.032	0.008	13.791	< .001
		Q-value difference[1]	1	0.031	0.010	4.213	.040
			2	0.042	0.008	17.520	< .001
	Dwell time ratio	Required action[1]	1	0.026	0.009	6.896	.009
			2	0.030	0.007	15.364	< .001
		Q-value difference[1]	1	0.016	0.011	0.951	.329
			2	0.034	0.007	13.051	< .001

[1]Controlling for first fixation and the stake difference. All effects are significant with required action/ Q-value difference as sole predictor.

4.6.2 S4.2: Results overview: Participants not above chance excluded

We report an overview over all major statistical results as reported in the main text and the supplementary material, but excluding the five (seven) participants in Sample 1 (2) that did not perform significantly above chance level, i.e., did not learn the task. For details on how mixed-effects regression were performed, see the Methods section of the main text. These analyses led to the same conclusions as the analyses based on the full samples reported in S01.

	DV	IV	Sample	b	SE	$\chi^2(1)$	p
Task performance	Response	Required action	1	1.230	0.076	68.376	< .001
			2	1.422	0.077	111.816	< .001
Effect of stake valence and magnitude on action (i.e., Pavlovian bias)	Response	Stake difference	1	0.130	0.030	14.830	< .001
			2	0.092	0.035	6.434	.011
		Reward stake	1	0.146	0.029	21.802	< .001
			2	0.078	0.030	6.072	.014
		Punishment stake	1	-0.058	0.030	3.543	.060
			2	-0.066	0.031	4.209	.040
	RT	Stake difference	1	-0.045	0.016	8.068	.005
			2	-0.031	0.013	5.828	.016
		Reward stake	1	-0.036	0.016	4.887	.027
			2	-0.015	0.011	1.208	.272
		Punishment stake	1	0.029	0.016	3.123	.077
			2	0.034	0.012	7.560	.006
Effect of attention on action (Go/ NoGo and Go RTs)	Response	Dwell time difference	1	0.142	0.037	10.442	.001
			2	0.205	0.032	30.129	< .001
		Dwell time ratio	1	0.144	0.035	12.762	< .001
			2	0.237	0.040	27.436	< .001
		Reward dwell time	1	0.033	0.040	0.593	.441
			2	0.078	0.033	5.158	.023
		Punishment dwell time	1	-0.202	0.038	19.051	< .001
			2	-0.301	0.043	35.949	< .001
		First fixation on rewards	1	-0.060	0.027	4.410	.036
			2	-0.064	0.023	7.490	.006
	RT	Dwell time difference	1	-0.009	0.026	0.122	.727
			2	-0.029	0.013	4.557	.033
		Dwell time ratio	1	-0.014	0.024	0.335	.551
			2	-0.025	0.013	3.731	.053
		Reward dwell time	1	-0.005	0.027	0.042	.838
			2	-0.016	0.016	0.977	.323
		Punishment dwell time	1	0.012	0.029	0.165	.685
			2	0.031	0.014	5.175	.023
		First fixation on rewards	1	-0.003	0.018	0.023	.881
			2	0.009	0.012	0.478	.490
Effect of required action on attention (first fixation and dwell time)	First fixation	Required action	1	0.106	0.034	9.417	.002
			2	0.097	0.032	6.955	.008
		Q-value difference	1	0.095	0.035	5.693	.017
			2	0.136	0.043	7.941	.005
	Dwell time diff.	Required action[1]	1	0.037	0.011	9.913	.002
			2	0.034	0.010	11.465	< .001
		Q-value difference[1]	1	0.028	0.012	3.965	.046
			2	0.040	0.008	15.803	< .001
	Dwell time ratio	Required action[1]	1	0.035	0.010	15.359	< .001
			2	0.032	0.008	14.013	< .001
		Q-value difference[1]	1	0.020	0.012	2.062	.151
			2	0.035	0.008	11.862	< .001

[1]Controlling for first fixation and the stake difference. All effects are significant with required action/ Q-value difference as sole predictor.

4.6.3 S4.3: Effect of stake magnitudes on responses and reaction times

Given that stake differences (reward minus punishment stake) affected both Go/ NoGo responses and reaction times, we additionally tested for separate effects of the reward and punishment stake magnitude on responses and reaction times using in mixed-effects logistic regressions (for Go/ NoGo responses) and linear regressions (for reaction times). We coded reward and punishment stake magnitudes as separate regressors (rather than their difference).

The effect of reward stake magnitude on responses was significant in both samples (Sample 1: $b = 0.14$, $SE = 0.03$, $\chi^2(1) = 20.8$, $p < .001$; Sample 2: $b = 0.08$, $SE = 0.03$, $\chi^2(1) = 8.2$, $p = .004$; Fig. S4.3A), while the effect of punishment stake magnitude was only significant in Sample 2 (Sample 1: $b = -0.05$, $SE = 0.03$, $\chi^2(1) = 3.3$, $p = .069$; Sample 2: $b = -0.06$, $SE = 0.03$, $\chi^2(1) = 4.7$, $p = .030$; Fig. S4.3B). In contrast, for RTs, higher reward stake magnitude predicted faster responses only in Sample 1 (Sample 1: $b = -0.03$, $SE = 0.01$, $\chi^2(1) = 4.0$, $p = .046$; Sample 2: $b = -0.01$, $SE = 0.01$, $\chi^2(1) = 0.03$, $p = .861$; Fig. S4.3C), while higher punishment stake magnitude consistently predicted slower responses (Sample 1: $b = 0.03$, $SE = 0.02$, $\chi^2(1) = 4.0$, $p = .045$; Sample 2: $b = 0.03$, $SE = 0.01$, $\chi^2(1) = 7.4$, $p = .006$; Fig. S4.3D). Note that RTs are only available for Go responses; hence, the amount of data (and resulting statistical power) are lower compared to the Go/ NoGo response data.

In conclusion, effects of stake magnitude on driving Pavlovian biases reported in the main manuscript were driven by variations in both the reward and the punishment stake. These effects resemble effects of Pavlovian biases reported before, but in this study emerged in a graded fashion, i.e., more and faster Go responding the larger the reward stake was, and less and slower Go responding the larger the punishment stake was.

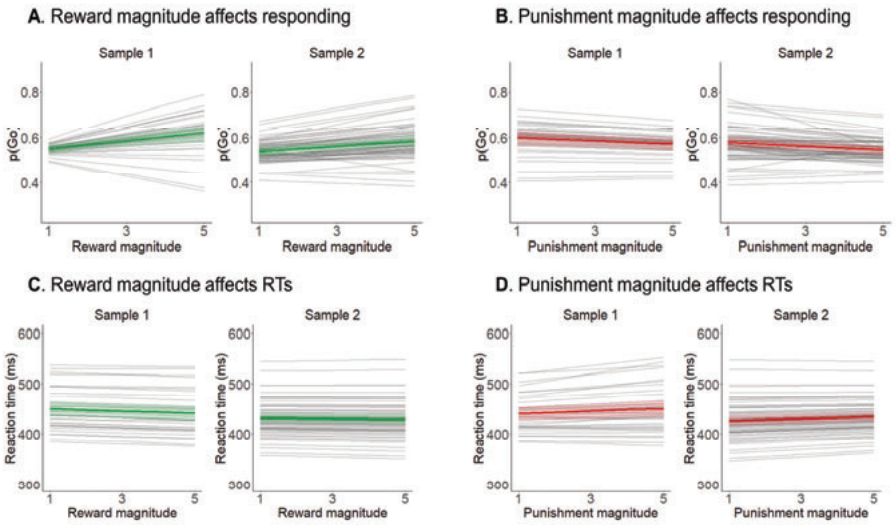


Figure 4.5. S4.3. Effect of stake magnitudes on responses and reaction times.
A higher reward stake magnitude led to a higher proportion of Go responses (**A**; significant in both studies), while a higher punishment stake magnitude led to a lower proportion of Go responses (**B**; only significant in Study 2). Similarly, a higher reward stake magnitude tended to speed up reaction times (**C**; significant only in Study 1), while a higher punishment stake magnitude tended to slow down reaction times (**D**; significant in both studies).

4.6.4 S4.4: Effect of action plans on attentional measures

As our first key prediction, we tested whether attention allocation to reward and punishment stake was affected by action requirements. For this purpose, we regressed attention measures (first fixation and dwell time difference) on participants' trial-by-trial action plans (required action and Q-value differences) using mixed-effects logistic (first fixation) and linear (dwell time difference) regression. Results are reported in the main text as well as in S01. First fixations were more likely on rewards when a Go action was required/ Q-values favored Go over NoGo. Similarly, participants looked overall longer at the reward (compared to the punishment) stake when a Go action was required/ Q-values favored Go over NoGo. Taken together, all these results suggest that attention to rewards/ punishments was synchronized to participants' action plans.

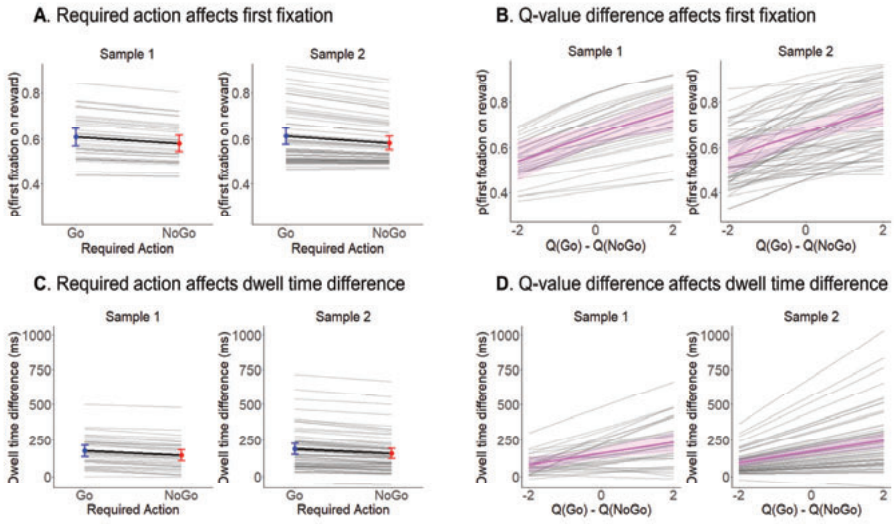


Figure 4.6. S4.4. Effect of action plans on attention measures.
Action requirements, i.e., whether participants should make a Go or a NoGo response based on the cue they see, biases participants' attention during the stakes phase: A Go compared to a NoGo requirements led to a higher proportion of first fixations on the reward stake (**A**) and longer dwell time on rewards (compared to punishments) (**C**). The same finding was obtained when fitting a Rescorla-Wagner model to participants' responses and using the Q-values based on responses from past trials to predict what participants should do on the current trial (**B** and **D**).

4.6.5 S4.5: Effect of dwell times on responses and reaction times

Given that dwell time differences (reward minus punishment dwell times) affected both Go/ NoGo responses and reaction times, we additionally tested for separate effects of reward and punishment dwell times on responses and reaction times using in mixed-effects logistic (for Go/ NoGo responses) and linear (for reaction times) regressions. Dwell time on rewards predicted a higher proportion of Go responses significantly only in Sample 2 (Sample 1: $b = 0.04$, $SE = 0.03$, $\chi^2(1) = 0.9$, $p = .331$; Sample 2: $b = 0.07$, $SE = 0.03$, $\chi^2(1) = 4.6$, $p = .032$; Fig. S4.4A). Dwell time on punishments significantly predicted a lower proportion of Go responses in both samples (Sample 1: $b = -0.19$, $SE = 0.04$, $\chi^2(1) = 18.0$, $p < .001$; Sample 2: $b = -0.28$, $SE = 0.04$, $\chi^2(1) = 35.1$, $p < .001$; Fig. S4.4B). Reward dwell time did not significantly predict RTs in either sample (Sample 1: $b = -0.03$, $SE = 0.03$, $\chi^2(1) = 1.6$, $p = .203$; Sample 2: $b = -0.01$, $SE = 0.02$, $\chi^2(1) = 0.8$, $p = .384$; Fig. S4.4C), but punishment dwell time predicted slower RTs in Sample 2 (Sample 1: $b = 0.03$, $SE = 0.03$, $\chi^2(1) = 0.9$, $p = .333$; Sample 2: $b = 0.04$, $SE = 0.01$, $\chi^2(1) = 7.7$, $p = .006$; Fig. S4.4D). Note that RTs are only available for Go responses; hence, the amount of data (and resulting statistical power) are lower compared to Go/ NoGo response data.

Interestingly, stake magnitudes and dwell times exerted highly similar effects on both responses and reaction times, with higher reward stake magnitude as well as more attention to them increased Go responding and speeded responses, while higher punishment stake magnitude as well as more attention to them decreased Go responding and slowed responses. Given that stake magnitudes and dwell times exerted such highly similar effects, one might expect them to operate through the same underlying mechanism. One consequence following from such a shared architecture is that the effects might influence each other, predicting an interaction effect. We hence performed exploratory analyses testing for such an interaction effect, reflecting whether effects of longer vs. shorter attention to the reward (punishment) stake were amplified when participants saw many vs. few potential rewards (punishments) or vice versa. The interaction between the stake difference and the dwell time difference on responses was not significant in either study (Sample 1: $b = -0.03$, $SE = 0.03$, $\chi^2(1) = 0.8$, $p = .379$; Sample 2: $b = -0.03$, $SE = 0.02$, $\chi^2(1) = 1.4$, $p = .229$), and neither was the case for RTs (Sample 1: $b = 0.04$, $SE = 0.02$, $\chi^2(1) = 2.3$, $p = .133$; Sample 2: $b = -0.003$, $SE = 0.011$, $\chi^2(1) = 0.03$, $p = .856$), providing no evidence for attention amplifying effects of stake magnitudes or vice versa.

In conclusion, longer dwell time on rewards led to more and faster responding while longer dwell time on punishments led to less and slower responding. However, effects on reaction times were only significant in the punishment domain. We did not find evidence for an interaction between stake magnitudes and dwell times, yielding no conclusive evidence whether both effects rely on the same underlying mechanism or not.

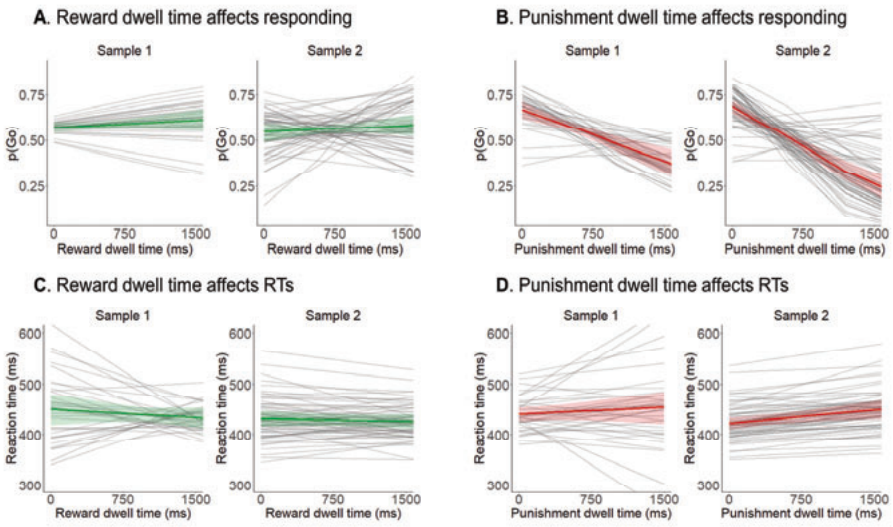


Figure 4.7. S4.5. Effect of dwell times on responses and reaction times.
Higher absolute dwell time on rewards led to a higher proportion of Go responses (**A**; only significant in Study 2), while higher absolute dwell time on punishments led to a lower proportion of Go responses (**B**; significant in both studies). Similarly, higher dwell time on rewards tended to speed up reaction times (**C**; though not significant in either study), and higher dwell time on punishment tended to slow down reaction times (**D**; only significant in Study 2).

4.6.6 S4.6: Supplementary online study manipulating attention to reward and punishment stakes

In the results from our eye-tracking studies reported in the main text, we observed an effect of (manipulated) action requirements on eye-gaze (first fixation and dwell time) and an effect of (measured) eye-gaze on the ultimate response. Given that both action requirements and eye-gaze predicted the ultimate response, on might wonder whether the link between eye-gaze and the ultimate response was spurious, induced by action plans as a "common cause" (an instance of the "third variable problem"). Note that all analyses regressing responses onto dwell time reported in the main text controlled for the action plans. In addition, we tested for a causal effect of attention to reward/ punishment information on responses in a separate online study in which we manipulated attention. This study was performed as a thesis project for Bachelor students at the beginning of the COVID-19 pandemic.

4.6.6.1 Participants and exclusion criteria

We collected data from 34 participants (M_{age} = 22.4, SD_{age} = 2.1, range 19–27; 18 female, 31 right-handed). Data collection and analyses were pre-registered (https://osf.io/kzdhm). Data was collected under a stopping rule of N = 55 as maximal sample size or May 10, 2020 as final data collection date (set by financial/ time constraints). As pre-registered, we conducted all analyses in two ways, once including all participants and once excluding participants who a) guessed the research hypotheses (zero participants) or b) did not significantly perform above chance (based on a per-participant logistic regression with response as dependent and required action as independent variable, with $p < .05$ as cut-off; three participants). Both ways led to identical conclusions.

We recruited participants via the SONA Radboud Research Participation System of Radboud University. Participants were required to be at least 18 years old, understand English at a sufficient level (self-reported), not be color-blind, perform the experiment on a PC with a keyboard (no phones or tablets) and complete the study within a maximum of 90 minutes (i.e., 1.5 times the expected completion time). The experiment was administered via the Gorilla platform (Anwyl-Irvine et al. 2021). After providing informed consent and demographic information on age, gender, and handedness, participants completed the "reversed-dot-probe" version of the Motivational Go/ NoGo Task for 30-40 minutes (see below). Afterwards, they filled out the brief (13-item) version of the Self-Control Scale (SCS) (Tangney et al. 2004) and the Behavioral Activation/ Behavioral Inhibition System Scales (BIS/BAS) (Carver and White 1994). Additionally, participants completed two vignettes (measuring omission bias) in which they rated the experienced regret and responsibility of two football coaches who won/ lost a match, afterwards changed/ kept their match plan, and then lost the next game (adapted from (Zeelenberg et al. 2002). Finally, participants performed a debriefing questionnaire asking them to a) guess the hypotheses of the experiment, b) report any (non-instructed) strategies they used, and c) guess whether the additional instructions helped them perform the task better. Participants were then debriefed about the purposes of the study. In compensation for participation, participants received 1 hour of course credit. Furthermore, participants with at least 60% accuracy in the Go/ NoGo task received tickets (proportional to performance) for a lottery featuring two 20€ gift card vouchers. Research was approved by the local ethics committee of the Faculty of Social Sciences at Radboud University (proposal no. ECSW-2018-171).

4.6.6.2 Task

Participants performed an adapted version of the Motivational Go/ NoGo learning task termed "reverse-dot-probe version" (Fig. S4.6A). On each trial, they first saw how many points they could win for a correct response (printed in green font with a "+") or lose for an incorrect response (printed in red font with a "-", termed "stakes"). Stakes varied between 10 and 90 points drawn from a uniform distribution. Reward and punishment stake were presented on the left/ right side of the screen, with positions counterbalanced across blocks. Participants were instructed to attend to the stakes because these were relevant for a catch task implemented on some of the trials (see below). After 500 ms, in addition to the stakes, one out of four action cues (letter from the Agathodaimon alphabet) appeared on the screen, which required either a Go response (space bar press) or a NoGo response (no button press). Participants had to learn the correct response from trial-and-error and respond within 1,500 ms. The action cue was presented in close proximity to either the reward stake or the punishment stake, nudging participants to direct more attention to one of the two stakes. Cue position was counterbalanced across trials and orthogonal to action requirements. After a brief fixation cross screen (700 ms), participants received the outcome (either the reward or the punishment stake previously shown) displayed for 1,500 ms. Feedback was probabilistic in that 86% (12 out of 14) trials were "valid" with a correct response winning points and an incorrect response losing points, while the remaining 14% of trials were "invalid" with a correct response losing points and an incorrect response winning points. Trials ended with a variable inter-trial interval (uniform distribution from 1,100 ms till 1,900 ms in steps of 100 ms).

On 12 trials within the first two blocks, after the outcome phase, a catch task occurred. Reward and punishment stake magnitudes were presented together with a "decoy" number (all numbers printed in white font on black boxes without +/- signs, random assignment of numbers to positions). Participants had to indicate the "other" outcome they could have received (i.e., points-to-be-won in case they lost points, points-to-be-lost in case they won points) by clicking on it with the mouse within 20 seconds. The catch task required participants to memorize the exact stake magnitudes seen earlier in the trial, incentivizing attention to them. For the latter two blocks, we did not include any catch trials to not interfere with participants applying the additional instructions (see below).

After the second block, participants received additional instructions that explicitly encouraged them to look at the reward stake in case they planned to perform a Go response, and look at the punishment stake in case they planned to perform a NoGo response. In this way, we aimed to test whether participants could voluntarily align their attention with their action plans and in this way reduce the effect of the action cue's position on responses.

Participants completed 224 trials split into four blocks à 56 trials, each blocks featuring four novel cues with 14 repetitions. Trial features (action cue identity, action requirement, stake magnitudes and positions, ITI) were controlled by one of ten pseudo-randomly drawn "spreadsheets" (preventing cue to repeated on more than two consecutive trials) randomly allocated to participants.

4.6.6.3 Data preprocessing

In line with the pre-registration, we excluded reaction times shorter than 300 ms from all analyses (as those are too fast to be induced by the presented cue). Using 200 ms as alternative cut-off (as used in our eye-tracking samples) did not change the conclusions.

4.6.6.4 Analyses

We analyzed participants' responses (Go/ NoGo) using mixed-effects logistic regression models and their reaction times using mixed-effects linear regression as implemented in the lme4 package in R (Bates et al. 2015). For all categorical independent variables, sum-to-zero coding was used. Continuous dependent and independent variables were standardized such that regression weights can be interpreted as standardized regression coefficients. We included all possible random intercepts, slopes, and correlations to achieve a maximal random effects structure (Barr et al. 2013). *P*-values were computed using likelihood ratio tests with the package afex (Singmann et al. 2018). We considered *p*-values smaller than $\alpha = 0.05$ as statistically significant.

As pre-registered (https://osf.io/kzdhm), firstly, we tested whether the action cue position (i.e., the cue being closer to the reward stake or to the punishment stake) as a proxy for participants' induced attention affect their Go/ NoGo responses and reaction times, expecting a main effect of cue position. Secondly, we tested whether instructing people to attend to stake that matched their action plan reduced the effect of cue position, expecting an interaction between cue position and instructions. We tested both hypotheses in a single model (a logistic regression model for responses, a linear regression model for reaction times) featuring the regressors required response (Go/ NoGo), cue position (on the reward/ punishment side), and instructions (before /after) as well as all possible interactions. As mentioned in the pre-registration, we also report the interaction between required action and instructions as well as the three-way interaction between required action, cue position, and instructions.

Furthermore, we specified two exploratory analyses in our pre-registration. Firstly, we tested whether the difference in stakes (reward minus punishment stake) affected participants' responses and reaction times, expecting more positive differences to lead to more and faster Go responses. For this purpose, we fitted a model with stake difference as sole regressor. Secondly, we calculated participants' mean score on the self-control scale (SCS), BIS and BAS scales and regret judgements and tested whether these scores modulated participants' cue position effect. For this purpose, we fitted a new model for each score featuring cue position, the respective score, and their interaction.

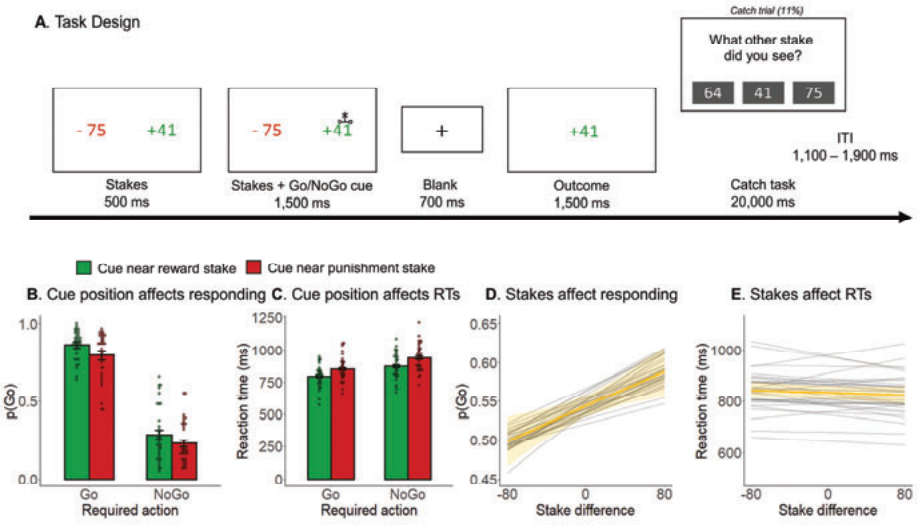


Figure 4.8. S4.6. Task design and results from the online study manipulation attention to reward and punishment information.
A. Task design. On each trial, participants saw many points they could win for correct responses or lose for incorrect responses ("stakes"). After 500 ms, a Go/ NoGo action cue was displayed either next to the reward or the punishment stake, nudging participants to direct more attention to the respective stake. Participants learned whether a cue required a Go or NoGo response from trial-and-error. Outcomes are delivered in a probabilistic manner (86% feedback validity). On catch trials, participants indicated which other stake (i.e., the one they did not receive as an outcome) they had seen before. **B.** Proportion of Go responses as a function of action requirement and cue position. Participants performed significantly more Go responses to Go cues than NoGo cues and when cues were presented next to the reward stake compared to the punishment stake. **C.** Reaction times as a function of action requirement and cue position. Participants showed significantly faster responses to Go cues than NoGo cues and when cues were presented next to the reward stake compared to the punishment stake. **D.** Proportion of Go responses as a function of stake difference (reward minus punishment stake). As net stakes became more positive, participants performed significantly more Go responses. **E.** Reaction times as a function of stake difference (reward minus punishment stake). As net stakes became more positive, participants became faster, but this effect was not significant.

4.6.6.5 Results

Overall, participants learned the Go/ NoGo task (% correct: M = 79.0, SD = 12.0, range 52.7–94.2), performing significantly more Go responses to Go cues than NoGo cues (main effect of required action: b = 1.60, SE = 0.14, $\chi^2(1) = 54.5$, $p < .001$). Three participants did not perform significantly above chance (per-participant logistic regression with response as dependent and required response as independent variable, which is significant for accuracy levels of at least 56%). In line with our pre-registration, we report results with and without these participants. Performance in the catch task was above chance (3 response options imply a chance level of 33.3%; a one-sided binomial test based on 12 trials is significant for 63% accuracy and higher) in only 25 out of 34 participants. Also, the group-level performance was hardly above chance (% correct: M = 66.4, SD = 18.6, range 25.0–81.7), likely reflecting that this task was very demanding.

Firstly, in line with our pre-registration, we tested whether the cue position (action cue on the reward/ punishment side) affected participants' Go/ NoGo responses. Participants performed more Go responses when the action cue was on the side of the reward stake compared to the side of the punishment stake (main effect of cue position: b = 0.19, SE = 0.05, $\chi^2(1) = 10.9$, $p < .001$;

Fig. S4.6B), suggesting that increased attention to rewards (compared to punishments) induced more Go responses. Similarly, participants performed faster Go responses when the action cue was on the side of the reward stake compared to the side of the punishment stake (main effect of cue position: b = -0.03, SE = 0.01, $\chi^2(1) = 25.7$, $p < .001$; Fig. S4.6C). These results suggested that more attention directed to reward/ punishment stake causally affects participants' responses and reaction times in the fashion of Pavlovian biases.

Secondly, in line with our pre-registration, we tested whether the effect of cue position became smaller after participants were instructed to attend to the stake that matched their action plan. The interaction effect between cue position and instructions was not significant (b = -0.03, SE = 0.04, $\chi^2(1) = 0.6$, $p = .458$), providing no evidence for responses becoming less affected by the cue position once participants tried to voluntarily deploy their attention. In fact, the sign of the effect suggested the effect of cue position to become stronger (instead of weaker) after additional instructions were administered. However, there was a significant interaction between required action and instructions (b = -0.38, SE = 0.07, $\chi^2(1) = 29.3$, $p < .001$), suggesting that participant overall performed better after receiving instructions. In absence of a control group, this effect cannot be disentangled from an increase in performance over time, providing inconclusive evidence for whether instructions affected participants' responses or not. The three-way interaction effect between required action, cue position, and instruction was not significant (b = 0.01, SE = 0.03, $\chi^2(1) = 1.8$, $p = .182$). Apart from responses, also the effect of cue position on reaction times was not significantly changed by instructions (b = 0.01, SE = 0.01, $\chi^2(1) = 1.7$, p = .199), and neither was the interaction between required action and instructions (b = 0.001, SE = 0.01, $\chi^2(1) = 0.04$, $p = .840$) nor the three-way interaction effect between required action, cue position, and instruction (b = -0.0003, SE = 0.005, $\chi^2(1) = 0.004$, $p = .948$) significant.

Thirdly, as part of the exploratory analyses mentioned in the pre-registration, we tested whether the difference in stakes (reward minus punishment stake) affected participants' responses or reaction times. As expected, as the difference in stakes increased (relatively more points to win than to lose), participants performed significantly more Go responses (b = 0.08, SE = 0.02, $\chi^2(1) = 8.1$, $p = .004$; Fig. S4.6D), suggesting that the difference in available rewards/ punishments biased their responses in the fashion of Pavlovian biases. Reaction times were not significantly affected by the stake difference (b = -0.004, SE = 0.004, $\chi^2(1) = 1.0$, $p = .316$; Fig. S4.6E).

Fourthly, as part of the exploratory analyses mentioned in the pre-registration, we tested whether the effect of cue position on responses was predicted by participants' score on the self-control scale (SCS), the BIS and BAS scales, or the regret and responsibility ratings in the omission bias vignettes. We did not find any significant modulation of the cue position effect by SCS scores (b = -0.03, SE = 0.03, $\chi^2(1) = 0.7$, $p = .403$), BAS Drive scores (b = -0.04, SE = 0.04, $\chi^2(1) = 1.0$, $p = .310$), BAS Reward Responsiveness scores (b = -0.01, SE = 0.03, $\chi^2(1) = 0.1$, $p = .756$), rated regret for changing the match plan after a previous football win (b = -0.02, SE = 0.04, $\chi^2(1) = 0.1$, $p = .710$), rated responsibility asymmetry when changing/ keeping the match plan after a previous football win (b = 0.02, SE = 0.03, $\chi^2(1) = 0.4$, $p = .532$), rated regret for changing the match plan after a previous football defeat (b = -0.01, SE = 0.03, $\chi^2(1) = 0.1$, $p = .750$), or rated responsibility asymmetry when changing/ keeping the match plan after a previous football defeat (b = -0.004, SE = 0.04, $\chi^2(1) = 0.01$, $p = .933$). However, the cue position effect was significantly modulated by BIS scores (b = -0.07, SE = 0.03, $\chi^2(1) = 4.3$, $p = .038$) with participants with higher BIS scores

showing weaker cue position effects, and by BAS Fun Seeking scores (b = -0.07, SE = 0.03, $\chi^2(1)$ = 4.6, p = .031) with participants with higher BAS scores showing again weaker cue position effects. Given the sample only comprised 34 participants and several between-participants analyses were run, these results should be interpreted with caution.

We repeated all analyses while excluding three participants who did not perform significantly above chance in the Go/ NoGo task. Firstly, still, participants performed more (b = 0.18, SE = 0.05, $\chi^2(1)$ = 9.0, p = .003) and faster (b = -0.04, SE = 0.01, $\chi^2(1)$ = 27.5, p < .001) Go responses when the action cue was on the side of the reward stake compared to side of the punishment stake. Secondly, the effect of cue position on responses was again not significantly different after compared to before additional instructions were administered (b = -0.02, SE = 0.04, $\chi^2(1)$ = 0.2, p = .623), but the effect of required action was again stronger after compared to before responses (b = -0.41, SE = 0.07, $\chi^2(1)$ = 23.39, p < .001), with again no significant three-way interaction (b = 0.01, SE = 0.04, $\chi^2(1)$ = 0.06, p = .800). Regarding reaction times, again, neither the effect of cue position (b = 0.01, SE = 0.01, $\chi^2(1)$ = 2.0, p = .159) nor the effect of required action (b = 0.003, SE = 0.01, $\chi^2(1)$ = 0.3, p = .597) was significantly modulated by instructions, and neither was the three-way interaction significant (b = -0.001, SE = 0.01, $\chi^2(1)$ = 0.1, p = .779). Thirdly, as the stake difference increased, participants again performed significantly more Go responses (b = 0.06, SE = 0.02, $\chi^2(1)$ = 5.7, p = .017), but not significantly faster responses (b = -0.006, SE = 0.004, $\chi^2(1)$ = 2.3, p = .127). Fourthly, we again did not find any significant modulation of the cue position effect by SCS scores (b = -0.04, SE = 0.03, $\chi^2(1)$ = 1.2, p = .264), BAS Drive scores (b = -0.02, SE = 0.04, $\chi^2(1)$ = 0.3, p = .582), BAS Reward Responsiveness scores (b = -0.01, SE = 0.04, $\chi^2(1)$ = 0.1, p = .751), rated regret for changing the match plan after a previous football win (b = 0.02, SE = 0.04, $\chi^2(1)$ = 0.3, p = .603), rated responsibility asymmetry when changing/ keeping the match plan after a previous football win (b = 0.02, SE = 0.04, $\chi^2(1)$ = 0.2, p = .632), rated regret for changing the match plan after a previous football defeat (b = -0.004, SE = 0.03, $\chi^2(1)$ = 0.1, p = .909), or rated responsibility asymmetry when changing/ keeping the match plan after a previous football defeat (b = 0.007, SE = 0.04, $\chi^2(1)$ = 0.02, p = .877). The modulation by BIS scores was not significant any more (b = -0.06, SE = -0.03, $\chi^2(1)$ = 2.33, p = .127), while the modulation by BAS Fun Seeking scores was still significant (b = -0.06, SE = 0.03, $\chi^2(1)$ = 4.2, p = .041). Overall, analyses excluding the three participants who did not perform the Go/ NoGo task significantly above chance led to identical conclusions as analyses including all participants.

4.6.6.6 Discussion

In this study, we manipulated attention by displaying Go/ NoGo action cues next to either the reward or punishment stake, nudging participants to pay relatively more attention to the stake that we next to the action cue. We obtained causal evidence that attention to reward information (compared to punishment information) leads to more Go (compared to NoGo) responses as well as to faster responses. We did not find evidence for instructions to voluntarily deploy attention in line action plans reducing the attentional effect. Potentially, the task was too demanding and the trial time course too fast for participants to voluntarily steer attention in a way that supported their action plans. Future studies might use different instructions or an altered task design that gives participants more time to deploy attention before they perform an action.

Furthermore, we found evidence for overall stake differences (reward minus punishment stake) biasing responses (but not reaction times) in the fashion of Pavlovian biases. These results

support the effect of stake differences on responses reported in the main text. Finally, we did not find any strong modulation of the attentional effect by self-reported measures such as the Self-Control Scale, the BIS/ BAS scales, or regret and responsibility ratings in two vignettes measuring omission biases. Although there was some evidence for stronger BIS and BAS Fun Seeking scores predicting weaker attention effects, these results should be treated with caution given the limited sample size and the higher number of tests. Future studies should test for such links in larger samples. In sum, the core conclusion is that the results of this study support a causal effect of attention on Go/ NoGo responses.

4.6.7 S4.7: Effects of stake magnitudes and dwell times on responses predict interindividual differences in task performance

Both stakes and dwell times affected Go/ NoGo responses (and reaction times) in a similar way, i.e., a higher reward stake as well as more attention to it increased Go responding and speeded responses, while a higher punishment stake as well as more attention to it decreased Go responding and slowed responses. Given such highly similar effects, one might expect them to operate through the same underlying mechanism. First, one consequence following from such a shared architecture is that effects should influence each other, i.e., the presence of a higher stake could alter the impact of dwell times on responses, or vice versa, which predicts an interaction effect. However, we observed no evidence for such an interaction effect (see S05), tentatively suggesting that effects operate independently of each other (though curiously with highly similar consequences).

An alternative way of assessing how comparable these effects are is to probe their consequences for task performance across participants: Does letting responses be strongly guided by stake differences (reward minus punishment stake magnitudes) vs. strongly guided by dwell time differences (reward minus punishment dwell times) have similar or different consequences for overall performance in the Go/ NoGo task? For this purpose, we re-fitted regression models across both samples, extracted per-participant regression coefficients (fixed-effect plus participant-specific random effect), and correlated these coefficients with participant overall performance (% correct responses).

Performance was significantly lower in those participants in which stake difference more strongly shaped their responses (Figure S08A, B). This finding was in stark contrast to significantly higher performance in those participants in which dwell time differences (reward minus punishment dwell time) more strongly affected response. It is noteworthy that the stake differences are experimentally controlled, and thus purely "bottom-up", while in contrast, dwell time differences were under participants' control and synchronized to action plans, both directly (effect on dwell time difference) and indirectly (effect on first fixations).

We performed control analyses to exclude the possibility that the association between attentional effects on responses and task performance was driven by better performing participants showing higher eye-tracking data quality. First, we computed the number of trials with any (opposed to no) fixation on any of the two stakes. This number was significantly positively correlated with performance, $r(97) = 0.23$, $p = .025$, but not with the attentional effect on responses, $r(97) = 0.13$, $p = .208$. When using both task performance and number of trials with any fixation to predict attention effects in a multiple linear regression, the effect of task performance was still strongly significant, $t(96) = 4.79$, $p < .001$. Second, we calculated the total

time (in ms) that people attended to any of the two stakes objects. This number was neither significantly correlated with performance, $r(97) = 0.09$, $p = .389$, nor with the attentional effect on responses, $r(97) = 0.13$, $p = .183$, and when using both task performance and total fixation time to predict attention effects in a multiple linear regression, the effect of task performance was still strongly significant, $t(96) = 4.90$, $p < .001$. In sum, it is unlikely that the correlation between performance and attentional effects on responses is driven by more accurate participants providing higher-quality eye-tracking data.

Furthermore, we performed control analyses checking whether performance, being associated with how many rewards (rather than punishments) participants received, was associated with differential fixation patterns (more first fixations or longer fixations) to reward vs. punishment stakes. It is possible that performance affects information search: high performing participants can reasonably expect to receive rewards most of the time, so they might be more interested in and attend more to reward stakes. Vice versa, lower performing participants might expect occasional punishments and thus also attend to punishment stakes. There was no significant correlation between task performance and the number of first fixations on rewards vs. punishments, $r(97) = -0.11$, $p = .298$ and the association between task performance and the attentional effect on responses remained significant when controlling for the number of first fixations, $t(96) = 4.97$, $p < .001$. There was however though a significantly negative correlation between task performance and overall dwell time difference (dwell time on reward stakes minus dwell time on punishment stakes), $r(97) = -0.27$, $p = .007$: participants with higher performance showed a more variable (i.e., less biased towards reward stakes) gaze pattern and attended relatively more to punishments compared to participants with low performance. The association between task performance and the attentional effect on responses remained significant when controlling for the this overall dwell time difference, $t(96) = 5.20$, $p < .001$. In sum, we found no evidence for high performing participants exclusively focusing on reward stakes and low performing participants also attending to punishment stakes. If anything, we found the opposite pattern of high performing participants showing a more variable gaze pattern (also attending to punishment stakes), which chimes with the idea that these participants could rely their response on their (more adaptive/ flexible) gaze pattern.

Note that all these performance-dependent results are exploratory and should be interpreted with caution.

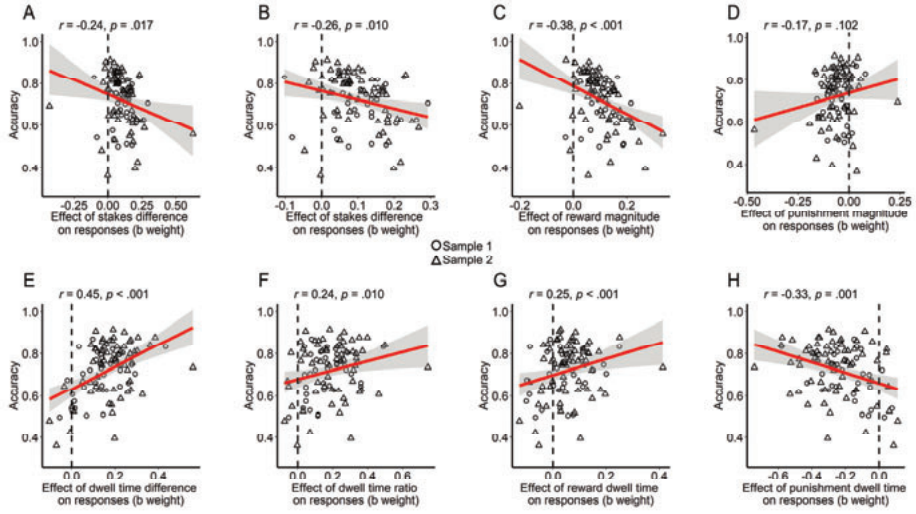


Figure 4.9. S4.7. Association between interindividual variability of accuracy and in the effects of stake magnitudes and dwell times on responses.

Participants' mean accuracy correlated significantly negatively with their respective effect of stake differences on responses (**A**), also when two outliers removed (**B**), which was driven both by a negative correlation with the effect of the reward stake (**C**; note that these effects tend to be positive) as well as a positive correlation with the effect of the punishment stake (**D**; note that these effects tend to be negative, i.e., participants with stronger negative effects showed worse performance). These correlations suggest that participants with strong stake difference effects showed poor performance. The opposite pattern occurred for the effect of dwell time on responses: This effect correlated significantly positively with accuracy, both for the difference between reward and punishment dwell times (**E**) as well as the relative dwell time (ratio) on rewards (**F**). Again, this effect was driven by reward dwell times (**G**) rather than punishment dwell times (**H**). These correlations suggest that participant with strong attention effects showed high performance.

Chapter 5

Neural mechanisms of action-attention synchronization during recruitment of Pavlovian biases

5 NEURAL MECHANISMS OF ACTION-ATTENTION SYNCHRONIZATION DURING RECRUITMENT OF PAVLOVIAN BIASES

5.1 ABSTRACT

Rewards and punishments influence action selection even when they are merely expected: Reward prospect invigorates action, while punishment prospect suppresses it, a phenomenon called "Pavlovian biases". We recently showed that humans can recruit these biases in an adaptive manner, selectively attending to reward/ punishment information that is congruent with their action plans, eliciting a bias towards the planned action. However, the neural mechanisms of how latent action plans direct attention allocation to reward and punishment information—and in this way elicit Pavlovian biases—are still unclear. In this study, we used MEG recordings, affording high temporal precision to track ongoing action preparation, to test different hypotheses on how latent action preparation—indexed by central beta power desynchronization—influences covert spatial attention—indexed by posterior alpha power lateralization. Participants performed a motivational Go/ NoGo learning task in which they could preview potential rewards or punishments ("stakes") in-between action selection and execution. We observed that stake magnitude biased participants' responses as well as ongoing beta power desynchronization in the fashion of Pavlovian biases. We did not find conclusive evidence for action plans affecting covert spatial attention to reward and punishment stakes as indexed by alpha power lateralization. However, we did find evidence for overt attention—as indexed by saccades—to reflect action plans and bias eventual responses. We discuss potential differences in how overt vs. covert attentional mechanism as well as proactive vs. reactive attentional strategies might be informed by ongoing action plans.

5.2 INTRODUCTION

Cues signaling rewards or punishments automatically invigorate or inhibit behavior (Dayan et al. 2006; Guitart-Masip, Duzel, et al. 2014; Swart et al. 2017). These *Pavlovian (or "motivational") biases* have been interpreted a as particularly "fast-and-frugal" decision heuristic, reflecting globally adaptive response strategies (O'Doherty et al. 2017). Findings that these biases are altered in psychiatric disorders such as alcohol addiction, depression, and social anxiety (Garbusow et al. 2016; Huys et al. 2016; Mkrtchian, Aylward, et al. 2017) suggest an important role for well-being and smooth goal pursuit in everyday life. Recently, we found evidence that participant are not just passively "enslaved" to these biases, but can actively recruit them to invigorate actions in line with their action plans by using selective visual attention (Algermissen and den Ouden 2022). Here, we tested through which neural mechanisms Go/ NoGo action plans affect visual attention to reward and punishment information in order to selectively recruit Pavlovian biases.

Previous research has speculated that Pavlovian biases are adaptive as they embody global environmental statistics (Dayan et al. 2006; O'Doherty et al. 2017): most rewards need to be actively acquired, and most threats can be evaded by staying still. Nonetheless, some contexts require waiting for a reward or acting to avoid a punishment. In such cases, these biases are maladaptive and need to be suppressed—an ability that humans only partially master (Cavanagh et al. 2013; Swart et al. 2018). We recently tested a different putative adaptive role that Pavlovian biases might serve (Algermissen and den Ouden 2022): While reward/ punishment cues themselves almost "ballistically" invigorate/ suppress actions, participants have control over whether they attend to and process such cues. Through directed attention, they might strategically recruit these biases based on ongoing their Go vs. NoGo action plans. Indeed, we observed that Go/ NoGo action plans led to preferential attention to reward vs. punishment information, respectively (Algermissen and den Ouden 2022). Vice versa, attention to reward cues (compared to attention to punishment cues) increased the proportion of Go responses, implying that attention to these cues modulated Pavlovian biases. These findings suggest a potential role of Pavlovian biases in supporting goal-directed action plans by seeking out information that matches those plans plan, while ignoring information that stands in conflict with it.

The neural mechanisms by which ongoing action preparation could inform visual selective attention to reward and punishment cues—and in this way set agents on a ballistic track towards the action that they are preparing—are unclear. Conversely, it is unclear through which mechanisms attention to reward and punishment cues biases ongoing action preparation. A well-known electrophysiological marker of action preparation are beta oscillations (13–33 Hz) in sensorimotor cortices: beta band oscillations desynchronize both when action plans are formed and executed (Salmelin and Hari 1994; Neuper et al. 2006; Boettcher et al. 2021) and resynchronize when action plans are prematurely cancelled before they can be implemented (Walsh et al. 2010; Gluth et al. 2013). An electrophysiological marker of (covert) visual attention are asymmetries in alpha oscillations (8–13 Hz) in occipital and parietal cortices: Alpha power decreases stronger contralaterally (relative to ipsilaterally) to an attended stimulus (Worden et al. 2000; Thut et al. 2006; Rihs et al. 2007), supposedly reflecting the disinhibition of early sensory cortices for the processing of visual input (Klimesch et al. 2007; Jensen and Mazaheri 2010). Alpha power also reflects when participants' attention is drawn to reward- or punishment-associated distractor stimuli (Marshall et al. 2018). Taken together, in this study, we tracked both central beta band power and posterior alpha band power to investigate how action preparation unfolds over time and when it affects (and, vice versa, is affected) by attention to rewards and punishments.

In the present task design, we informed participants about the location of upcoming reward and punishment information, which allowed us to dissociate proactive attention (alpha power lateralization occurring before cues appeared) from reactive attention (alpha power lateralization after cues appeared) (Geng 2014). While proactive attention requires planning even before cues appear, reactive attention means waiting until stimuli appear and then deciding where to direct the focus of attention. Furthermore, apart from directing attention, action plans could also selectively boost the early bottom-up processing of certain cues that match action plans. Hence, we also tested whether the bottom-up processing of reward and punishment stakes—indexed by gamma power increases (Marshall et al. 2018) and event-related fields (ERFs) (Hickey et al. 2010; Luque et al. 2017)—was selectively altered based on participants' action plans.

Taken together, in this study, we investigated how Go/ NoGo action preparation mechanisms influenced covert visual attention towards (as well as bottom-up processing of) reward and punishment information. Also, we investigated how, vice versa, attention influenced eventual responses. Both mechanisms are crucial for understanding how humans can potentially recruit Pavlovian biases in a way that boosts action plans. We recorded magnetencephalography (MEG) while participants performed an adapted motivational Go/ NoGo learning task in which action selection and action execution were separated. Between selection and execution, they could covertly attend to reward and punishment outcomes ("stakes") that did not confer information about the correct response. We hypothesized that action preparation—indexed by learned action values derived from a reinforcement learning model as well as by early beta power desynchronization—would influence whether participants attended more towards available rewards and punishments—indexed by alpha power lateralization, gamma power lateralization, and ERFs. Vice versa, alpha lateralization should predict both the eventual response and late beta power desynchronization facilitating its release. These hypotheses were pre-registered. Such interactions between ongoing action preparation and attention allocation could shed new light on how Pavlovian biases can be recruited in an adaptive fashion.

5.3 Results

Participants performed a motivational Go/ NoGo learning task ("Oyster Farming Task", Fig. 5.1A). In this task, action selection and action execution were separated by a phase in which they could see prospective rewards and punishments for the correct/ incorrect response, respectively. Each trial started which a cue (an oyster) that required either a Go or an NoGo response (feeding/ no feeding) to be executed at a later stage. Next, participants could pre-view the number of rewards (pearls) for a correct response and number of punishments (tumors) for an incorrect response while instructed to keep fixation at the central fixation cross. Finally, an imperative response cue appeared, instructing which button (left/ right) should be pressed to feed the oyster. Correct responses led to rewards (pearls) with 77% validity, otherwise to punishments (tumors); for incorrect responses, the opposite relationship held. A catch task implemented on some trials required participants to report whether the oyster featured more pearls or more tumors, incentivizing attention to these task features.

In addition to the Go/ NoGo task, participants performed a Posner task in which cues instructed them whether targets were to appear on the left or the right side of the screen (see S5.1). We used this task as a localizer for selecting frequencies and sensors at which power was modulated by spatial attention.

5.3.1 Behavior

Overall, participants (N = 40) learned the correct (Go/ NoGo) response for each stimulus (% correct: M = 77.3, SD = 11.7, range 49.9–92.5). Only 36 participants performed significantly above the chance level (of 56%, determined with a one-sided binomial test based on 240 trials); data from the other four participants was excluded from all analyses. On trials with Go responses, participants pressed the incorrect button (i.e., the side where the oyster was not "open") on only a minority of trials (% errors: M = 2.60, SD = 0.03, 0–17.2). In addition to performing more Go responses to Go cues than NoGo cues (b = 1.66, SE = 0.10, $\chi^2(1) = 78.9$, $p < .001$, Fig. 5.1B), participants were also faster at performing correct than incorrect Go responses (b = -0.009, SE = 0.002, $\chi^2(1) = 17.7$, $p < .001$). Most importantly, responses were biased in a Pavlovian manner by the magnitude of available reward and punishment stakes, with more Go responses (b = 0.08, SE = 0.04, $\chi^2(1) = 4.3$, $p = .039$, Fig. 5.1C, D) and faster Go responses (b = -.003, SE = 0.001, $\chi^2(1) = 3.90$, $p = .049$) on trials with relatively higher stakes (rewards minus punishments). Participants also performed well in the catch task (% correct: M = 88.5, SD = 13.7, range 41.7–100, Fig. 5.1E; 35 participants performed significantly above the chance level of 69% determined by a one-sided binomial test based on 24 trials). Taken together, participants successfully learned the task while exhibiting Pavlovian biases in both responses and reaction times. For behavioral results of the Posner task used for localizing alpha power lateralization, see S5.1.

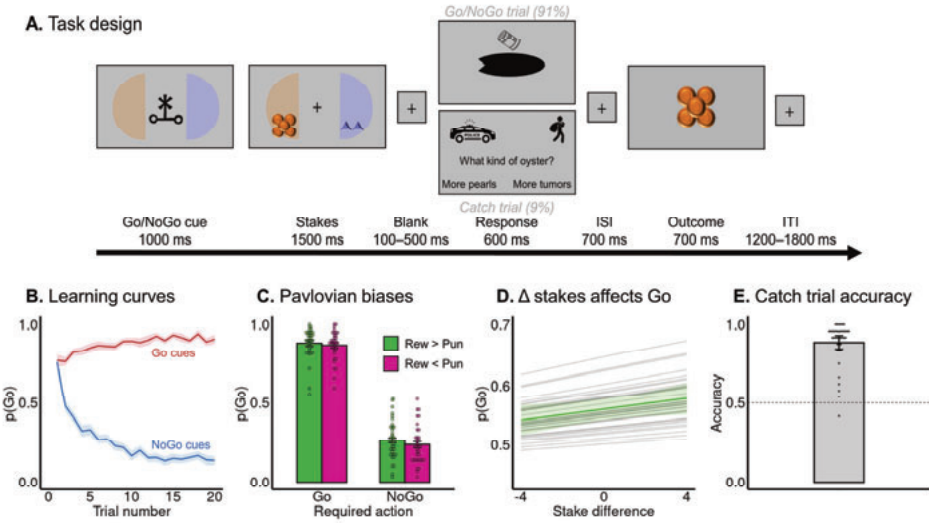


Figure 5.1. Task design and behavioral results.
A. *Task design*. Participants learned Go/ NoGo responses to various cues (cover story: feed/ not feed various oyster types to maximize pearls and minimize toxic tumors). Cue presentation (instructing the correct action) and action execution were separated by a phase in which rewards (pearls, here orange) and punishments (toxic tumors, here blue) for correct/ incorrect responses could be pre-viewed. After stakes disappeared, the oyster (black ellipse) could be fed before it closed, and for feeding it (Go responses), participants had to press the button on the side on which the oyster was still open (open end of the ellipse). Outcomes were delivered in a probabilistic manner (77% feedback validity). On catch trials, participants had to indicate whether the oyster featured more pearls or tumors (to retrieve their stolen oyster from the police department). **B.** *Performance in the Pavlovian Go/ NoGo task*. Trial-by-trial proportion of Go actions (±SEM across participants) for Go cues (red lines) and NoGo cues (blue lines). Participants clearly learned whether to make Go actions or not (red lines go up; blue lines go down). **C.** *Pavlovian biases*. Participants performed more Go responses on trials on which the reward stake was higher than the punishment stake ("Win" trials, green bars) than vice versa ("Avoid trials", pink bars). Individual data points reflect response proportion per participant. **D.** *Stake differences biased responding in a continuous fashion*. A higher stake difference (i.e., reward stake minus punishment stake) resulted in a higher proportion of Go responses. Faint grey lines represent regression lines per participant as predicted by the mixed-effects regression model; the green line represents the group-level regression line; green shading represents 95% confidence intervals. **E.** *Performance in the catch trials*. Individual data points reflect accuracy per participant. Most participants performed very well on the catch trials that required them to recall whether an oyster featured more pearls or more tumors.

5.3.2 Central beta power tracks action preparation

A large body of research has observed decreases in beta power in sensorimotor cortex when humans perform a manual action (Salmelin and Hari 1994; Neuper et al. 2006; Donner et al. 2009). What is relevant for our purposes is that beta power decreases have been observed already several seconds before an eventual response (Boettcher et al. 2021), qualifying beta power as a potential latent marker of ongoing action preparation. We first verified that beta power predicted upcoming Go/ NoGo responses also in the Oyster Farming Task, which was a pre-requisite for further hypotheses linking beta power signatures of action preparation to alpha power signatures of attention allocation.

First, we locked the signal relative to the Go/ NoGo cue and tested for differences in beta power (13–33 Hz) at left and right central sensors over the entire trial duration (0–3.5 sec. relative to cue onset), contrasting trials with Go responses to trials with NoGo responses. A cluster-based permutation test on the frequency band-averaged and sensor-averaged beta power yielded significantly lower beta power on trials on which participants performed a Go response compared

to trials on which participants performed a NoGo response. This result was driven by two clusters above threshold (Fig. 5.2A): one cluster early during the Go/ NoGo cue presentation (0.400–0.950 sec., $p = .036$) and a second cluster emerging during stakes presentation until the eventual response (1.375–3.500 sec., $p = .002$). These results indicate that already during the presentation of the Go/ NoGo cue, beta power revealed the eventual response (Go/ NoGo) that participants would perform, with lower beta power for trials with Go responses compared to trials with NoGo responses. This difference in beta power transiently disappeared at the presentation onset of the stakes, only to reappear 0.375 sec. after stakes onset up until the eventual response. After the response, beta power increased again (the "beta rebound", see S5.2), with stronger decreases and subsequent rebounds after at the hemisphere contralateral to the response hand (see S5.3). These signatures were selectively present in beta power, but not in alpha or gamma power (see S5.2, S5.3). Taken together, (only) beta power at central sensors revealed participants' latent action plans ahead of responses.

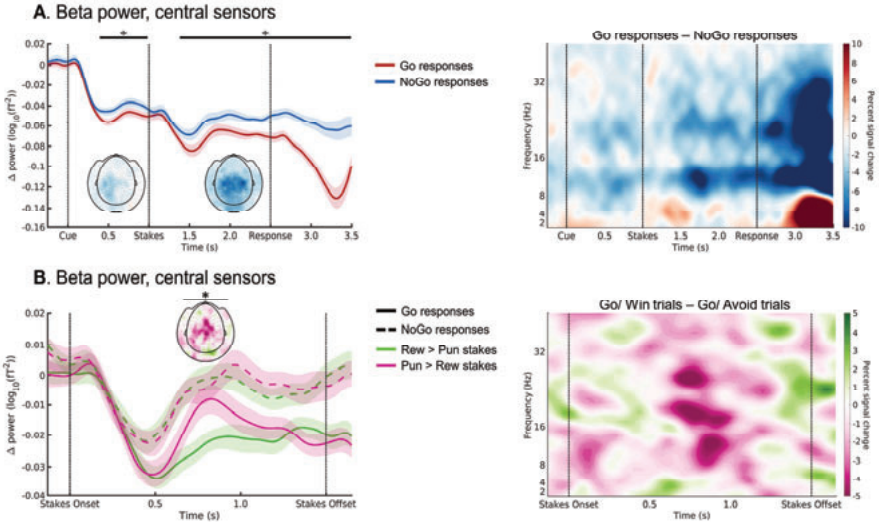


Figure 5.2. Beta power as a latent index of online action preparation.
A. *Beta power at the trial distinguishes future Go from NoGo responses.* Mean (±SEM across participants) beta power (13–33 Hz) at central sensors was lower on trials on which participants eventually performed a Go response (red lines) compared to trials with a NoGo response (blue lines). Differences emerged first around 0.400–0.950 sec. after cue onset—i.e., more than two seconds before participants could perform the response—and again from 0.375 sec. after stakes onset until the eventual response. Black horizontal lines indicate the time range for which the cluster driving significance was above threshold. **B.** *Beta power reflects influences of stakes differences on action preparation.* During stakes presentation, mean beta power was lower for Go than NoGo responses, but transiently resynchronized around 0.675–0.900 sec. on trials on which participants eventually performed a Go response, but the punishment stake exceeded the reward stake (pink lines).

5.3.3 Central beta power reflects influences of stakes on action preparation

Given that participants' propensity to perform Go responses was significantly affected by the stake differences (reward minus punishment stake), we asked whether also beta power—as a latent index of action preparation—was sensitive to stake differences. We locked data to stakes onset and contrasted (band-averaged) beta power at central sensors (sensor-averaged) in trials on which the reward stake exceeded the punishment stake ("Win trials") with trials on which the punishment stake exceeded the reward stake ("Avoid trials"). Beta power transiently re-synchronized on Avoid

(relative to Win) trials around 0.675–0.900 sec. (p = .024) (Fig. 5.2B). Visual inspection of the per-condition time courses indicated that this difference occurred on trials on which participants eventually performed a Go response. Note that this transient resynchronization did not prevent the eventual Go response. This resynchronization occurred before beta power predicted RTs (see S5.4), suggesting that it might have played a role in slowing down RTs on trials where the punishment stake was greater in magnitude, yet a Go response was executed.

5.3.4 Fronto-temporal delta power tracks reward and punishment stake magnitudes

Stake magnitudes biased participants' responses and reaction times. The influence of stakes was also reflected in the neural signal with a transient beta power resynchronization on Avoid trials. We asked next which other MEG signals encoded stake magnitudes, potentially revealing upstream regions that modulated action preparation as visible in the beta power resynchronization and slower RTs. Based on previous literature (Hunt et al. 2012), we expected such a "value" signal in lower frequencies (delta/ theta power) at anterior sensors. To test for a correlate of stake magnitudes, first, we performed a weakly constrained cluster-based permutation test over lower frequencies (1–8 Hz, averaged) at all anterior sensors (frontal, temporal, and central sensors, not averaged), contrasting Win trials with Avoid trials. Power was significantly lower for Win trials than Avoid trials (p = .036), driven by a cluster around 0.700–1.500 sec. peaking at left frontal and temporal sensors (Fig. 5.3A). Differences between these two conditions were above the cluster-forming threshold in delta power (1–4 Hz), but not in theta power. In sum, we found fronto-temporal delta power to encode the stakes (i.e., whether a trial was a "Win" or an "Avoid" trial). Notably, such a signature was not present when outcomes were presented at the end of the trial (see S5.5), suggesting differential electrophysiological signatures for expected vs. obtained outcomes.

Next, we tested whether this frontotemporal signal distinguishing Win and Avoid trials tracked stakes in a continuous fashion, i.e., kept track of the magnitudes of the stakes, or merely indexed the categorical judgment of whether the reward stake exceeded the punishment stake (i.e., the decision relevant for the catch trials). A process encoding stakes in a continuous fashion should correlate with both reward and punishment stake magnitudes, but with opposite signs (e.g., encode the reward stake positively and punishment stake negatively). To test for such continuous process, we performed a multiple linear regression across trials at each sensor, frequency, and time point separately for each participant, followed by a sign-flipping cluster-based permutation test across participants. We focused on delta power (1–4 Hz) around 0.700–1.500 sec. at left frontal and temporal sensors. We indeed observed that frontotemporal delta power correlated significantly negatively with reward magnitude (p = .010, cluster above threshold around 1.150–1.500 sec.) and significantly positively with punishment magnitude (p =.002, 0.925–1.400 sec.). These results indicated that frontotemporal delta power tracks stake differences in a continuous fashion (Fig. 5.3B), integrating both the reward and the punishment stake values.

Taken together, we found that delta power at left frontal and temporal sensors encoded stakes value parametrically, encoding reward magnitudes negatively and punishment magnitudes positively. Notably, these modulations were significant at a later time point than modulations of central beta power by the stakes (Fig. 5.2B), which questions their involvement in modifying action plans as reflected in central beta power. Note however that this value signal in frontotemporal data power was paralleled by modulations of occipital alpha (see S5.6) and gamma (see S5.7) power by the stakes around the same time. Furthermore, note that MEG correlates of value during the stakes

phase differed markedly from value correlates during the outcome phase (S5.5). In sum, stakes differences, i.e., the net "value" of a given trial, was encoded in frontotemporal delta power. These value correlates were significant at a later time point than beta power modulations by these values, questioning their putative causal role in biasing beta power and eventual responses.

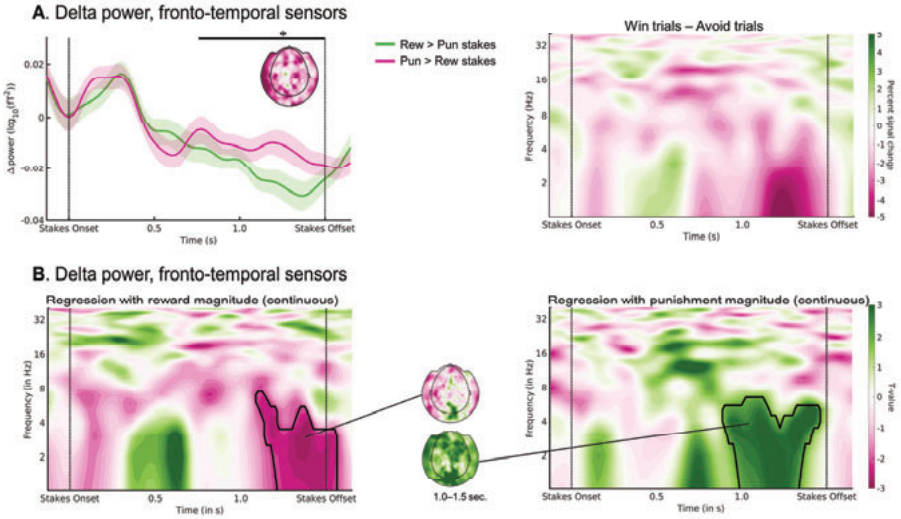


Figure 5.3. Delta power at fronto-temporal sensors reflects stake differences and magnitudes.
A. Mean (±SEM across participants) delta power at left fronto-temporal sensors distinguished Win trials (the reward stake exceeding the punishment stake) than Avoid trials (the punishment stake exceeding the reward stake). Delta power (1–4 Hz) at left fronto-temporal sensors around 0.7–1.5 sec. after stakes onset was lower on Win trials (green lines) than Avoid trials (pink lines). The black horizontal line indicates the time range for which the cluster driving significance was above threshold. **B.** Delta power at left fronto-temporal sensors correlated negatively with the reward stake and positively with the punishment stake. The same delta power signal around 0.9–1.5 sec. after stakes parametrically tracked both the reward stake magnitude (negatively) and the punishment stake positively (positively). Solid black lines indicate clusters above threshold.

5.3.5 Action plans do not modulate posterior alpha power lateralization

Next, we tested our primary hypothesis that Go/ NoGo action plans draw attention to reward/ punishment information, respectively. We tested whether alpha power (7–14 Hz) lateralization at posterior (i.e., parietal and occipital) sensors—as a latent index of covert spatial attention—was influenced by action requirements. All hypotheses were pre-registered (https://osf.io/kn7gj).

To compare left and right sensors and to enhance statistical power, we computed a trial-by-trial index of alpha power lateralization, the *attentional lateralization index* (ALI), which is the difference in power between sensors ipsilaterally to the reward stake minus power ipsilaterally to the punishment stake, divided by their sum (Thut et al. 2006; Marshall et al. 2018). Positive ALI_{α} values reflect a stronger focus on rewards (i.e., relatively lower alpha power contralateral to rewards) and negative values reflect more focus on punishments. To enhance sensitivity, each frequency bin and sensor was weighted based on the condition difference in an independent localizer (Posner) task when participants attended to a target on the left vs. on the right side of the screen (see S5.1). For robustness checks, we also performed analyses with i) a mask simply averaging over sensors and frequencies (i.e., a weight of 1 for each sensor and frequency bin), ii)

excluding participants with saccades on > 33% of trials (leaving $N = 31$; see further details in S5.8), and iii) excluding all trials with any detected saccade to one of the stakes (16% of trials).

Action plans were operationalized in two ways: first as the correct action required for the respective cue ("the ground truth"), second as the difference in Q-values ($Q_{DIF} = Q_{Go} - Q_{NoGo}$) based on a Rescorla-Wagner model fitted to participants' past choices and outcomes. The latter approach accounted for the fact that, at the beginning of each block, participants did not know the correct action for each cue and thus could not be expected to synchronize their attention to their action plans. Furthermore, the difference in Q-values reflected the degree to which participants had (not) learned the correct action for each individual cue. Both approaches were pre-registered. As a pre-registered covariate, we included the stake valence mapping (i.e., rewards appearing on the left or right) to account for participant-specific side biases in attention allocation.

First, we tested whether alpha power lateralization around 0.400–0.800 sec. after cue onset, i.e., before stakes appeared on the screen, was modulated by action plans (pre-registered hypothesis 1A). Notably, during this time, colored semi-circles in the background foreshadowed on which side the reward/ punishment stake would appear. We observed in our previous work (Algermissen and den Ouden 2022) that participants' first fixations were systematically biased towards the stake that matched their action plans, suggestive of the recruitment of proactive, anticipatory attention. Accordingly, we expected that alpha power should be lower contralaterally relative to ipsilaterally to the stakes that matched their action plan, reflective of proactive attention directed at those stakes. However, ALI_{α} before stakes onset was not significantly influenced by whether participants planned a Go or NoGo response (required action: $b = -0.0003$, $SE = 0.006$, $\chi^2(1) = 0.002$; $p = .96$; Q_{DIF}: $b = 0.005$, $SE = 0.007$, $\chi^2(1) = 0.55$; $p = .50$; Fig. 5.4A). Results were also not significant either when averaging across frequencies and sensors (required action: $p = .87$; Q_{DIF}: $p = .58$), excluding participants with saccades on > 33% of trials (required action: $p = .93$; Q_{DIF}: $p = .63$), or excluding all trials with detected saccades (required action: $p = .87$; Q_{DIF}: $p = .58$). Visual inspection of time-frequency plots (Fig. 5.5) suggested that, immediately before stakes onset, alpha power tended to transiently increase at both left and right posterior sensors without any clear difference between hemispheres. Taken together, there was no evidence for action plans influencing anticipatory covert attention—as reflected in posterior alpha power lateralization —before stakes onset.

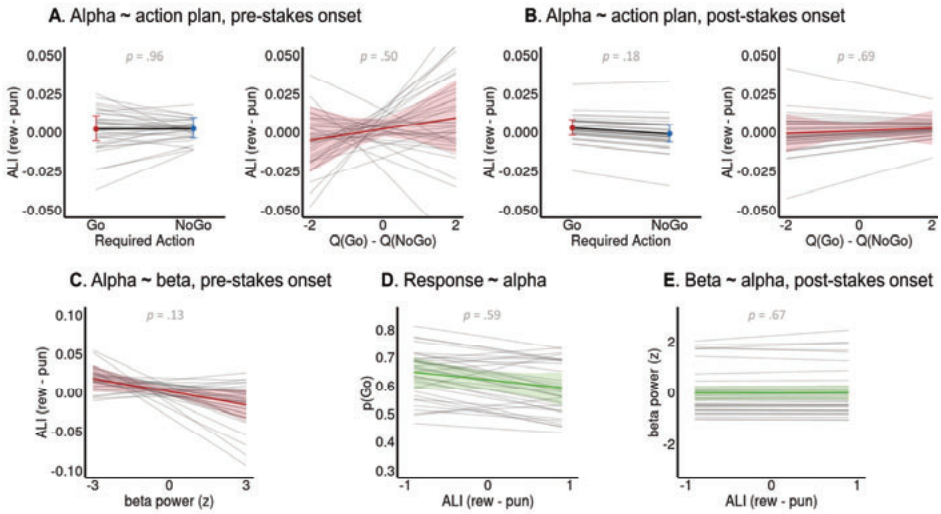


Figure 5.4. Group-level and per-participant regression lines of relationships between action selection/ beta power and attentional focus/ alpha power lateralization.
A. Neither Go/ NoGo action requirements nor the difference in action (Q) values learned from past choices and outcomes predicted posterior alpha power lateralization (parietal and occipital sensors) *before* stakes onset (attentional lateralization index, ALI: ipsilateral minus contralateral sensors relative to the location of the reward stake; positive values indicate higher focus on rewards). Faint grey lines represent regression lines per participant as predicted by the mixed-effects regression model; the colored line represents the group-level regression line; colored shading represents the 95% confidence intervals. **B.** Neither Go/ NoGo action requirements nor the difference in Q-values predicted posterior alpha power lateralization *after* stakes onset. **C.** Beta power decreases at central sensors before stakes onset, reflecting action preparation, did not significantly predict posterior alpha power lateralization before stakes onset. **D.** Posterior alpha power lateralization after stakes onset did not predict participants' eventual Go/ NoGo responses. **E.** Posterior alpha power lateralization after stakes onset did not predict beta power decreases at central sensors around the same time.

Second, we tested whether, instead of the ground-truth action requirement or the Q-value difference based on past learning history, the trial-by-trial degree of initial beta power desynchronization at central sensors, taken as an index of latent action preparation, predicted attentional focus on the reward vs. punishment stake (pre-registered hypothesis 1B). We expected that a stronger beta power desynchronization, reflecting a stronger Go (compared to NoGo) action plan, should drive attention to the reward (rather than the punishment) stake. Trial-by-trial beta power was quantified as the mean power around 0.400–0.950 sec. after cue onset (see Fig. 5.2A), with each sensor and frequency bin weighted by the strength of the respective difference between Go and NoGo trials on a group-level. ALI_{α} before stakes onset was not significantly predicted by beta power desynchronization (b = -0.012, SE = 0.007, $\chi^2(1) = 2.30$; p =.13; Fig. 5.4C). Hence, we did not find evidence for early action preparation signals influencing covert attention allocation to the expected locations of reward and punishments.

As an exploratory analysis, we tested whether alpha power not before, but *during* stakes presentation (0–1.5 sec. relative to stakes onset) was modulated by action plans. Possibly, participants did not direct their attention to the expected location of stakes in an anticipatory manner, but merely did so reactively after stakes appeared on the screen. Stakes that match action plans could preferentially capture attention. However, we found that ALI_{α} after stakes onset was

not significantly modulated by action plans (required action: b = 0.005, SE = 0.003, $\chi^2(1)$ = 1.84; p =.18; Q_{DIF}: b = 0.002, SE = 0.004, $\chi^2(1)$ = 0.16; p =.69; Fig. 5.4B). Similarly, ALI_α after stakes onset was not predicted by the degree of beta power desynchronization earlier during the cue phase (b = -0.006, SE = 0.007, $\chi^2(1)$ = 0.42; p =.52). Results were neither significant when excluding participants with saccades on > 33% of trials (required action: p =. 11; Q_{DIF}: p = .64; beta power: p = .60) nor when excluding all trials with detected saccades (required action: p =.07; Q_{DIF}: p = .56; beta power: p = .48). Visual inspection of time-frequency plots (Fig. 5.5) suggested that, after stakes onset, alpha power decreased at both left and right posterior sensors, and tended to do so slightly more strongly at right sensors when participants' action plans should have led them to focus on the left of the screen. This observation was in line with our hypothesis and confirmed by cluster-based permutation test at right posterior sensors (p = 0.004, cluster above threshold 0.400–0.775 sec.), but there was no corresponding difference at left posterior sensors. Exploratory analyses yielded that alpha power around the same time was significantly modulated by stake magnitudes (see S5.6), but effects were again restricted to right sensors. Visual inspection also suggested that lateralization was stronger at frontal and central rather than parietal and occipital sensors (Fig. 5.5). Taken together, there was no significant evidence for action plans influencing covert attention allocation to rewards and punishments once these had appeared on the screen. Visualization of condition differences suggested that modulations only occurred at right sensors and tended to be stronger at anterior rather than posterior sensors.

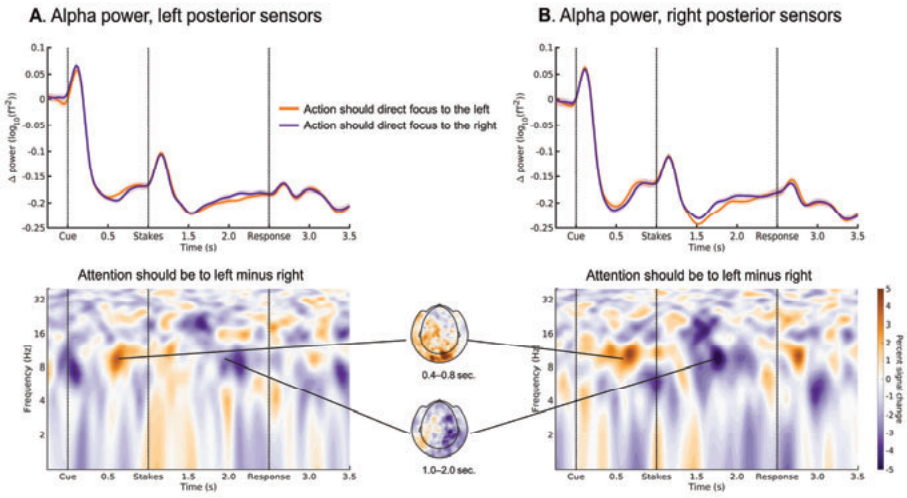


Figure 5.5. Attention to reward/ punishment stakes as a function of action plans.
Mean (±SEM across participants) alpha (8–13 Hz) power at left (A) and right (B) posterior (parietal and occipital) sensors as a function of whether, given the Go/ NoGo action requirement and rewards/ punishment stakes appearing to the left/ right of the screen, participants should attend to the left or right side of the screen. Alpha power at posterior sensors was not significantly different between trials on which action plans should have made participants attend to the left/ right of the screen, neither before nor after stakes onset. Before stakes onset, alpha power transiently resynchronized at both left and right sensors. After stakes onset, alpha power desynchronized again, which tended to be stronger at right sensors when participants were supposed to focus on the left (i.e., contralateral) side of the screen.

While action requirements did not significantly modulate alpha power either before or after stakes presentation, in contrast, there was a strong, but intricate effect of the stake valence mapping on the ALI_α. Attention to rewards as indexed by the ALI_α was stronger when rewards were presented on the right side of the screen, both before stakes onset (b = -0.61, SE = 0.11, $\chi^2(1)$ = 9.93; p = .002) and after stakes onset (b = -0.66, SE = 0.11, $\chi^2(1)$ = 24.21; p < .001). Results did not change when excluding participants with saccades on >33% of trials (before onset: $\chi^2(1)$ = 21.26, p < .001; after onset: $\chi^2(1)$ = 21.89; p < .001), or excluding all trials with detected saccades (before onset: $\chi^2(1)$ = 36.89, p < .001; after onset: $\chi^2(1)$ = 24.82; p < .001). Recoding ALI_α from a reward minus punishment contrast to a left minus right contrast indicated that the latter was systematically above zero, suggesting overall higher power at left relative to right sensors. This result is difficult to interpret. On the one hand, it could reflect the fact that participants focused more on the right side of the screen, in general (see also S5.8 for a right-ward bias in eye-movements). On the other hand, it could derive from overall lower power at right sensors. See also Fig. 5.5 as well as S5.1 and S5.6 for stronger alpha power modulations over right compared to left sensors. In sum, more variable power levels over right sensors—leading ALI_α to indicate an overall focus towards the right—might be responsible for stronger modulations of alpha power by task factors over right compared to left sensors.

Taken together, there was no conclusive evidence for the hypothesis that action plans—operationalized as action requirements, Q-value differences based on past learning, and early beta power desynchronization—affected attention allocation indexed by posterior alpha power

lateralization. See S5.6 for alpha modulation by other task factors, suggesting that any alpha power modulation occurred rather late, i.e., shortly before stakes disappeared from the screen, and were restricted to right posterior sensors.

5.3.6 Action plans do not modulate bottom-up processing of stakes

In addition to affecting attention allocation reflected in ALI, a second putative mechanism through which action plans could affect processing of stakes is through the modulation of bottom-up processing of stakes, i.e., "setting the stage" for processing of incoming stake information. As bottom-up processing of incoming sensory information is typically visible in the gamma band, we tested whether gamma power (45–65 Hz in line with) (Marshall et al. 2018) at occipital sensors was stronger contralaterally (relative to ipsilaterally) to the stakes that matched participants' action plans (preregistered hypothesis 1C). As stronger stimulus processing is typically associated with higher gamma power, we tested for higher power contralaterally compared to ipsilaterally to the focus of attention. ALI_{γ} was thus defined as power contralaterally minus ipsilaterally to rewards, with positive values again indicating stronger focus on rewards. We focused on occipital sensors because only those showed a notable gamma power increase after stakes onset.

In line with our hypothesis, there was a trend for gamma power to be higher contralaterally to stakes that matched action plans (see Fig. 5.6A), but this effect was not significant (b = 0.006, SE = 0.003, $\chi^2(1) = 3.19$; $p = .078$). The relationship with the Q-value difference was far from significant (b = 0.004, SE = 0.003, $\chi^2(1) = 1.30$; $p = .26$). Visual inspection of time frequency plot (Fig. 5.6A, B) suggested only weak modulation of gamma power, which was heterogeneous across frequency and sensors. In sum, there was no significant evidence for bottom-up processing of reward and punishment stakes—as indexed by evoked gamma power—being influenced by action plans.

As an alternative test for strengthened bottom-up processing contralaterally to action-matching stakes, we tested for modulations of the event-related fields (ERFs; pre-registered hypothesis 1D). We performed separate cluster-based permutation tests at left and right occipital sensors around 0–0.5 sec. after stakes onset, contrasting trials on which participants' action plans should have directed their focus to the left with trials on which focus should have been on the right of the screen (preregistered hypothesis 1D). There was no significant modulation, neither at left ($p = .55$) nor right ($p = 1$) occipital sensors. Visual inspection of ERF time courses (Fig. 5.6C, D) did not suggest any modulation by action plans. Taken together, there was no evidence for bottom-up processing of reward and punishment stakes—as indexed by ERFs—being influenced by action plans.

In sum, we did not find conclusive evidence for action plans—operationalized as action requirements and Q-value differences based on past learning—influencing early bottom-up processing indexed via occipital gamma power and ERFs. See S5.7 for modulations of gamma power and ERFs by other task factors; gamma power modulation was not constrained to the first 500 milliseconds, but continued until stakes disappeared.

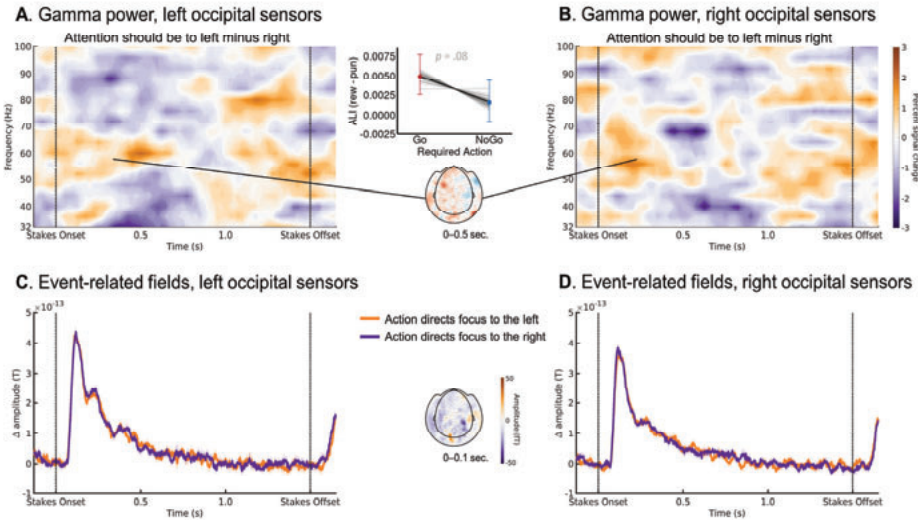


Figure 5.6. Bottom-up processing of reward/ punishment stakes as a function of action plans.
Mean (±SEM across participants) gamma (45–65 Hz) power at left (**A**) and right (**B**) occipital sensors as a function of whether, given the Go/ NoGo action requirement and reward/ punishment stakes appearing on the left/ right side of the screen, participants should attend to the left/ right of the screen. Gamma power was not significantly different between trials on which action plans should have led participants to attend to the left vs. the right side of the screen. Similarly, event-related fields (ERFs) at left (**C**) and right (**D**) occipital sensors as a function of whether, given the Go/ NoGo action requirement and reward/ punishment stakes appearing on the left/ right side of the screen, participants should have attended to the left/ right side of the screen. ERFs were not significantly different between trials on which action plans should have led participants to attend to the left vs. the right side of the screen.

5.3.7 Alpha power lateralization does not predict eventual responses

Finally, we tested whether the posterior alpha power lateralization, reflecting the relative amount of attention rewards vs. punishments received during the stakes presentation phase, predicted participants' eventual response (pre-registered hypothesis 2A). We have previously found that longer dwell times on the reward compared to the punishment stake predicted a higher propensity of eventual Go responses (Algermissen and den Ouden 2022). Similarly, we hypothesized that relatively lower alpha power contralaterally to compared to ipsilaterally to rewards, suggestive of a higher focus on the rewards, would lead participants to perform more Go responses. We controlled for both stake valence mapping and Q-value differences to test whether beta power changes predicted trial-by-trial variation in responses that was not yet captured by participants' learning history. We did not find evidence for alpha lateralization predicting eventual responses (b = -0.05, SE = 0.04, $\chi^2(1)$ = 0.29; p =.59; Fig. 5.4D), also not when excluding participants with saccades on > 33% of trials (p = .75) or excluding any trials with a detected saccade (p = .75). Hence, there was no evidence for covert attention allocation towards rewards and punishments influencing eventual responses.

Finally, we tested whether the posterior alpha power lateralization during stakes presentation predicted the degree of beta power desynchronization at central sensors at the same time (pre-registered hypothesis 2B), which might be a more sensitive measure of action invigoration than the eventual binary response. We expected that a stronger focus on the reward compared to the

punishment stake (i.e., relatively lower alpha power contralaterally to rewards) would disinhibit motor cortex and make participants more likely to eventually perform a Go response as reflected in (continuing) beta power desynchronization. We again controlled for both stake valence mapping and Q-value differences. We did not find any significant relationship between ALI_{α} and beta power desynchronization during the stakes phase required action: (b = 0.004, SE = 0.010, $\chi^2(1)$ = 0.18; p =.67; Fig. 5.4E), also not when excluding participants with saccades on > 33% of trials (p = .85) or excluding any trials with a detected saccade (p = .81). In sum, there was no evidence for covert attention to rewards and punishments influencing ongoing action preparation.

Taken together, our analyses suggested no effect of covert attention allocation to the stakes—as operationalized by posterior alpha power lateralization—on ongoing action preparation or eventual responses. Hence, covert attention might not have been sufficient to invigorate Pavlovian biases.

5.3.8 Uninstructed saccades reflect action plans and bias eventual responses

Although participants were instructed to maintain fixation at the central fixation cross throughout the trial, 27 participants showed "impulsive" saccades to the stakes on at least some trials (n = 1,576). Given null-findings in covert attention indexed in alpha power lateralization, we attempted to replicate our previous finding that overt attention as indexed by eye movements reflected action plans and biased eventual responses (Algermissen and den Ouden 2022).

First fixations were not significantly affected by Go vs NoGo action requirements (b = .09, SE = 0.09, $\chi^2(1)$ = 0.93, p = .34; Fig. 5.7A) nor by the Q-value difference (b = .07, SE = 0.08, $\chi^2(1)$ = 0.83, p = .36). However, participants tended to attend longer to the reward stake compared to the punishment stake when a Go action was required and when the Q-value difference was high, although neither effect was significant (required action: b = .04, SE = 0.02, $\chi^2(1)$ = 2.74, p = .100; Q_{DIF}: b = .02, SE = 0.02, $\chi^2(1)$ = 0.72, p = .40; Fig. 5.7B).

Finally, we tested whether participants' dwell times on reward minus punishment stakes predicted their eventual responses and reaction times (RTs). We only analyzed trials from the Go/NoGo task, discarding catch trials. The difference in dwell time (reward minus punishments) significantly predicted eventual responses (b = .17, SE = 0.06, $\chi^2(1)$ = 4.49, p = .034; Fig. 5.7C), with more responses after relatively longer dwell times on rewards compared to punishments. There was no significant relationship between dwell time differences and RTs of eventual responses (b = .004, SE = 0.004, $\chi^2(1)$ = 0.54, p = .46; Fig. 5.7D).

In sum, analyses of impulsive saccades on a subset of trials suggested that overt attention—operationalized as total dwell time on the reward minus the punishment stake—tended to be influenced by action plans. Also, dwell time differences significantly predicted eventual responses. Unlike our previous eye-tracking study, participants were instructed to keep fixation at the central fixation cross, resulting in only relatively few saccades that were available for analysis. Thus, these analyses had substantially lower statistical power compared to our previous eye-tracking study.

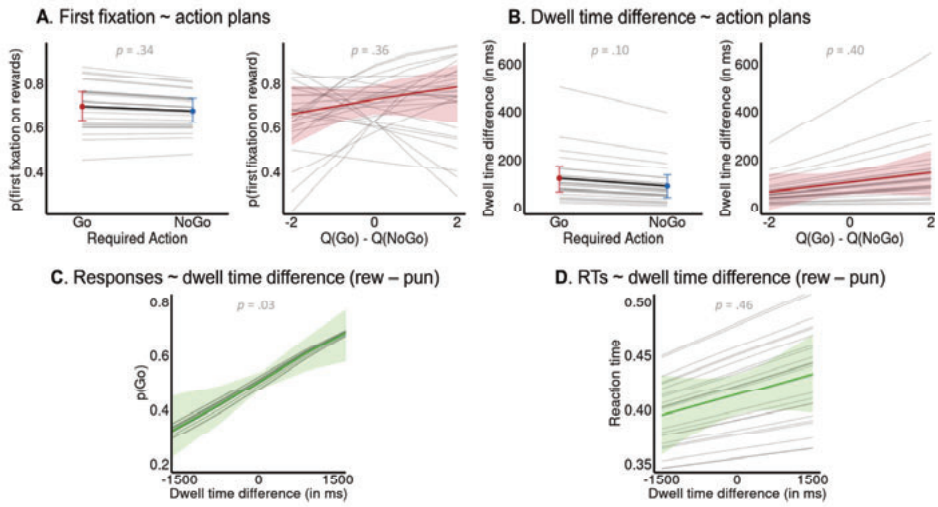


Figure 5.7. Effects of action plans on impulsive saccades and converse effects of saccade dwell time on eventual responses.
A. Action plans did not direct first fixations. Faint grey lines represent regression lines per participant as predicted by the mixed-effects regression model; the colored line represents the group-level regression line; colored shading represent mean and 95% confidence intervals. **B.** Action plans did not affect dwell time differences. While dwell times tended to be longer for rewards when a Go action was required and Q-values favored Go over NoGo actions, these effects were not significant. **C.** Dwell time differences predicted eventual responses. The longer participants fixated the reward stake (compared to the punishment stake), the more likely were they to perform a Go response. **D.** Dwell time differences did not predict RTs.

5.4 Discussion

In this study, we investigated neural mechanisms of how instrumental action plans can selective recruit Pavlovian biases via visual selective attention. Participants performed a Go/ NoGo learning task in which action selection and execution were separated by a phase in which potential reward/ punishment outcomes ("stakes") could be pre-viewed. We expected that Go/ NoGo action plans would direct visual attention to stakes that "matched" their action plan such as to set participants on a ballistic track towards their planned response, guided by the Pavlovian system. Specifically, we expected more attention to the reward stake under a Go action plan and more attention to the punishment stake under a NoGo action plan. We operationalized participants' action plans via trial-by-trial central beta power desynchronization and their covert attention via posterior alpha power lateralization. Vice versa, we expected that the relative attention to the reward vs. the punishment stake would affect the eventual response, invigorating participants' original action plan in the fashion of Pavlovian biases.

Behaviorally, the net difference between the reward and punishment stake biased Go/ NoGo action selection in the fashion of Pavlovian biases. Neurally, beta power at central sensors, taken as an index of latent action preparation, revealed participants' eventual response already several seconds before response onset. Beta power desynchronization was sensitive to stake differences, resynchronizing when stake magnitudes biased action preparation. Net stakes were parametrically encoded in delta power at left frontotemporal sensors, but these correlations were significant at a later time point than the beta power modulation by the stakes. While stake magnitudes influenced both eventual actions and MEG correlates of action preparation, there was no conclusive evidence

for either i) covert attention to these stakes being synchronized to participants' action plans or ii) covert attention influencing eventual actions. Participants' action plans—both indexed via their past learning history and via early beta central power synchronization during Go/ NoGo cues—did not significantly predict the amount of attention participants directed at the reward vs. the punishment stake—as indexed by posterior alpha power lateralization. Such a link between action preparation and attention was absent both before and during stakes presentation. Apart from alpha lateralization, gamma power lateralization evoked by the stakes suggested a numerically (though not significantly) higher focus on stakes that matched action plans. There was no evidence for attention alpha power lateralization biasing eventual responses, i.e., neither participants' eventual action nor beta power desynchronization. In contrast to alpha power lateralization, analyses of saccades (made in violation of the instruction to maintain central fixation) showed that participants tended to attend more to stakes that action plans. Also, dwell time on these stakes significantly predicted their eventual responses. Taken together, we did not find conclusive evidence for action plans affecting covert attention to reward and punishment stakes in a way that would strategically invigorate Pavlovian biases. However, there was some evidence for overt attention as indexed by eye-movements reflecting action plans and biasing eventual actions.

5.4.1 Stakes modulate beta power desynchronization

Beta power desynchronization as a latent index of action preparation was sensitive to the stake magnitudes. While beta power desynchronized in preparation of Go responses over the time course of the trial, it showed a transient resynchronization around 0.7–0.9 sec. after stakes onset when punishment stakes exceeded reward stakes. This transient resynchronization even occurred on trials on which participants eventually performed a Go response (and beta power desynchronized again). Notably, this modulation by stakes occurred before the time point where beta power predicted participants' reaction times. This observation is in line with the idea that, beyond modulating Go/ NoGo responses, stakes might have modulated beta power in a way that also slowed down reaction times. In sum, these results are indicative of the time point when action preparation—as indexed by beta power desynchronization—is affected by the stake magnitudes.

5.4.2 Frontotemporal delta power encodes the overall stakes value

Neural encoding of the reward and punishment stake magnitude was present in a parametric fashion in delta power at left frontal and temporal sensors. Reward magnitude was encoded negatively, while punishment magnitude was encoded positively. The process underlying delta power modulations was thus sensitive to both stake valence and magnitude, integrating both stakes into the net "value" of the trial-by-trial stakes. The (left-) lateralized nature of this signal tentatively suggests a neural source not only medial prefrontal, but also lateral prefrontal or temporal regions. It is tempting to even speculate that this signal could arise from subcortical structures such as amygdala or hippocampus, which would be in line with our previous fMRI findings from a very similar task in which BOLD responses in these regions were modulated by the valence of a trial (Algermissen et al. 2022). Furthermore, delta/ theta power in amygdala and hippocampus has previously been found to reflect threat probabilities and regulate the expression of fear responses (Stujenske et al. 2014; Karalis et al. 2016; Manssuer et al. 2022), which is in line with the observation of higher delta power for more negative stakes. Recent studies have suggested that, in principle, MEG can reflect activity from subcortical areas (Dalal et al. 2013; Pizzo et al. 2019), but source reconstruction techniques would be needed to more precisely localize the source of this signal. Notably, this stakes signature was very different from signatures of outcomes once those were obtained at the end of the trial, tentatively suggesting different electrophysiological signatures—and underlying neural processes—for expected compared to obtained outcomes.

The neural process driving modulations in frontotemporal delta power is a candidate process for mediating the effect of stake magnitudes on action preparation as indexed in beta power desynchronization. However, delta power significantly encoded stake valence and magnitudes only around 0.9–1.5 sec. after stakes onset, i.e., after the transient resynchronization in beta power induced by high punishment stakes occurred. It is thus unclear whether and how the neural process reflected in left frontotemporal delta power is related to this beta power resynchronization, and whether it is involved in driving Pavlovian biases in action selection. We speculate that this process is already present earlier than can be detected with our linear regressions performed on average time-frequency power. Potentially, multivariate decoding analyses (King and Dehaene 2014) could shed further light on the temporal dynamics of this signal. Notably, this frontotemporal delta power modulation by the stakes occurred at the same time as occipital alpha and gamma power modulations by the stakes, potentially suggesting a second, late processing wave that follows (rather than drives) those processes that bias beta power.

5.4.3 Modulation of overt and covert attention by action plans

Our primary hypothesis was that covert spatial attention to reward and punishment stakes—indexed by posterior alpha power lateralization—would be affected by participants' action plans, drawing attention to those stakes that matched the action plan. Participants did attend to the stakes, evidenced by the fact that their actions were biased by stake magnitudes and that they performed the catch task with high accuracy. However, alpha power lateralization did not reflect any systematic bias in attention to the reward vs. punishment stake depending to participants' action plans. Similarly, also occipital gamma power lateralization and ERFs were not significantly influenced by participants' action plans.

One possible interpretation of the lack of systematic alpha lateralization is that humans indeed do not use attention in a selective manner that would match their action plans. This interpretation would explain the absence of alpha lateralization results in the current study, but be at odds with our previous eye-tracking study (Algermissen and den Ouden 2022) in which we found that overt attention (i.e., eye gaze) to reward/ punishment stakes was influenced by participants' action plans and predicted their eventual actions. Replicating these results in the current study, we found evidence that participants' impulsive saccades (which occurred despite instructions to keep their focus at the center of the screen) tended towards those stakes that matched their action plans and that, vice versa, the duration of participants' fixations on those stakes predicted their eventual responses. Possibly, these results point at a dissociation between covert attention mechanisms as reflecting in alpha power lateralization and overt attention mechanisms reflected in eye movements. Only the latter might be informed by action plans and selectively directed at stakes that matched ongoing action plans. Notably, our previous study employed a gaze-contingent design requiring participants to saccade to stakes in order to render them visible. Although somewhat artificial for a lab experiment, this setup might be closer to real life situations in which cues are spread across the environment and have to be actively sought out. In contrast, in this study, participants could passively monitor upcoming cues and then decide whether certain cues deserved special attention. Hence, task designs that require or incentivize active exploration of potential stakes via eye-movements—instead of studies merely requiring covert monitoring—might be more suited to observe such effects.

The time course of alpha power during stakes presentation was largely parallel over left and right sensors. Notably, before stakes onset, alpha transiently increased over both left and right posterior sensors, potentially reflecting a global suppression of visual cortex to prevent any

distractor interference (Rihs et al. 2007). After stakes onset, alpha power decreased again, and tended to do so more strongly over right posterior sensors when the stakes that matched participants' action plans appeared on the left (contralateral) side. While this effect was in line with our hypothesis, it occurred rather late and was not paralleled by an inverse effect over left sensors. Similarly, stake valence and magnitude modulated alpha power around the same time point, again selectively over right sensors. Possibly, right sensory cortices (or participants' attention to the left hemifield) were more sensitive to task factors. The late time point of these modulations is in line with a reactive attentional strategy. Potentially, this task did not incentivize participants sufficiently to plan their attentional trajectory ahead of time. Instead, they solved it in a merely reactive manner determined not (only) by top-down plans, but also bottom-up factors (such as stake valence and magnitude mapping) that varied across trials (Wolfe et al. 2000; Geng 2014). It remains to be tested whether higher task demands that require attentional trajectories to be planned ahead of time would reveal those trajectories to be influenced by Go/ NoGo action plans.

5.4.4 Conclusion

In sum, we found evidence for the magnitude of the reward and punishment stakes biasing action selection in the fashion of Pavlovian biases. We did not find evidence for humans covertly attending to these stakes (indexed by alpha power lateralization) in a way that would invigorate ongoing action plans. However, we did find evidence for overt attention to these stakes (as indexed by eye movements) in line with participants' action plans as well as overt attention biasing eventual responses, replicating our previous findings. This discrepancy between findings in alpha power and eye-movements might be explained by differences between overt and covert attentional processes as well as proactive and reactive attentional strategies. More challenging task designs might be needed to unravel the neural mechanisms of how humans use proactive, top-down attentional strategies to selectively invigorate Pavlovian biases.

5.5 Methods

5.5.1 Sample and exclusion criteria

We collected MEG data from 40 volunteers (M_{age} = 25.38, SD_{age} = 5.47, range 18–45; 21 female, 37 right-handed, all normal or corrected-to-normal vision). Sample size was determined by computing the effective sample size (Aarts et al. 2014) for 35 participants performing 264 trials (9,240 trial in total), which, assuming an intra-cluster coefficient of 0.10 (estimated on previous EEG data), effectively yielded 2,100 trials and 80% power to detect effects with standardized regression coefficients of $\beta > 0.06$. We additionally collected data from five extra participants to account for low performance and data quality. Data collection, sample size, and primary analyses were pre-registered (https://osf.io/kn7gj).

We recruited participants via the SONA Radboud Research Participation System of Radboud University. Inclusion criteria were age 18–45; self-reported English language mastery; normal or corrected vision (lenses permitted, but not glasses); no previous treatment for neurological or psychiatric disorders, epilepsy, severe concussions, or brain surgery; and no metal objects in the body such as plates or screws, vascular clips, active implants or pacemakers, permanent medical patches, non-removable piercings, metal splinters, or metal dental wires.

After data collection, in line with our pre-registered exclusion criteria, we excluded data from four participants who did not perform significantly above chance on the Oyster task (i.e., correct response on less than 134 trials, which is 56% of trials and not significantly above chance based

on a one-tailed binomial test based on $p < .05$) from all analyses. None of the participants exhibited MEG artifacts on more than 33% of trials. All participants exhibited head movement of at most 10 mm relative to the starting position of task blocks (33 participants < 5 mm, 21 participants < 3 mm). Five participants performed saccades towards (at least) one of the stakes on more than 33% of trials (i.e., 88 trials). Analyses of posterior alpha power lateralization were performed both with and without these participants as well as with and without any trial containing detected saccades; exclusions did not alter the results.

Participants took part in a 1.5h session consisting of informed consent and instructions, preparations and de-metallization, set-up of the MEG scanner and head localization, performance of first the Posner localizer task (15 min.) and then the Oyster Farming Task (45 min.), debriefing, and scanning of participants' head shape. Furthermore, 17 participants (for whom no anatomical MRI scans were available from previous studies run at the center) took part in a separate 15 min. session in which a structural T1 MRI scan was acquired. Participants received €17.50 for participation (another €5 for the anatomical MRI scan) plus a performance-dependent bonus between 1–5€ (M = 3.81, SD = 1.19, range 1.28–5.00). Research was approved by the local ethics committee (CMO Arnhem-Nijmegen) under the general ethical approval for the Donders Centre for Cognitive Neuroimaging (Imaging Human Cognition, CMO2014/288) and participants provided written informed consent before the experiment.

5.5.2 Oyster Farming Task

To assess neural mechanisms of how action selection biased attention allocation to rewards/ punishments and how attention biased eventual action execution, participants performed an adapted motivational Go/ NoGo learning task framed as Oyster Farming task. In the cover story, participants were told that they would farm oysters that can grow pearls (rewards) or tumors (punishments) based on whether they were fed (Go) or not (NoGo). Pearls gained money as they could be sold, while hazardous tumors cost money for waste disposal. Participants could preview the potential pearls and tumors an oyster might eventually grow (depending on the given response) between action selection (Go/ NoGo cue presentation) and action execution (imperative response cue). Participants performed 264 trials split into three blocks of 88 trials (80 Go/ NoGo, 8 catch), each block with a new set of cues.

Each trial started with one of four abstract cues (Agathodaimon letter representing a distinct oyster type) presented for 1,000 ms. Each oyster type either needed to be fed (Go) or not be fed (NoGo), which participants had to learn from trial-and-error. Showing the correct response led to pearls (rewards) on 77% of trials (valid trials), and tumors (punishments) otherwise (invalid trials), with reversed contingencies for incorrect responses. Response execution was only possible at a later phase of the trial and had to be withheld until then.

Next, potential pearls and tumors were presented for 1,500 ms on both sides of the fixation cross (displacement of 3.6°), displaced below the horizontal meridian by 2.1° to increase alpha power (Hillebrand and Barnes 2002). Participants were instructed to keep fixation at a centrally presented fixation cross throughout the trial while covertly attending to pearls and tumors which were relevant for an unrelated catch task (see below). Fixation was tracked via eye-tracking recordings (see S5.8 for eye-data quality). Pearls and tumors were colored in orange (RGB = [200, 100, 7]) and blue (RGB = [104, 104, 255]), respectively, with color assignment counter-balanced across participants. Whether tumors/ pearls were presented on the left/ right or vice versa varied across trials (orthogonal to action requirements), but upcoming positions were already indicated by faint half-circles of the respective color while Go/ NoGo cues were on the screen.

After a variable delay (100–500 ms, uniform distribution in steps of 100 ms), an imperative response cue appeared, which was a black ellipse (representing the closing oyster) still open on the respective left/ right side. To feed the oyster, participants now had to press the a left/ right button—depending on the side whether the oyster was still open—within a deadline of 600 ms. Keeping the button required for feeding uncertain until this moment precluded premature responding. Button presses in time were confirmed by a small food can positioned above the ellipse falling over in the respective direction. Late responses were registered, but counted as NoGo when determining outcomes.

Finally, after a delay of 500 ms, participants received the previously presented number of pearls of tumors as an outcome depending on their response (correct/ incorrect) and trial validity (valid/ invalid). Outcomes were presented for 700 ms. The next trial proceeded after a variable delay (1,200–1,800 ms; uniform distribution in steps of 100 ms).

One 24 out of 264 trials, instead of a response execution phase, a catch task was implemented to incentivize participants to pay attention to potential pearls and tumors. One these trials, participants had to indicate whether the oyster featured more pearls or tumors (to retrieve it from the police department after it was stolen) within a period of 4,000 ms.

See S5.1 for a description of the Posner localizer task.

5.5.3 Behavioral analyses

For behavioral analyses of the Oyster Farming task, in line with our pre-registration (https://osf.io/kn7gj), we excluded all catch trials. We furthermore excluded all reaction times (RTs) below 200 ms, as we reasoned that these responses were too fast to reflect processing of the Go/ NoGo cue (% of Go responses per participant: M = 0.4, SD = 1.1, range 0–6.2). Also, for trials with RTs above 800 ms, RTs were excluded and responses counted as NoGo (% of Go responses per participant: M = 13.9, SD = 3.9, range 4.9–24.5). We reasoned that 200 ms after response cue offset, participants should have realized that the permitted response window was over.

Responses were analyzed with mixed-effects logistic regression (function glmer) and RTs with mixed-effects linear regression (function lmer) with the *lme4* package in R. All independent variables that were treated as factors were zero-sum coded. All continuous independent or dependent variables were z-standardized such that regression weights can be interpreted on the same scale as standardized regression coefficients. To achieve a maximum random effects structure (Barr et al. 2013), we included all possible random intercepts, slopes, and correlations. *P*-values were computed via likelihood ratio tests (function mixed) with the *afex* package. We considered tests with p-values < 0.05 as significant.

We analyzed Go/ NoGo responses and RTs as function of required response, reflecting learning the task, and of stake differences, reflecting Pavlovian biases of stakes on behavior. The latter analyses included the required response and the reward valence mapping (whether the reward stake appeared on the left or the right) as covariates.

5.5.4 Computational modeling of behavior

We approximated participants' unobservable action intentions in two ways: First as a rough proxy, we used the action required by the specific cue presented on a given trial. However, participants could not know this action at the start of each block or when they never learned the correct response (for certain cues). Hence, as a more proximal measure, we fitted a Rescorla-

Wagner model (Rescorla and Wagner 1972) to each participant's choice data and simulated the values of Go and NoGo responses (Q-values) based on the outcomes obtained on past trials. In this model, on each trial *t*, an agent performs an action *a* to a stimulus *s* and obtains an outcome *r* (+1, -1). Based on this outcome, they update the Q-value for the respective action for the given stimulus:

$$Q_t(a_t, s_t) = Q_{t-1}(a_t, s_t) + \alpha * (r - Q_{t-1}(a_t, s_t)) \quad (1)$$

Q-values were then translated to action probabilities using a Softmax choice rule:

$$p(Go, s_t) = \frac{\beta * e^{Q_t(Go, s_t)}}{\beta * e^{Q_t(Go, s_t)} + \beta * e^{Q_t(NoGo, s_t)}} \quad (2)$$

This model featured two free parameters estimated based on the data, namely the learning rate α, reflecting how much past Q-values our updated based on recent outcomes, and the inverse temperature β, reflecting how deterministically participants' actions follow the estimated Q-values. We fitted parameters using a grid search, with α constrained to the range [0, 1] in steps of 0.01 and β constrained to the range of [1, 20] in steps of 0.1. Based on each participants' fitted parameters as well as their responses and obtained outcomes, we then simulated Q-values for Go and NoGo responses on each trial using one-step-ahead predictions (Steingroever et al. 2014).

5.5.5 MEG data collection

We collected MEG data with a 275-channel axial gradiometer MEG system (VSM/CTF Systems) in a magnetically shielded room. Six sensors (MRF66, MLC11, MLC32, MLC61, MLO33, MRO33) were permanently disabled due to high noise levels, leaving 269 intact sensors. During scanning, participants' head position was tracked over the three fiducials with online head motion tracking (Stolk et al. 2013). In case of a large amount of head motion, participants were asked to re-position their head to the starting position during breaks. Eye position and blinks were recorded using an Eyelink 1000 eye- tracker (SR Research). All data online low-pass filtered at 300 Hz and digitized at a sampling rate of 1,200 Hz. After the MEG session, participants' head shape was recorded relative to three fiducial coils using a 3D tracking device (Polhemus) to facilitate co-registration of MEG and structural MRI data for source reconstruction.

5.5.6 MR data collection

For aiding future source reconstruction analyses, we acquired anatomical MRI scans using a 3T MRI system (Siemens) using a T1-weighted MR-RAGE sequency with a GRAPPA acceleration factor of 2 (TR/TE = 2.3/ 3.03 ms, voxel size 1 mm isotropic, 192 transversal slices, 8° flip angle).

5.5.7 MEG pre-processing

MEG data for both tasks was preprocessed with Fieldtrip (Oostenveld et al. 2011) and MATLAB custom code (version 2019b). We mostly followed our pre-registered pre-processing steps (https://osf.io/kn7gj), but deviated by computing third-order synthetic gradients and rejecting additional IC components to account for an unexpected 20 Hz artifact induced by a camera in the magnetically-shielded room. First, in preparation of an independent component analysis (ICA), data were epoched in consecutive data bins of 1 sec; bins off task (i.e., during instructions, practice phase, and breaks) were excluded. We then computed third-order synthetic gradients to correct for a 20 Hz (and harmonics) artefact induced by electric circuits, demeaned the data, rejected bins with high variance based on visual inspection (to exclude SQUID jumps

and muscle artifacts characterized by increased high-frequency variance), and performed ICA. Components associated with blinks, saccades, muscle activity, heart rate, and non-biological noise (no 1/f shape, but peaks at 20 Hz and harmonics) were later removed (Posner task: *M* = 9.90, *SD* = 4.12, range 4–25; Oyster task: *M* = 17.73, *SD* = 4.60, range 9–30).

After artifact identification, raw data was re-epoched into trials, adding 1.5 sec. before and after events of interest to prevent edge artifacts in the time-frequency decomposition (Posner task: -1.5–3.25 sec. cue-locked, -1.5–2.0 sec. target-locked; Oyster Farming task: -1.5–4.0 sec. cue-locked, -1.5–3.0 sec. stakes-locked; -1.5–2.1 sec. response cue-locked; -1.5–2.2 sec. outcome-locked). We again computed synthetic gradients and demeaned the data. Trials containing data bins previously identified as noise-contaminated were rejected (number of rejected trials: Posner task: cue-locked: *M* = 25.33, *SD* = 16.75, range 6–98; target-locked; *M* = 21.56, *SD* = 14.36, range 6–83; Oyster task: cue-locked: *M* = 21.44, *SD* = 11.26, range 4 – 51; stakes-locked: *M* = 16.31, *SD* = 9.47, range 2–41; response cue-locked: *M* = 38.69, *SD* = 7.51, range 27–57; outcome locked: *M* = 39.25, *SD* = 7.64, range 27–60). Finally, noise components estimated from the ICA were removed from the data.

5.5.8 Time-frequency decomposition

For time-frequency decomposition, as pre-registered (https://osf.io/kn7gj), we zero-padded trials to 8.0 sec. and computed synthetic planar gradients. For low frequencies (2.5–40 Hz), we used Hanning tapers of 400 ms width for every 25 ms with 1 Hz resolution (400 ms time windows factually afford only 2.5 Hz resolution; however, data were interpolated to 1 Hz resolution, which affords integer frequency bins and circumvents the arbitrary selection of wider frequency bins). For high frequencies (32–100 Hz), we used a DPSS multi-taper approach with a sliding time window of 250 ms and steps of 25 ms with 8 Hz smoothing and a resolution of 4 Hz. After frequency decomposition, we combined the planar gradients.

5.5.9 Cluster-based permutation tests on condition-averaged data

Trial-by-trial time-frequency (TF) data was sorted into different conditions and averaged separately per condition per participant. Condition-averaged power was log10-transformed (to account for between-participant differences in overall power and to make power more normally distributed) and then averaged across participants. We tested for differences in time-frequency power between two respective conditions using two-sided cluster-based permutation tests with a cluster-forming threshold of $|t| > 2$ and $p < .05$ as criterion for statistical significance (Maris and Oostenveld 2007). For plots, data of two conditions was subtracted from each other and the difference was converted back to percent signal change by raising 10 to the respective power value, subtracting 1, and multiplying the resulting value by 100.

5.5.10 Multiple linear regression on single-trial time-frequency power

To test whether fronto-temporal delta power reflected both reward stake magnitude and punishment stake magnitude (1–5 pearls/ tumors) in a parametric fashion, we performed multiple linear regressions across trials at each sensor, frequency, and time point, using TF power as dependent variable and trial-by-trial reward and punishment magnitudes as independent variables. All variables were demeaned such that the intercept became zero. Such a multiple linear regression was performed for each participant, resulting in a time-frequency-sensor-predictor *b*-map reflecting the association between stake magnitudes and TF power at each time-frequency-sensor bin. *B*-maps were Fisher-z transformed, which makes the sampling distribution of correlation coefficients approximately normal and allows for combining them across participants. We then

performed sign-flipping cluster-based permutation tests across participants to test whether regression weights were significantly different from zero. Plots depict the *t*-values at each time/ frequency bin across participants averaged over sensors.

5.5.11 Trial-by-trial indices of beta power desynchronization and posterior alpha and gamma power lateralization

To test our pre-registered hypotheses about interactions between action preparation and attention allocation in the Oyster Farming task, we estimated the following trial-by-trial indices (all pre-registered under https://osf.io/kn7gj): 1) central beta power desynchronization during the cue phase; 2) central beta power desynchronization during the stakes phase; 3) posterior alpha power lateralization during the cue phase; 4) posterior alpha power lateralization during the stakes phase, 5) occipital gamma power lateralization during the stakes phase. Values for these indices were only computed for trials not rejected during MEG processing; values for rejected trials were set to NA.

For central *beta* power desynchronization both during *cue* and *stakes* presentation, we had initially pre-registered to fit a linear trend to beta power during Go/ NoGo cue presentation for every trial, quantifying the trial-by-trial variation in early beta power decrease. However, this metric did not distinguish trials with Go from trials with NoGo responses significantly (b = -0.02, SE = 0.02, $\chi^2(1) = 0.66$, $p = .42$), rendering it inadequate for capturing latent action plans. Instead, for quantifying beta power desynchronization during the *cue* phase, we computed the trial-by-trial average beta power in the cluster that distinguished Go from NoGo responses shortly after cue onset (Fig. 5.2A), i.e., between 0.400–0.950 sec., at central sensors in the beta range (13–33 Hz). Frequencies and sensors were not weighted equally, but each weight was set to the *t*-value obtained from a *t*-test contrasting power for Go vs. NoGo responses (averaged per condition per participant) across participants for the respective frequency bin and sensor. Absolute *t*-values below |2| were set to zero. This metric strongly and significantly distinguished Go and NoGo responses (b = -0.03, SE = 0.01, $\chi^2(1) = 11.81$, $p < .001$). For central beta power desynchronization during the *stakes phase*, we performed the same procedure, but in the time range of 0–1.5 after stakes onset, using *t*-values obtained from *t*-tests in the same window.

For quantifying *alpha* power lateralization both during cue and stakes presentation, we computed the *attentional lateralization index* (ALI) (Thut et al. 2006; Marshall et al. 2018) relative to the reward stake such that positive values reflected a stronger focus on the reward stake, while negative values reflected a stronger focus on the punishment stake. Given that alpha power was expected to decrease contralaterally to attended stakes, the alpha power ALI_α was thus computed as

$$ALI_\alpha = \frac{\alpha_{ipsilateral} - \alpha_{contralateral}}{\alpha_{ipsilateral} + \alpha_{contralateral}} \quad (3)$$

In line with our pre-registration, during the *cue* phase, we computed the trial-by-trial average alpha (7–14 Hz) power 0.4–0.8 sec. after cue onset (i.e., second half of cue presentation while, with tapers of 400 ms width, such a window excludes any signal from after stakes onset) separately for left and right posterior (i.e., parietal and occipital) sensors. Given known heterogeneity across participants in alpha power peak and the spatial topography of alpha lateralization (Haegens et al. 2014), we did not merely average across frequencies and sensors, but weighted each frequency bin and sensor based on a *t*-value mask from the independent localizer (Posner) task. In our pre-registration, we had initially planned to compute participant-specific *t*-value masks by contrasting trials on which cues pointed towards the left vs. trials on which cues pointed towards the right side

of the screen, separately for each participant. However, this contrast did not induce strong alpha power lateralization, not even at the group-level (see S5.1). In contrast, targets presented on the left vs. right side of the screen did induce strong alpha power lateralization (see S5.1) on a group-level, though not a single-participant level (see S5.1). We thus decided to use a common mask for all participants and to use the target side (target-locked) rather than the cue side (cue-locked) as contrast of interest. We computed a mask by performing a *t*-test at each sensor and frequency bin, contrasting power for left target vs. right targets (averaged per condition per participant) across participants. Absolute *t*-values below $|2|$ were set to zero. We then computed trial-by-trial average alpha power at left and right sensors while weighting each frequency and sensor by its respective *t*-value. For posterior alpha power lateralization during the *stakes* phase, we performed the same procedure, but in the time range of 0–1.5 after stakes onset.

For posterior *gamma* lateralization during the *stakes* phase, we again computed an attentional lateralization index. We computed the trial-by-trial average gamma power 0–0.5 sec. after stakes onset in the range of 45–65 Hz (Marshall et al. 2018), with a weight of 1 for each frequency and at sensor, separately for left and right occipital sensors. Unlike alpha, we focused on only occipital (and not parietal) sensors given that only these sensors showed a clear increase in gamma power after stakes onset. Given that gamma power was expected to increase (rather than decrease) contralaterally to attended stakes, having the ALI_γ pooled the same way as the ALI_α (with positive value reflecting more processing of rewards) implied the following calculation:

$$ALI_\gamma = \frac{\gamma_{contralateral} - \gamma_{ipsilateral}}{\gamma_{contralateral} + \gamma_{ipsilateral}} \quad (4)$$

5.5.12 Hypothesis testing on beta power desynchronization and alpha power lateralization using mixed-effects regression

We tested the following pre-registered hypotheses (https://osf.io/kn7gj):

1. *Hypothesis 1A*. Action plans (Go/ NoGo action requirement, Q-value difference) influence attention allocation (posterior alpha lateralization) during the cue phase.
2. *Hypothesis 1B*. Early action preparation (central beta power desynchronization) during the cue phase predicts attention allocation (posterior alpha lateralization) during the cue phase.
3. *Hypothesis 1C*. Action plans (Go /NoGo action requirement, Q-value difference) influence bottom-up processing (posterior gamma power lateralization) during stakes presentation.
4. *Hypothesis 1D*. Action plans (Go/ NoGo action requirement) influence bottom-up processing (ERF amplitude) during stakes presentation.
5. *Hypothesis 2A*. Attention allocation (posterior alpha power lateralization) during stakes presentation predicts the eventual response (Go/ NoGo).
6. *Hypothesis 2B*. Attention allocation (posterior alpha power lateralization) during stakes presentation predicts the ongoing action preparation (beta power desynchronization) during stakes presentation.

For hypotheses 1A, 1B, 1C, and 2B we used mixed-effects linear regression (*lmer* function) and for hypothesis 2A mixed-effects logistic regression (*glmer* function) via the *lme4* package in R. All independent variables that were treated as factors were zero-sum coded. All continuous

independent or dependent variables were z-standardized such that regression weights can be interpreted on the same scale as standardized regression coefficients. To achieve a maximum random effects structure (Barr et al. 2013), we included all possible random intercepts, slopes, and correlations. *P*-values were computed via likelihood ratio tests (function mixed) with the *afex* package. We treated tests with p-values < 0.05 as statistically significant. In all analyses, we included stake valence mapping (whether the reward stake appeared on the left or the right) as a covariate to account for participant-specific side biases in attention allocation. In addition, in analyses featuring beta power as independent variable (hypotheses 1B and 2B), we used the Q-value difference as a covariate to test whether beta power predicted trial-by-trial variation responses that was not yet captured by participants' learning history.

5.5.13 Cluster-based permutation tests on event-related fields

To test whether action requirements affected early event-related fields (ERF; hypothesis 1D), we performed a pre-registered (https://osf.io/kn7gj) cluster-based permutation tests (Maris and Oostenveld 2007). We split trials into four conditions based on whether a Go or NoGo action was required and whether the reward stake appeared on the left or the right side of the screen. Based on these four conditions, we formed two new conditions: one condition on which we expected higher amplitudes at *right* sensors, namely when a Go response was required and the reward stake presented on the left side or when a NoGo was required and the reward stake presented on the right side, and another condition in which we expected higher amplitudes at *left* senosrs, namely when a Go response was required and the reward stake presented on the right side or when a NoGo response was required and the reward stake presented on the left side. We then performed a cluster-based permutation test over time (0–0.5 sec. after stakes onset).

5.6 SUPPLEMENTARY MATERIALS FOR CHAPTER 5

5.6.1 S5.1: Behavioral results and posterior alpha power modulation in the Posner localizer task

Participants performed a Posner task for the participant-specific localization of posterior alpha power lateralization as an effect of attention to the left/ right side of the screen (Haegens et al. 2014). Furthermore, this task trained participants at keeping their attention fixated at the fixation cross in the middle of the screen while covertly attending to stimuli appearing on the left/ right side of the screen. In this task (Worden et al. 2000), participants first saw an arrow pointing to the left/ right for 500 ms and were instructed to covertly attend to the direction indicated by the arrow while maintaining fixation at the fixation cross in the center of the screen (see Fig. S5.1A panel A). After a variable delay (750–1,250 ms; uniform distribution in steps of 50 ms), a target (+ or x) appeared on the screen for 200 ms (programmed to be 20 ms, in fact 200 ms due to a hardware error as visible from log files), namely in 80% of trials on the side previously indicated by the arrow and in 20% of trials on the other side (horizontal displacement of 3.6°; 2.1° below the horizontal meridian to enhance stimulation of the upper bank of the calcarine fissure closer to MEG sensors (Hillebrand and Barnes 2002); exact same positions as for stakes in the Oyster Farming Task). Participants were instructed to classify the target as a + or x using left/ right buttons (response assignment counter-balanced across participants) as fast as possible (response deadline of 1,200 ms). Keeping target validity below 100% discouraged participants from saccading to the anticipated target location. After a variable inter-trial-interval (1,250–1,750 ms; uniform distribution in steps of 50 ms), the next trial started. At the start of the task, participants received instructions and performed 20 practice trials during which response feedback (correct/ incorrect/ press faster) was given. Afterwards, they performed 200 trials (two blocks of 100 trials) without feedback.

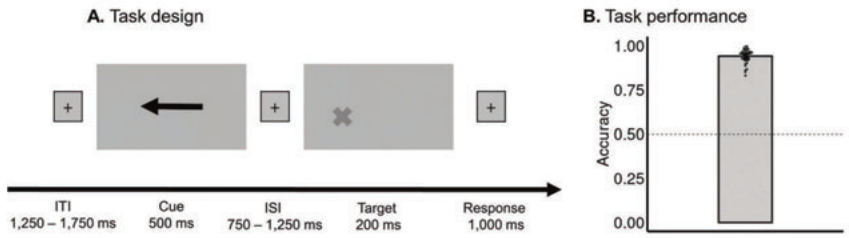


Figure 5.8. S5.1A. Posner task design and task performance.
A. On each trial of the Posner task, participants first saw a cue (i.e., an arrow) pointing to the left/ right side of the screen. The side to which the arrow pointed matched the side on which a target would later appear on 80% of trials. After cue offset and a variable inter-stimulus interval, a target stimulus (+ or x) appeared for 200 ms either on the left or the right side of the screen. Participants used left/ right button presses to classify the identity (+ or x) of the target. **B.** All participants performed the Posner task with high accuracy.

Participants performed well in the Posner task (% correct: M = 94.0, SD = 4.0, range 83.0–99.5; see Fig. S5.1A panel B). All participants performed significantly above the chance level (57% determined with a one-sided binomial test based on 200 trials).

First, we tested whether alpha (8–13 Hz) power at posterior (parietal/ occipital) sensors decreased more strongly contralaterally to the side whether the cue (i.e., arrow) was pointing at. A cluster-based permutation test (band- and sensor averaged) was not significant at left posterior

sensors (p = .086; Fig. S5.1B panel A), but at right posterior sensors (p = .016, cluster above threshold around 0.400–0.650 sec.), suggesting a stronger decrease when the cue was pointing to the right vs. left side of the screen (Fig. S5.1B panel B). Overall, cue-induced posterior alpha power lateralization was rather weak and not consistently found across participants, rendering this contrast an inappropriate localizer for alpha power lateralization (Fig. S5.1C panel A).

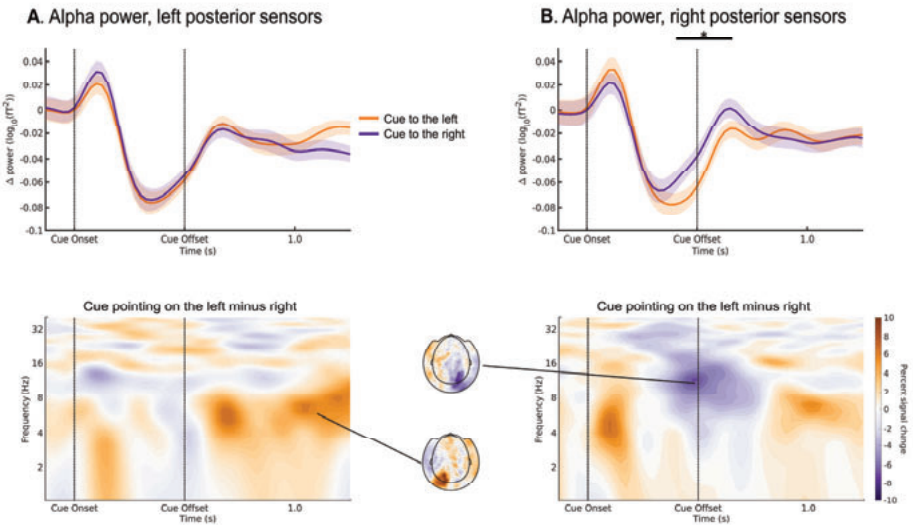


Figure 5.9. S5.1B. Alpha power modulation by cue location at left and right posterior sensors.
A. Mean (±SEM across participants) alpha power (8–13 Hz) at left posterior sensors was not significantly modulated by cue location. **B.** Mean alpha power at right posterior sensors decreased significantly more strongly when the cue pointed at the left (contralateral) compared to the right side of the screen (cluster above threshold around 0.400–0.650 sec.). TF plots display power for cues pointing to the left minus power for cues pointing to the right side of the screen.

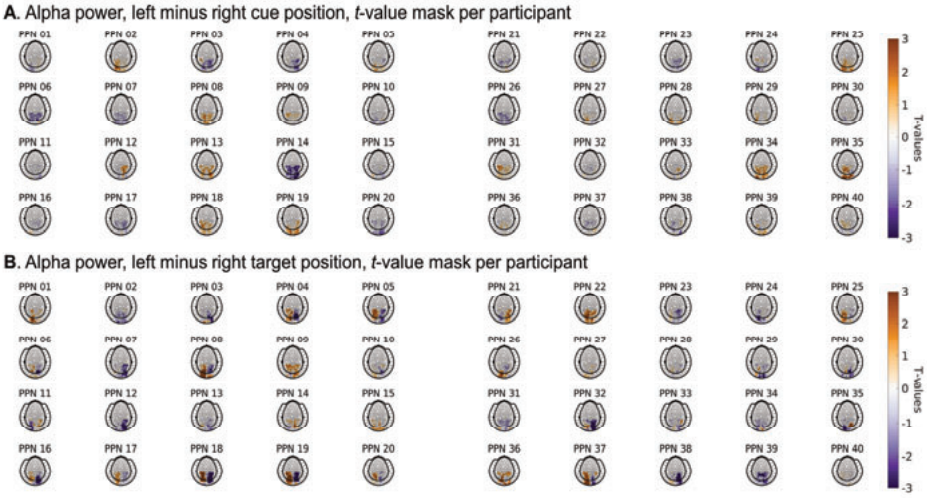


Figure 5.10. S5.1C. Alpha power modulation by cue location (A) and target location (B) for each individual participant.
Alpha power (7–14 Hz) 0-0.8 sec. after cue/ target onset showed lateralization (i.e., lower power contralateral to the cue/ target) only for a minority of participants. Target location appeared to induce much stronger modulations than cue location. Still, not every participant showed a clear target-induced lateralization. *T*-values were obtained by computing a between-trials *t*-test at each sensor and frequency bin in the range of 0–0.800 sec. relative to cue/ target onset, with cues/ targets on the left minus cues/ targets on the right side of the screen.

Next, we tested whether alpha power locked to the target onset showed a significant modulation by the cue side. There was no significant modulation of alpha power by the cue location in the last 0.5 sec. *before* target onset, neither at left nor right posterior sensors (no clusters above threshold). Only *after* target onset, significant modulations—not by the *cue* location, but rather the *target* location—appeared: Alpha power at left posterior sensors decreased more strongly when the target appeared on the right (contralateral) side (p = .006, cluster above threshold around 0.250–0.675 sec.; Fig. S5.1D panel A), and vice versa, alpha power at right posterior sensors decreased more strongly when the target appeared on the left (contralateral) side (p = .002, 0.200–0.750 sec.; Fig. S5.1D panel B). Although strongly significant at the group-level, only few participants showed significant alpha power modulation at an individual participant-level (Fig. S5.1C panel B).

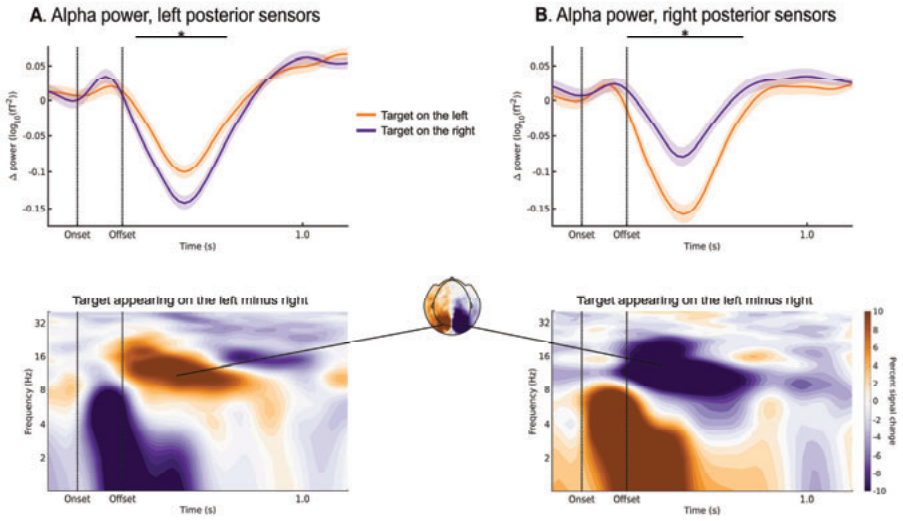


Figure 5.11. S5.1D. Alpha power modulation by target location at left and right posterior sensors.
A. Mean (±SEM across participants) alpha power (8–13 Hz) at left posterior sensors decreased more strongly when the target appeared on the right (contralateral) compared to the left side (around 0.250–0.675 sec.). **B.** Vice versa, mean alpha power at right posterior sensors decreased significantly more strongly when the target appeared on the left (contralateral) compared to the right side (around 0.200–0.750 sec.). TF plots display power for left targets minus right targets.

Given that alpha power was only weakly modulated by cue location, but strongly modulated by target location, we chose to use the group-level t-mask of left vs. right targets (thresholded at $|t| < 2$) at posterior sensors (0.2–0.8 sec.) as a localizer for our single-trial analyses. The primary reason for this choice was a robust modulation by target location (i.e., t-values consistently exceeding $|2|$ for all sensors). Limitations of this approach is that it cannot capture individual differences in peak frequency bins or sensors (Haegens et al. 2014). Furthermore, at the time of the target, visual stimulation was provided on the screen; hence, this contrast does not reflect anticipatory/ proactive attention, but reactive attention to a stimulus. Still, given the absence of strong alpha modulations by either cue location or target location on the single participant-level, this group-level mask was the only analyses that showed robust alpha power lateralization as needed for testing our pre-registered hypotheses.

5.6.2 S5.2: Central alpha, beta and gamma power modulation around responses

In line with our expectation of beta power (but not alpha or gamma power) constituting a latent index of online action preparation and execution, we also tested for differences between trials with Go responses and trials with NoGo responses *around the time of responses*. In line with previous findings (Salmelin and Hari 1994; Salmelin, Hämäläinen, et al. 1995; Pfurtscheller et al. 2003; Neuper et al. 2006; Donner et al. 2009), we expected such an index to emerge *several seconds before* response execution, *peak* at response execution, and then *show a rebound* after the response. We tested whether these expectations were fulfilled by beta power, but also alpha or gamma power.

Relative to the onset of the response cue, beta power (13–33 Hz; band-averaged) at central sensors (sensor-averaged) was first significantly lower for Go than NoGo responses ($p < .001$, cluster above threshold around 0–0.675 sec.; Fig. S5.2 panel A), but later significantly higher for Go than NoGo responses ($p < .001$, cluster above threshold around 0.800–1.500 sec.), consistent with previous reports of a beta "rebound". These findings are in line with previous literature and, in conjunction with Go/ NoGo differences emerging several seconds before the eventual response (see main text), suggest that beta power at central sensors is a suitable measure of latent action preparation.

Besides beta power, we also tested whether alpha and gamma power at central sensors showed signatures of latent action preparation. We focused on two signatures: i) exhibiting differences between Go and NoGo responses already before onset of the response cue; ii) showing a rebound after the response has been made. For a third signature (stronger effects contralateral to the response hand), see S5.3.

With respect to alpha power (8–13 Hz), first, we tested whether power (band-averaged) at central sensors (sensor-averaged) was significantly different between trials with Go responses and trials with NoGo responses in-between the onset of the Go/ NoGo cue and the eventual response. Alpha power was significantly lower for Go than NoGo responses ($p < .001$, cluster above threshold around 1.500–3.500 sec. relative to Go/ NoGo cue onset), with differences above threshold emerging around the time of stakes presentation and thus later than differences in beta power. Second, we tested whether alpha power rebounded at the time of responses (locked to onset of the response cue). Alpha power was lower for Go than NoGo responses ($p < .001$, cluster above threshold around 0–1.050 sec.), but showed no rebound after the response (Fig. S5.2 panel A). Instead, Go /NoGo differences in alpha power continued for several hundred milliseconds longer than differences in beta power, extending beyond the response deadline. Taken together, alpha power began to distinguish trials with Go and NoGo responses already seconds before the response cue, but did not rebound after the response. Together with the fact that alpha effects were not stronger contralaterally to the response hand (see S5.3), these findings suggest that alpha power is not a latent index of action preparation.

With respect to gamma power (32–100 Hz), again, we first tested whether power (not band-averaged given expectable heterogeneity across the broad gamma band) at central sensors (sensor-averaged) was significantly different between trials with Go responses and trials with NoGo responses in-between the onset of the Go/ NoGo cue and the eventual response. The respective cluster-based permutation test was not significant ($p = .15$), providing no evidence for gamma power indexing the upcoming response already before response onset. Second, we tested whether gamma power rebounded at the time of responses (locked to onset of the response cue). Gamma power was significantly higher for Go than NoGo responses around the time of responses ($p < 0.001$, cluster above threshold around 0.075–1.500 sec.), but did not show a "rebound"-like

reversal in the power time courses. Visual inspection of the TF plot (Fig. S5.2 panel B) revealed that this result was driven by two separate clusters, one around 0.075–0.675 sec. in the upper gamma range (64–88 Hz) and one around 0.550–1.500 sec. in the lower gamma range (32–56 Hz). The latter cluster occurred at the same time as the beta power rebound, suggesting that the beta power rebound might actually have extended into lower gamma band frequencies. Notably, (upper) gamma power showed no (negative) rebound after the response. Taken together, gamma power did not distinguish trials with Go and NoGo responses before the response cue and did not rebound after the response. These findings suggest that (upper) gamma power is not a latent index of action preparation. Note Fig. 5.2A for a similar pattern in delta/ theta power around the time of responses, also not distinguishing Go/ NoGo responses at any earlier time points, but strongly after the onset of the response cue.

Taken together, only beta power, but not alpha or gamma power exhibited features of an index tracking latent action preparation (i.e., predicting responses before response execution, rebound directly after response execution). These findings ascertain the quality of our data, being able to replicate patterns consistently found in previous literature, and motivate our focus on beta power as an index of latent action preparation.

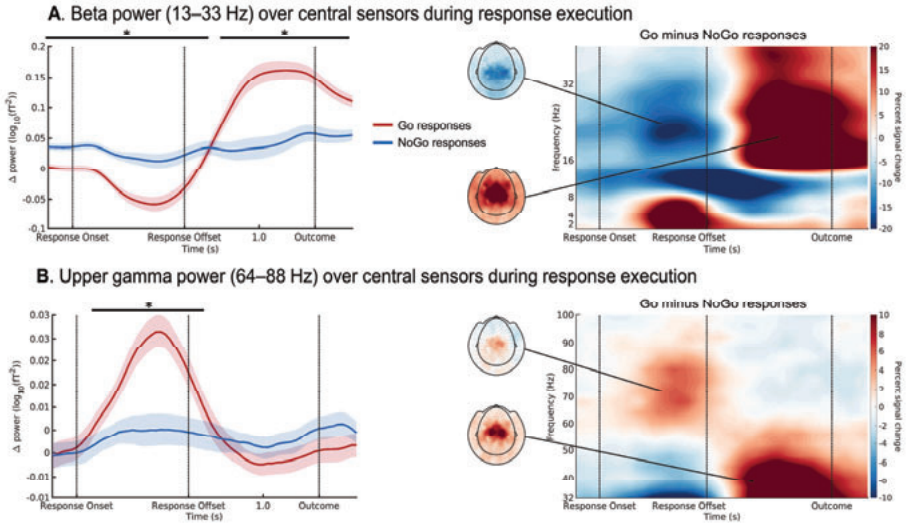


Figure 5.12. S5.2. Beta power and gamma power modulation by Go/ NoGo responses around the time of responses.
A. *Beta power around responses decreases, then rebounds for Go responses.* Mean (±SEM across participants) beta power (13–33 Hz) at central sensors was first lower on trials on which participants eventually performed a Go response (red line) compared to trials on which they performed a NoGo response (blue line; 0–0.675 sec. relative to response cue onset), but later showed a rebound with higher beta power after Go than NoGo responses (0.800–1.500 sec.). No such rebound was visible in the alpha (8–13 Hz) band. The black horizontal line indicates the time range for which the cluster driving significance is above threshold. **B.** *Gamma power increases around Go responses, without a rebound.* Mean upper gamma (64–88 Hz) power increased around Go compared to NoGo responses, but there was no subsequent "rebound"-like dip below the power level of NoGo responses. Lower gamma (32–56 Hz) power was higher after Go compared to NoGo responses (0.550–1.500 sec.), potentially reflecting a spreading of the beta power rebound into lower gamma frequencies.

5.6.3 S5.3: Central alpha, beta and gamma power modulation by the response hand

Apart from decreasing at the time of responses and increasing ("rebounding") afterwards, beta power has typically been found to show stronger decreases and increases at central sensors contralateral to the response hand (Salmelin and Hari 1994; Salmelin, Hämäläinen, et al. 1995; Pfurtscheller et al. 2003; Neuper et al. 2006; Donner et al. 2009). We tested for this response hand-lateralization in the alpha, beta, and gamma band at central sensors.

First, the decrease in beta power (13–33 Hz; band-averaged) at left central sensors (sensor-averaged) was not significantly different between left and right responses. However, at later time points, left central beta power increased significantly more strongly for right (contralateral) compared for left (ipsilateral) hand responses ($p < .001$, cluster above threshold around 0.700–1.500 sec.; Fig. S5.3A panel A). Vice versa, beta power at right central sensors initially decreased more strongly for left (contralateral) compared to right (ipsilateral) hand responses ($p < .001$, cluster above threshold around 0.225–0.725 sec.; Fig. S5.3A panel B), but later also increased significantly more strongly after left compared to right hand responses ($p < 0.001$, cluster above threshold around 0.875– 1.500 sec.). Taken together, beta power decreases tended to be stronger contralateral to the response hand, with differences being significant only at right central sensors. In contrast, stronger subsequent increases ("rebounds") contralateral to the response hand occurred at both left and right central sensors.

Second, alpha power (8–13 Hz; band-averaged) at left central sensors (sensor-averaged) decreased significantly more strongly for right (contralateral) compared to left (ipsilateral) hand responses ($p = .008$, clusters above threshold around 0.350–0.575 sec. and 0.900–1.275 sec.). There was however no rebound in alpha power (see Fig. S5.3 panels A, B). Vice versa, there was no significant difference between left hand and right hand responses in alpha power at right central sensors ($p = .052$). Taken together, alpha power decreases around the time of the response tended to be stronger for responses of contralateral compared to the ipsilateral hand, but this difference was only significant at left central sensors.

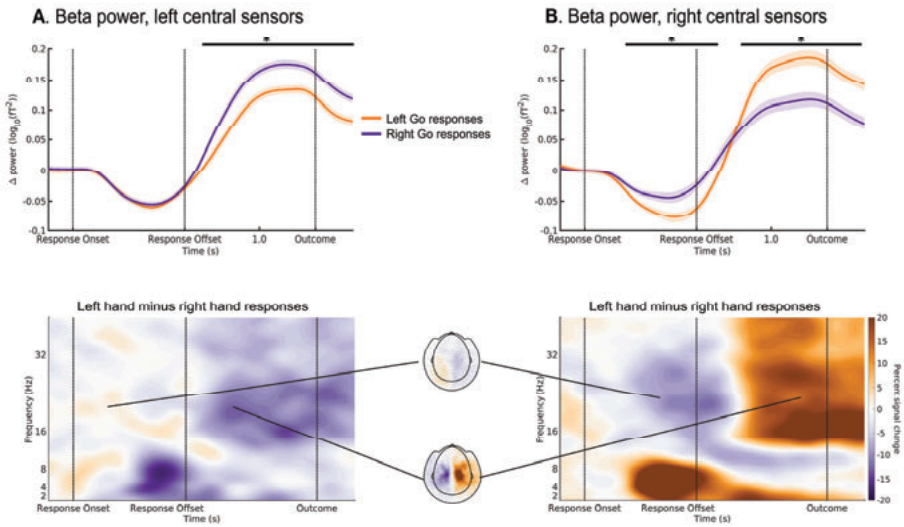


Figure 5.13. S5.3A. Beta power at left and right central sensors shows stronger changes for contralateral responses.
A. The decrease in mean (±SEM across participants) beta power (13–33 Hz) at left central sensors was not modulated by participants' response hand, while the subsequent increase ("rebound") was stronger for responses of the right (contralateral) hand. No such lateralization was visible in the alpha (8–13 Hz) band. The black horizontal line indicates the time range for which the cluster driving significance was above threshold. **B.** Vice versa, the decrease as well as the subsequent increase ("rebound") in mean beta power at right central sensors was stronger for responses of the left (contralateral) hand.

Third, upper gamma power (64–88 Hz; not band averaged) at left central sensors (sensor-averaged) increased significantly more strongly during right (contralateral) compared to left (ipsilateral) hand responses ($p < .001$, cluster above threshold around 0.325–0.675 sec.; Fig. S5.3B panel A), while upper gamma power at right central sensors increased significantly more strongly during left (contralateral) compared to right (ipsilateral) hand responses ($p < .001$, cluster above threshold around 0.075–0.925 sec. ; Fig. S5.3B panel B). Conversely, lower gamma power (32–56 Hz) showed the same pattern, but at a later time point (left central sensors: $p < .001$; cluster above threshold around 0.375–1.500 sec.; right central sensors: $p < .001$; cluster above threshold around 0.750–1.500 sec.), i.e., at the time when also rebounds in beta power occurred. Taken together, gamma power increased more strongly contralaterally to the response hand.

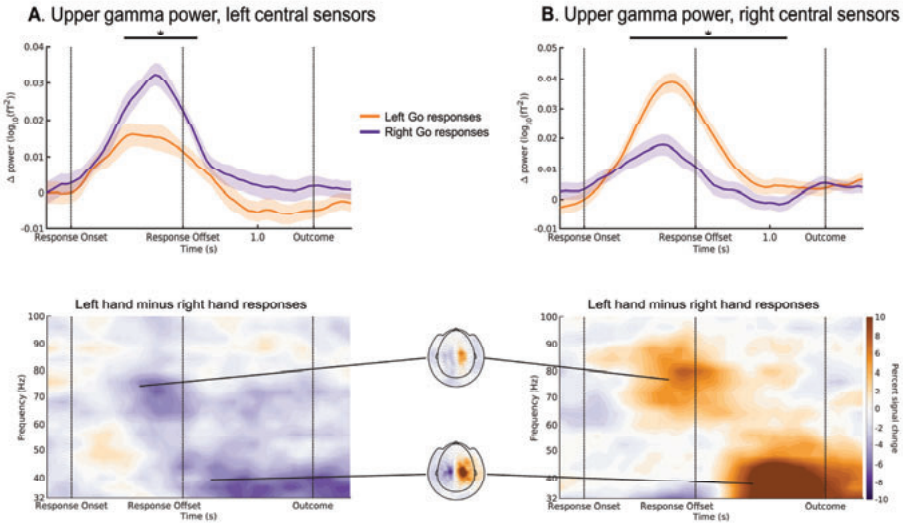


Figure 5.14. S5.3B. Upper gamma power at left and right central sensors increases more strongly during contralateral responses.
A. The increase in mean (±SEM across participants) upper gamma power (64–88 Hz) at left central sensors was stronger for responses of the right (contralateral, purple line) hand. At later time points, lower gamma power (32–56 Hz) was higher for contralateral responses, similar to beta power (see Figure S5.3A). The black horizontal line indicates the time range for which the cluster driving significance was above threshold. **B.** Vice versa, the increase in mean upper gamma power at right central sensors was stronger for responses of the left (contralateral, orange line) hand. At later time points, lower gamma power was higher for contralateral responses, similar to beta power (see Figure S5.3A).

In sum, all three power bands showed a stronger modulation contralateral to the response hand. These findings ascertain the quality of our data, being able to replicate patterns consistently found in previous literature.

5.6.4 S5.4: Central beta power predicts reaction times

We observed that on trials on which the punishment stake exceeded the reward stake ("Avoid" trials), beta power transiently resynchronized again, even when participants eventually performed a Go response. Even though this resynchronization was apparently not sufficient to induce NoGo responses, it might have still slowed down reaction times (RTs) for Go responses. We thus investigated when beta power started correlating with eventual RTs and if beta power already before, or only after the transient resynchronization induced on Avoid trials was predictive of RTs.

We addressed this question with two approaches. In the first approach, we performed a median split on the RTs for Go responses for each participant and averaged data for each condition for each participant. A cluster-based permutation test across participants on beta power (13–33 Hz, band-averaged) at central sensors (sensor-averaged) suggested that beta power was significantly lower on trials with fast RTs compared to trials with slow RTs ($p < .001$), driven by cluster above threshold around 0.950–1.500 sec. (Fig. S5.4 panel A). This difference between trials with fast and slow responses occurred later than the transient resynchronization on Avoid trials.

In the second approach, we performed a multiple-linear regression at each sensor, frequency, and time point, using RTs (z-standardized) as regressor. A sign-flipping cluster-based permutation test suggested significant positive correlations with beta power ($p = .006$), driven by cluster above threshold around 1.050–1.550 sec. in the range of 9–26 Hz (Fig. S5.4 panel B). Similar to condition differences based on a median split, this correlation was only significant at time points later than the transient resynchronization on Avoid trials.

In sum, two complementary approaches suggested that beta power before the transient resynchronization that was induced by high punishment stakes was not yet predictive of eventual RTs. Only after the resynchronization did a significant correlation (and differences between fast and slow trials) occur. These findings are in line with the notion that RTs were not yet determined before the impact of stakes on beta power, but only afterwards, with stake magnitudes potentially playing a role in slowing RTs.

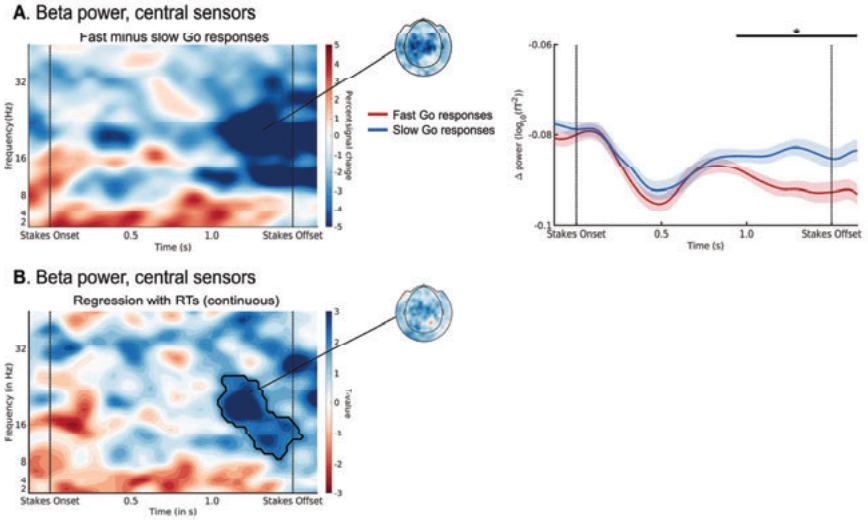


Figure 5.15. S5.4. Beta power at central sensors predicts RTs only late in the trial.
A. Mean (±SEM across participants) beta power at central sensors was significantly lower on trials with fast RTs compared to trials with slow RTs around 0.950–1.500 sec. The black horizontal line indicates the time range for which the cluster driving significance was above threshold. **B.** Beta power around 1.075–1.650 sec. was significantly positively correlated with RTs (i.e., lower beta power predicted faster responses; note that the color axis is flipped to achieve the same color pattern as in panel A). Solid black lines indicate clusters above threshold.

5.6.5 S5.5: Whole-scalp power modulations by outcome valence and magnitude

Besides the processing of stake magnitudes when stakes were merely expected, we also asked whether time-frequency power measured with MEG exhibited similar correlates of rewards and punishments once participants obtained these outcomes. fMRI studies have observed similar BOLD responses to both expected and obtained outcomes in the same regions, e.g., striatum and PFC (Bartra et al. 2013). However, while in this study, we found stake magnitudes at the time of stakes presentation to be reflected in left-lateralized frontotemporal delta power, in our previous EEG-fMRI study (Algermissen et al. 2021), we found correlates of obtained outcome in delta/ theta power at frontal midline electrodes and beta power at posterior midline electrodes. Given this discrepancy between the encoding of expected and obtained outcomes, we aimed to replicate our EEG-fMRI findings using this MEG data set.

We performed two weakly constrained cluster-based permutation tests: One test on delta/ theta power (1–8 Hz; band averaged) at anterior (i.e., frontal/ temporal/ central) sensors (sensor-averaged) and one test on alpha/ beta power at posterior (i.e., parietal/ occipital) sensors. Delta/ theta power at anterior sensors was first significantly higher for punishment compared to reward outcomes ($p < 0.001$, 0.175– 0.500 sec.) and then higher for reward than punishment outcomes ($p = .004$, 0.675–1.000 sec.; Fig. S5.5A panel A). Vice versa, alpha/ beta power at posterior sensors was significantly higher for reward compared to punishment outcomes ($p < 0.001$, 0.175–1.000 sec.; Fig. S5.5A panel B). Visual inspection of the time-frequency plot revealed two clusters: an early one in which delta/ theta power was higher for punishments than rewards, and a later one in which theta/ alpha/ beta power was higher for punishments than rewards. The first cluster peaked at frontal midline sensors, but spanned almost all sensors. The second cluster peaked at posterior midline sensors, but also spanned all sensors.

Furthermore, we explored whether gamma power (32–100 Hz) reflected outcome valence, which has been reported by past literature (Gueguen et al. 2021; Strube et al. 2021). We performed a weakly constrained cluster-based permutation test on gamma power (not band-averaged given expectable heterogeneity across the broad gamma band) at all sensors (sensor-averaged). The resulting permutation test was significant ($p = .018$), driven by a cluster around 0.225–0.675 sec. and 48–72 Hz (Fig. S5.5A panel C). Visual inspection of the topography of power differences revealed that differences were maximal at (right) parietal sensors.

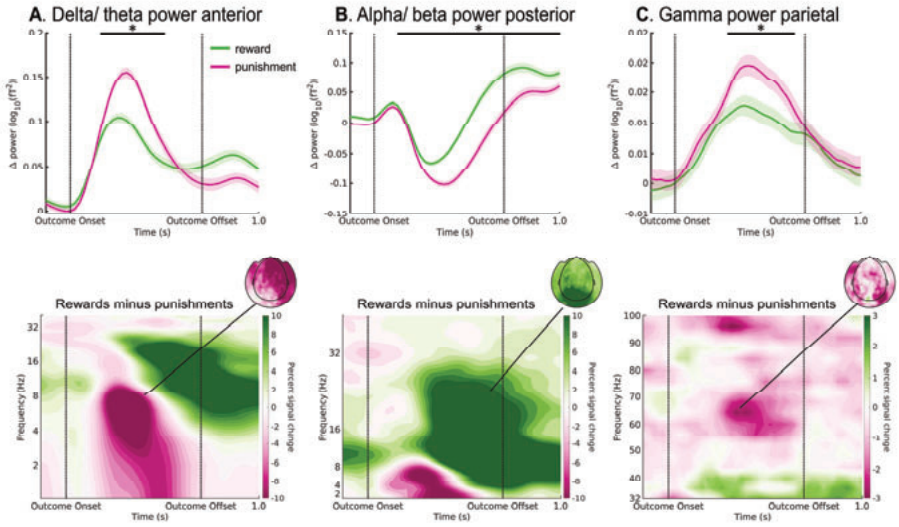


Figure 5.16. S5.5A. Delta/ theta, alpha/ beta, and gamma power modulation by outcome valence.
A. Mean (±SEM across participants) delta/ theta power (1–8 Hz) at anterior sensors (peak at right frontal sensors) was stronger for punishments (pink line) compared to rewards (green line) around 0.175–0.500 sec. after outcome onset. The black horizontal line indicates the time range for which the cluster driving significance was above threshold. **B.** Alpha/ beta (8–33 Hz) power at posterior sensors (midline occipital/ parietal sensors) was stronger for reward compared to punishment around 0.175-1.000 sec. after outcome onset. **C.** Gamma power (48–72 Hz) at posterior sensors (peak at right parietal sensors) was stronger for punishments compared to rewards around 0.225–0.675 sec. after outcome onset.

Finally, we performed multiple-linear regressions at each sensor, frequency, and time point, using both outcome valence (positive/ negative) and outcome magnitude (1–5) as regressors. Again, delta/ theta power ($p < .001$, 0.150 – 1.000 sec., 1–8 Hz, peaking at right frontal sensors; Fig. S5.5B panel A) and gamma power ($p = .040$, 52–68 Hz, 0.275–0.625 sec., peaking at right parietal sensors; Fig. S5.5B panel C) encoded outcome valence negatively, while alpha/ beta power ($p < .001$, -0.125–1.000 sec., 8–32 Hz, peaking at posterior sensors; Fig. S5.5B panel B) encoded outcome valence positively. These findings replicated the condition differences reported above. In contrast, outcome magnitude was primarily reflected in the signal at posterior sensors, being encoded positively in delta/ theta power ($p < .001$; 0.025–0.925 sec., 1–8 Hz; Fig. S5.5B panel A), negatively in alpha/ beta power ($p < .001$; 0.025– 1.000 sec., 8–33 Hz; Fig. S5.5B panel B), and positively in broadband gamma power ($p < .001$; 0–1 sec., 48–100 Hz; Fig. S5.5B panel C). These changes in signal over visual cortices likely reflect more visual input (e.g., color, contrast) provided by higher magnitude outcomes.

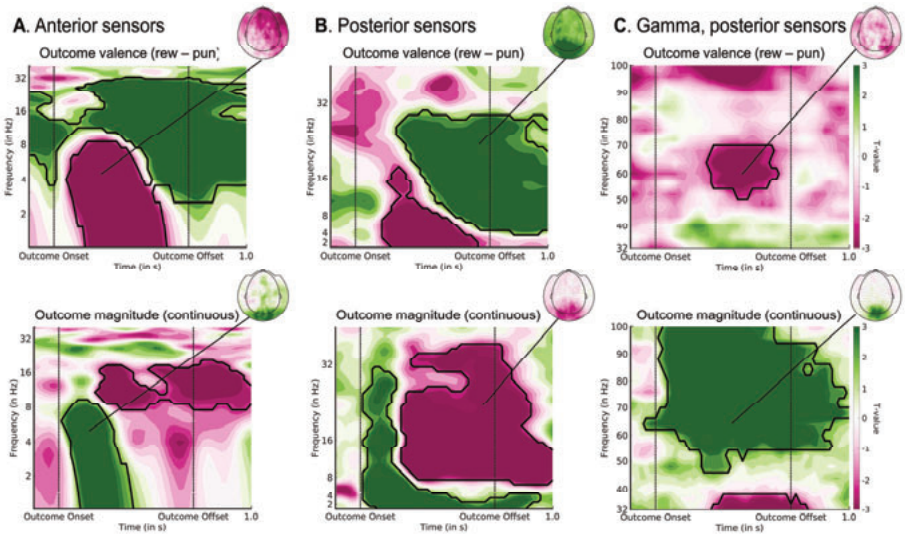


Figure 5.17. S5.5B. Delta/ theta, alpha/ beta, and gamma power modulation by outcome valence and outcome magnitude.
A. Outcome valence was negatively reflected in delta/ theta power (1–8 Hz) at anterior sensors (peak at right frontal sensors). In contrast, outcome magnitude was positively reflected in delta/ theta power at posterior sensors. Results are based on a multiple linear regression containing both outcome valence and outcome magnitude as regressors. Outcome valence reflected the sign of the prediction error, while outcome magnitude reflected the absolute prediction error size. Solid black lines indicate clusters above threshold. **B.** Alpha/ beta (8–33 Hz) power at posterior sensors reflected outcome valence positively, but the outcome magnitude negatively. **C.** Gamma power at posterior sensors reflected outcome valence negatively (52–68 Hz), but outcome magnitude positively (48–100 Hz).

In sum, frontal delta/ theta and parietal gamma power encoded outcome valence negatively, while posterior alpha/ beta power encoded outcome valence positively. The strong signal differences at midline frontal/ occipital sensors were in line with our previous EEG-fMRI findings (Algermissen et al. 2021), but markedly different from the encoding of stakes during the stakes presentation period. These findings ascertain the quality of our data, being able to replicate findings consistently found in previous literature, and also ascertain that similar outcome processing and learning processes took place in this study as in our previous EEG-fMRI study. Outcome magnitude was positively encoded in delta/ theta power and broadband gamma power as well as negatively encoded in alpha/ beta power all at posterior sensors, which might have simply arisen from higher outcome magnitudes being represented by more visual input on the screen.

5.6.6 S5.6: Posterior alpha power modulation by stake valence and magnitude

In addition to top-down influences on attention allocation—such as Go/ NoGo action requirements—we expected that also bottom-up features of the task might attract participants' covert attention and induce alpha power modulation. For example, participants might attend more to the reward stake than the punishment stake (or vice versa) or to the higher stake (i.e., stakes of higher magnitude irrespective of valence) than the low stake. The existence of such modulations would ascertain that alpha power in our task was reflecting participants' attention to features that commonly attract attention. Furthermore, comparing the effects of action plans to modulations by other task factors can be insightful for judging whether non-significant findings might stem for true null effects, suboptimal data quality, or features of alpha modulation that we did not consider in our pre-registered hypotheses. We tested for differences between rewards/ punishments and high/ low stakes appearing on the left or right by performing cluster-based permutation tests on alpha power (8–13 Hz, band-averaged) at left and right posterior (occipital/ parietal) sensors (sensor-averaged). Analyses were performed twice, once on the 36 participants who performed the Go/ NoGo task above chance level and once on the subset of 31 participants who showed saccades on less than 33% of trials. Both analyses led to identical conclusions.

First, we tested for differences in alpha power between trials on which the reward stake appeared on the left vs. trials on which the reward stake appeared on the right side the of the screen. There were no such differences, neither at left posterior sensors (36 participants: $p = .21$; 31 participants: $p = .0.21$; Fig. S5.6A panel A) nor right posterior sensors (36 participants: $p = .082$; 31 participants: $p = .18$; Fig. S5.6A panel B). Taken together, there was no evidence for stake valence mapping (rewards on the left vs. right) affecting alpha power lateralization. However, see S5.8 for effects of stake valence mapping on the dwell times of saccades to the stakes.

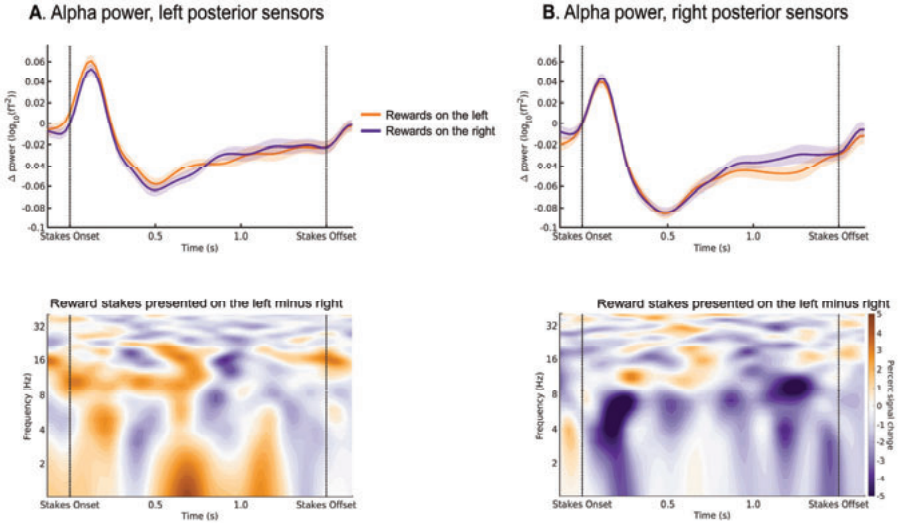


Figure 5.18. S5.6A. Modulation of alpha power at left (A) and right (B) posterior sensors by the stake valence.
Alpha power was not significantly modulated by whether the reward (punishment) stake appeared on the left or right side of the screen.

Second, we tested for differences in alpha power between trials on which higher stakes appeared on the left side vs. trials on which higher stakes appeared on the right side the of the screen. At left posterior sensors, there was no significant difference (both for 36 participants and 31 participants: no clusters above threshold). However, there was a significant difference at right posterior sensors (36 participants: p = .022; 31 participants: p =.016; Fig. S5.6B panel A), with lower alpha power contralateral to high stakes around 1.250–1.500 sec.; Fig. S5.6B panel B). Given that this difference at right sensors was not mirrored by an inverse effect at left sensors, evidence for high stakes attracting attention and modulating posterior alpha remained ambiguous. Notably, modulation of alpha power lateralization by stake magnitudes mapping occurred rather late, i.e., more than one second after stakes onset. Also, modulations occurred not only at (right) occipital, but also parietal and even central sensors. See S5.8 for effects of stake magnitudes mapping on eye-movement dwell times.

Taken together, alpha power was not modulated by the position of the reward/ punishment stake on the screen and only weakly modulated by the position of high/ low stakes. The latter modulation occurred very late, towards the end of the stakes presentation phase. These findings question the assumption that alpha power lateralization was reflecting participants' focus of attention in our task, given that we would have expected stronger modulations. In line with findings about alpha power modulation by Go/ NoGo action plans reported in the main text (see Fig. 5.4), alpha power modulations occurred at right, but not left sensors, and occurred not only at occipital, but also parietal and even central sensors. Hence, the only strong alpha modulation present in our data was the effect of target position in the Posner task, in which case only a single (but not two) lateralized stimulus was presented. Potentially, when participants face two lateralized stimuli, any modulation of alpha power is delayed and rather weak.

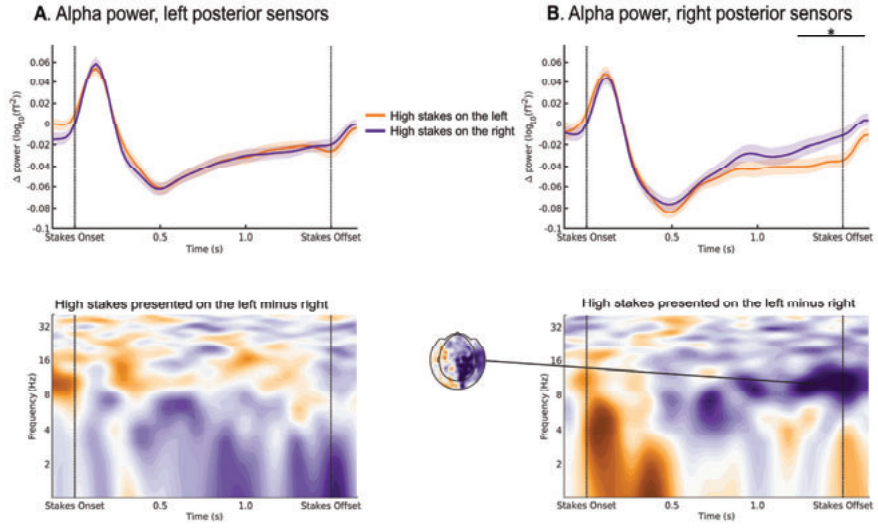


Figure 5.19. S5.6B. Modulation of alpha power at left and right posterior sensors by stake magnitude.
A. Alpha power at left posterior sensors was not significantly modulated by whether the higher (compared to the lower) stake was located on the left (orange line) or the right side (purple line) of the screen. **B.** In contrast, alpha power at right posterior sensors was significantly lower when the high stake appeared on the left (contralateral) side of the screen (cluster above threshold around 1.250–1.500 sec.). The black horizontal line indicates the time range for which the cluster driving significance was above threshold.

5.6.7 S5.7: Occipital gamma power and ERF modulation by stake valence and magnitude

In our pre-registered hypotheses, we tested whether gamma power and ERFs—as indices of bottom-up processing of visual input—were modulated by action plans (see main text). We did observe a trend towards higher gamma power contralateral to stakes that matched participants' action plans, which was however not significant. To assess whether the absence of these effects might derive from a true null effect or rather from suboptimal data quality, it can be insightful to compare the effects of action plans to other effects that should modulate gamma power and ERFs. One such task factor is whether the reward/ punishment stake appeared on the left or right side of the screen—possibly, participants processed rewards more strongly than punishments (or vice versa). Another factor is whether the higher/ lower stake appeared on the left or right side of the scren—higher stakes, coming with more color and contrast on the screen, should induce stronger bottom-up signal. We tested for differences between i) rewards/ punishments appearing on the left/ right and ii) higher/ low stakes appearing on the left/ right using cluster-based permutation tests on gamma power (32–100 Hz, not band-averaged given expectable heterogeneity across the broad gamma band) at left and right occipital sensors (sensor-averaged) in the range of 0–1.5 sec. after stakes onset. Also, we tested for similar differences in ERFs by performing cluster-based permutation tests on the mean time-domain signal at left and right occipital sensors (sensor-averaged) around 0–0.5 sec. after stakes onset.

First, we tested for differences in occipital gamma power between trials on which rewards appeared on the left or right side of the screen. Gamma power at left occipital sensors was significantly higher when *the reward stake* appeared on the right (contralateral) side of the screen (p = .006), with clusters above threshold between 80–96 Hz around 0.250–0.575 sec. and between 72–84 Hz around 1.050–1.500 sec. (Fig. S5.7A panel A). In contrast, there was no corresponding modulation at right occipital sensors (p = .349; Fig. S5.7A panel B). These results provided somewhat ambiguous evidence for whether gamma power lateralization was modulated by the location of the reward/ punishment stake on the screen. Furthermore, neither ERFs at left occipital sensors (p =.142) nor ERFs at right occipital sensors (p = .164) were significantly affected by whether rewards/ punishments appeared on the left or right side of the screen (Fig. S5.7A panel A and B). Visual inspection of the topography of magnetic field differences around 0.1–0.2 sec. suggested that ERFs were somewhat higher contralateral to the side on which the *punishment* stake appeared, which is the opposite to the gamma power modulation reported above. For this observation, see also higher delta/ theta power contralateral to punishments at posterior sensors in Fig. S5.6 panel A.

Taken together, rewards (compared to punishments) appeared to elicit stronger gamma power, but not larger ERF amplitudes.

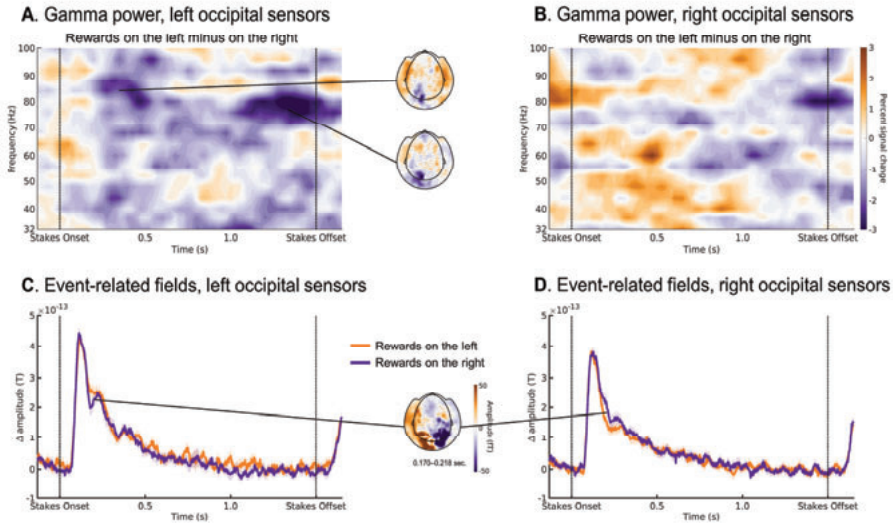


Figure 5.20. S5.7A. Gamma power and ERF modulation at occipital sensors by stake valence.
A. Gamma power at left occipital sensors was significantly higher when the reward stake appeared on the right (contralateral) compared to the left (ipsilateral) side, with clusters above threshold around between 80–96 Hz around 0.0250–0.575 sec. and between 72–84 Hz around 1.050–1.500 sec. **B.** In contrast, gamma power at right occipital sensors was not significantly modulated by whether rewards appeared on the left or right side of the screen. Furthermore, ERFs at neither left (**C**) nor right (**D**) occipital sensors were affected by whether the reward stake appeared on the left (orange line) or right (purple line) of the screen. Numerically, it appeared that ERFs tended to be higher contralateral to the side on which the punishment stake appeared.

Second, we tested for differences in occipital gamma power between trials on which higher stakes appeared on the left vs. right side of the screen. Gamma power at left occipital sensors was significantly higher when higher stakes appeared on the right (contralateral) side of the screen (p < .001), with clusters above threshold between 64–100 Hz around 0–0.850 sec. and between 60–72 Hz around 1.275–1.500 sec. (Fig. S5.7B panel A). In contrast, there was no corresponding modulation at right occipital sensors (p = .38; Fig. S5.7B panel B). These results provided somewhat ambiguous evidence for whether gamma power lateralization was modulated by high vs. low stakes. In contrast, ERFs were clearly stronger contralaterally to the side at which high stakes appeared: ERFs at left occipital sensors were higher when higher stakes appeared on the right (contralateral) vs. left side of the screen (p =.004, cluster above threshold around 0.088–0.148 sec.; Fig. S5.7 panel C). Vice versa, ERFs at right occipital sensors were higher when higher stakes appeared on the left (contralateral) vs. right side of the screen (p = 0.002; cluster above threshold around 0.085–0.178 seconds; Fig. S5.7 panel D). Taken together, ERFs were clearly stronger contralaterally to the side at which high stakes appeared, while gamma power was only higher contralaterally to high stakes appearing at left sensors. These modulations might plausibly reflect the fact that higher stakes came with more visual input (e.g., color, contrast) being presented contralateral to the respective sensors.

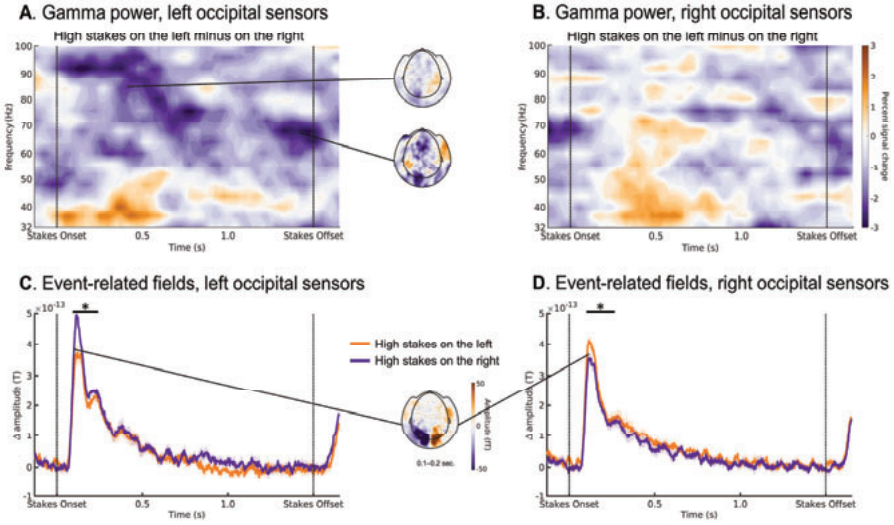


Figure 5.21. S5.7B. Gamma power and ERF modulation at occipital sensors by stake magnitude.
A. Gamma power at left occipital sensors was significantly higher when higher stakes appeared on the right (contralateral) compared to the left (ipsilateral) side of the screen, with clusters above threshold around between 64–100 Hz around 0–0.850 sec. and between 60–72 Hz around 1.275–1.500 sec. **B.** In contrast, gamma power at right occipital sensors was not significantly modulated by whether higher stakes appeared on the left or right side of the screen. **C.** ERFs at left occipital sensors were significantly stronger when higher stakes appeared on the right (contralateral, purple line) compared to the left (ipsilateral, orange line) side, with a cluster above threshold around 0.088–0.148 sec. **D.** ERFs at right occipital sensors were significantly stronger when higher stakes appeared on the left (contralateral) compared to the right (ipsilateral) side of the screen, with a cluster above threshold around 0.085–0.178 sec.

Taken together, gamma power at left, but not at right occipital sensors appeared to be modulated by the mapping of the stake valence (rewards vs. punishments) and the stake magnitude (higher vs. lower stakes), with higher gamma power at left sensors when rewards or higher stakes appeared on the right side of the screen. In both cases, differences in gamma power were focused on the higher gamma band (> 60 Hz) and occurred in two separate time windows, one around 0–0.800 sec. and one around 1.250–1.500 sec. The fact that gamma power only at left, but not at right occipital sensors was significantly modulated by task factors is puzzling and stands is contrast to alpha power modulations, which tended to occur selectively at right sensors (see Fig. S5.6B). However, the fact that modulations reoccurred in the last 500 ms of the stakes presentation phase is consistent with alpha modulations (see Fig. S5.6 panel B) as well as frontotemporal delta power modulations by reward and punishment stake magnitudes in the same time window (see Fig. 5.3C, D). One could speculate that stakes processing occurs in two phases, an early one that is only visible in the gamma band, followed by later one also visible in posterior alpha and frontotemporal delta power. These findings suggested that stakes encoding was not finished after the first few hundred milliseconds of the stakes presentation phase, but continued until stakes disappeared again. Although we were primarily interested in early gamma power reflecting (almost) pure bottom-up processing, it might be interesting to test whether alternatively, gamma power at later time points reflected participants' action plans. Overall, these findings ascertained that gamma power could be modulated by task factors, suggesting that any modulation by action plans was considerably weaker or non-existent.

Furthermore, ERFs at both left and right occipital sensors were significantly stronger contralaterally to the side on which higher (compared) to lower stakes appeared. Modulation of ERFs by the side on which rewards/ punishments appeared was considerably weaker and not significant. ERFs contralaterally to punishments tended to be higher than ERFs contralaterally to rewards. These findings ascertained that ERFs could be modulated by task factors, suggesting that any modulation by action plans was considerably weaker or non-existent.

5.6.8 S5.8: Saccade data quality and modulation by task factors

Of the 36 participants included in analyses, only 25 participants had eye-tracking data recorded for every trial. For the other eleven participants, calibration was lost on (at least some) trials. For one participant, no eye-tracking data at all was recorded. For the 35 participants with at least some data, on average, data was missing for 20 trials (M = 20.43, SD = 0.18, range 0–176).

Of the 35 participants included with any eye-tracking data, on average, saccades occurred on 45 trials (M = 45.03, SD = 60.17, range 0–235; Fig. S5.8 panel A). Ten participants showed saccades to at least one of the stakes on more than 20% of the trials, five participants on more than 33% on the trials, and three participants on more than 50% of trials. Analyses on alpha power were performed both with and without the five participants with saccades on at least 33% of trials. In the 30 participants left after excluding those five participants, on average, saccades occurred on 25 trials (M = 24.70, SD = 24.91, range 0–75) and lasted on average 10 ms per trial (M = 10.22, SD = 14.12, range 0 – 46.65).

Overall, dwell times tended to be longer for i) the *right stake* (mixed-effect linear regression with dwell time difference on left minus right stakes as dependent variable and only an intercept as independent variable; $t(26.72) = -2.27$, $p = .031$; Fig. S5.8 panel B), ii) the *reward stake* (mixed-effect linear regression with dwell time difference on the reward minus the punishment stake as dependent variable and only an intercept as independent variable; $t(29.64) = 4.26$, $p < .001$; Fig. S5.8 panel C), and iii) the *high stake* (mixed-effect linear regression with dwell time difference on high minus low stakes as dependent variable and only an intercept as independent variable; $t(29.31) = 2.69$, $p = .012$; Fig. S5.8 panel D). However, none of these effects was strong, suggesting that eye-movements were not overly systematic.

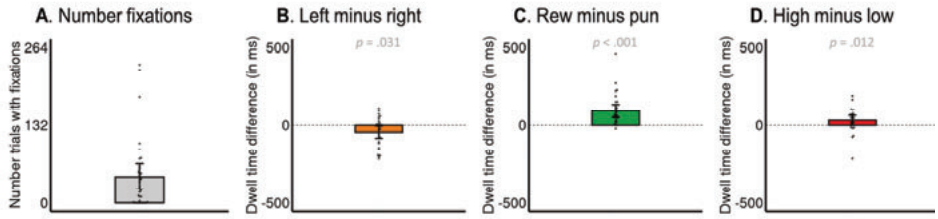


Figure 5.22. S5.8. Systematic biases in overt saccades.
A. Number of trials featuring (at least) one fixation to one of the stakes per participant. The individual data points display the number of trials with fixations for each participant, bar height displays the mean trial count. Out of the 36 participants who learned the task and were included in the analyses, 27 participants showed a saccade on at least one trial. Eye-tracking data from one participant was completely lost and data was incomplete for another ten participants. **B.** Overall, participants' dwell times showed significant bias towards the right stake. The individual data points display a mean dwell time difference for each participant, bar height displays the mean across participants. **C.** Overall, participants' dwell times showed a strongly significant bias towards the reward stake. **D.** Overall, participants' participants' dwell times showed a significant bias towards the respectively higher stake.

Taken together, participants performed impulsive saccades on a subset of trials although they were instructed to maintain fixation at the center of the screen. Saccades exhibited systematic biases towards right stakes, reward stakes, and high stakes. Note that unlike our previous eye-tracking study, participants were instructed to keep fixation at the central fixation cross, resulting in only relatively few saccades that were available for analyses. The analyses reported here had thus substantially lower statistical power compared to our previous eye-tracking study.

Chapter 6

General Discussion

6 GENERAL DISCUSSION

6.1 AIMS OF THIS THESIS

The two key aims of this thesis were (a) to better understand the **neural origin of Pavlovian biases in learning and decision-making** (chapters 2, 3, and 5), and (b) to test whether these biases **can be adaptively recruited to support goal-directed behavior** (chapters 4 and 5). Phenomena of "weakness of will", i.e., humans not acting in line with their best intentions, have often been explained by postulating different decision-making systems that compete for control over behavior. The Pavlovian system might be one such system, which is characterized by its fast and automatic, but also inflexible way of invigorating/ inhibiting actions in response to environmental cues that signal reward/ punishment availability. The seemingly automatic link between reward/ punishment cues and Go/ NoGo actions is termed "Pavlovian biases". While the Pavlovian system might provide sensible "priors" on actions in novel or seemingly uncontrollable environments, its pervasive influence on action selection even in well-known circumstances remains puzzling. The prevalence of Pavlovian biases across many species suggests a potentially ancient mechanism that is shared across the animal realm. However, the exact neural structures giving rise to Pavlovian biases remain elusive. Similarly, it is not yet clear how humans can suppress Pavlovian biases in situations in which these are maladaptive. Finally, it is not clear whether instrumental and Pavlovian action selection systems operate independently of each other or whether, alternatively, instrumental systems might be able to "recruit" the Pavlovian system to make both pursue the very same goal.

In this final chapter, I will summarize the questions posed at the outset of this thesis and how the empirical studies that I conducted begin to answer them. I will first briefly summarize the main findings of each empirical chapter. Next, I will discuss these findings in the context of the main aims of this thesis, evaluating how my findings shed new light on the neural origin of Pavlovian response and learning biases as well as how humans could benefit from a strong Pavlovian system when it is synchronized to their action plans. I will also briefly consider limitations and caveats that should be addressed by future research. Afterwards, I will consider the broader implications of Pavlovian biases in explaining various "motivational" phenomena. First, I will point out how Pavlovian biases may explain other decision anomalies such as framing effects as well as the effects on incentives on cognitive control recruitment. Second, I will discuss how Pavlovian biases can shed light on whether rewards and punishments are polar opposites on a unifying dimension or distinct categories encoded in separate brain regions. Third, I will discuss ways in which Pavlovian biases could be used to motivate and invigorate other behaviors as well as factors that are crucial for setting the balance between Pavlovian biases and other decision-making systems. As a general conclusion, I will draw implications for research on motivation, learning, and decision-making, more generally, for the understanding of the etiology and maintenance of psychiatric disorders, and for the design of environments that allow agents to effectively pursue their goals. I will close by reconsidering Pavlovian bias as a limiting as well as an empowering factor in instrumental goal pursuit.

6.2 MAIN FINDINGS

In Chapter 2, I used simultaneous EEG-fMRI recordings to investigate the origin of Pavlovian response biases as well as how these biases can be suppressed by top-down control. Prominent theories of the basal ganglia pathways have previously suggested that prediction errors about the

availability of rewards and punishments are sent from the dopaminergic midbrain (ventral tegmental area, VTA, and substantia nigra pars compacta, SNc) to the striatum, where they have differential consequences for the facilitation vs. suppression of action release. In line with these theories, I expected that striatal BOLD signal would encode state prediction errors (i.e., the cue valence signaling reward/ punishment availability), which would give rise to Pavlovian biases in behavior. Furthermore, I expected increased midfrontal theta power on trials on which Pavlovian biases were successfully suppressed in favor of alternative actions that were more likely to lead to the desired outcome. Lastly, I aimed to shed light on the trial-by-trial relationship between striatal BOLD signal and midfrontal theta power, expecting that stronger theta power on a trial-by-trial basis would predict attenuation of striatal value signals. Linking cortical conflict detection signals to subcortical value and action selection signals promised to shed further light on the exact cortico-subcortical interactions by which humans are able to suppress Pavlovian biases.

Results did not confirm any of these hypotheses. Firstly, while participants exhibited Pavlovian biases in behavior, cue valence was not consistently reflected in the striatal BOLD signal, which was instead dominated by the performed Go/ NoGo actions. These results replicated previous work, but left open which areas would encode cue valence and putatively bias striatal action selection. Cue valence was encoded in the BOLD signal of several prefrontal cortical (PFC) regions, including ventromedial prefrontal cortex (vmPFC), anterior cingulate cortex (ACC), and posterior cingulate cortex (PCC), as well as in subcortical regions such as the amygdala and the hippocampus. Trial-by-trial BOLD signal modulations in vmPFC and ACC predicted reaction times, consistent with a putative role of these regions mediating the invigorating effect of prospective outcomes on behavior. Secondly, not midfrontal theta power, but transient increases in midfrontal alpha power indexed the successful inhibition of Pavlovian biases. Similar to the striatal BOLD signal, theta power was instead strongly dominated by the performed Go/ NoGo action, and more specifically showed signatures of an evidence accumulation processes selectively accruing support for performing a Go action. Thirdly, striatal BOLD signal and midfrontal theta power were not only both dominated by the performed action, but also coupled on a trial-by-trial basis: the striatal BOLD signal predicted theta power over and above signals from other regions such as motor cortices or ACC. This relationship occurred close to the time at which participants made their response, suggesting the involvement of the striatum mostly at late stages when selecting the eventual response. In contrast, theta power at an earlier time point—shortly after cue onset—was significantly correlated with the vmPFC BOLD signal, which reflected the cue valence. Taken together, I propose that these signals are in line with a model in which the vmPFC encodes the cue valence following cue presentation, which then drives Pavlovian biases in striatal action selection processes leading up to the eventual response. Midfrontal theta power might give online insights into the unfolding action selection process.

In Chapter 3, I used simultaneous EEG-fMRI recordings to investigate the neural origin of Pavlovian biases in learning. Previous studies have described such biases in learning, in which rewards are preferentially attributed to Go actions, while punishments are only reluctantly attributed to NoGo actions. In particular, I tested whether striatal BOLD signal was better described by prediction errors incorporating such a learning bias compared to prediction errors that did not. This hypothesis derived from computational models of the asymmetric basal ganglia architecture that predict the presence of such learning biases. Additionally, I tested whether the BOLD signal in PFC regions also reflected biased prediction errors. I then quantified the relative timing of such learning signals by correlating the BOLD signal in regions that showed signatures of biased learning with the EEG signal, assessing putative prefrontal influences on the striatal

learning processes. Computational reinforcement learning models showed that participants' behavior was better described by a combination of Pavlovian response and learning biases than by either bias alone. In line with my hypotheses, I observed that the BOLD signal in the striatum, but also in various cortical regions including dorsal ACC (dACC), perigenual ACC (pgACC), and PCC was significantly better described by biased prediction errors than by "standard" prediction errors. In addition, midfrontal theta/ delta power significantly correlated with biased prediction errors, while the correlation with standard prediction errors was not significant. Putting these observations together, EEG correlates of prefrontal regions (most notably dACC and PCC) preceded EEG correlates of the striatal signal, in line with a mechanism in which early PFC signals bias striatal learning signals. Taken together, the results presented in this chapter are consistent with the idea of prefrontal cortical contributions to biased learning in the striatum.

In Chapter 4, I propose that a strong Pavlovian system might be adaptive not only in providing "priors" on actions in novel or uncontrollable environments, but additionally may support the instrumental system in reaching its action goals. When humans plan a Go/ NoGo action, they might preferably attend to incidental reward/ punishment cues in the environment that activate the Pavlovian system and, in this way, trigger a Go/ NoGo action in line with their action plans. Under this novel hypothesis, biases do not steer the "direction" of action selection, but once an action plan is formed, they are recruited to support the implementation of it. To test this hypothesis, I developed a novel version of the Go/NoGo Motivational Learning Task in which action selection and action execution were separated by an intermediate phase. In this phase, participants could pre-view potential reward and punishment "stakes" (of varying magnitude) that would be received for correct/ incorrect actions. I tested this idea in two independent samples using a gaze-contingent eye-tracking set-up. In this set-up, stakes were occluded and only rendered visible once participants actively fixated them, rendering the very first fixation a proxy of "pure" top-down processes uncontaminated by bottom-up input. I expected that participants' first fixations were more likely targeted at the reward (compared to the punishment) stake on trials which required a Go action compared to trials which required a NoGo action. Vice versa, I expected that total dwell time on reward compared to punishment stakes—indicative of the total amount of attention each stake received—would predict participants' eventual responses.

Participants achieved high performance on this novel paradigm while their Go/ NoGo choices were sensitive to the magnitude of reward and punishment stakes, reflecting Pavlovian biases. In line with my hypothesis on the adaptive recruitment of biases, cues requiring a Go response drove first fixations towards reward stakes more frequently than NoGo cues. Furthermore, the difference in dwell time on reward and punishment stakes strongly predicted responses, with relatively more attention to rewards invigorating Go responding. Exploratory analyses suggested that that those participants who were more sensitive to stake magnitude showed overall worse task performance. In contrast, participants who exhibited a stronger link between attention to the stakes and their eventual responses showed higher task performance. In sum, the findings presented in this chapter suggest that Go/ NoGo action plans do indeed inform participants' attention to incidental reward and punishment cues. Vice versa, attention to these cues activated Pavlovian biases that automatically triggered the action that was in line with participants' action plans. Stronger reliance on these biases was conducive to task performance. These results shed novel light on the putatively adaptive nature of strong Pavlovian biases: when synchronized with instrumental action plans, these biases can ensure the implementation of these plans in an automatic, less effortful and more robust manner.

Lastly, in Chapter 5, I used MEG recordings to investigate the potential neural mechanisms by which evolving Go/ NoGo action plans can influence (covert) attention to incidental reward and punishments and thus recruit Pavlovian biases to aid instrumental control. Button presses and reaction times only give limited insight into the evolving action preparation processes. Similarly, fixation times are only an imperfect proxy of how much processing a cue eventually received. MEG recordings allowed to track both evolving action preparation and covert spatial attention in real time. Specifically, I used beta power desynchronization over central sensors as an index of participants' latent action plans. Furthermore, I used alpha power lateralization over posterior sensors as an index of participants' covert attention allocation to reward and punishment stakes appearing on the left and right side of the screen. I tested six pre-registered hypotheses on how different markers of action preparation could affect attention allocation as indexed in alpha power lateralization and, vice versa, how alpha power lateralization might shape ongoing action preparation and eventual responses. Participants again learned the task successfully while their behavior was affected by the magnitude of reward and punishment stakes in the fashion of Pavlovian biases. Furthermore, beta power desynchronization was stronger for eventual Go than NoGo responses, starting immediately upon presentation of the Go/NoGo cue and thus several seconds before response onset. Beta power desynchronization was transiently disrupted by a high punishment stake, which led to a short period of resynchronization. Stake magnitudes were represented in frontotemporal delta power. The eye-tracking findings from Chapter 4 were replicated, such that participants tended to focus more on the reward stake under a Go compared a NoGo action plan and that, vice versa, attention to the stakes predicted the eventual action. However, there was no evidence for action plans (nor any other task factor) significantly affecting alpha power lateralization. Taken together, I find evidence for beta power indexing latent action plans and the effect of stake magnitudes on responses. Furthermore, I find evidence for overt gaze behavior, but not covert attention as indexed by alpha power lateralization to reflect participants' action plans.

6.3 INTERPRETATION OF THE FINDINGS

6.3.1 The origin of Pavlovian biases

6.3.1.1 Prefrontal and amygdalar inferences on outcome availability

During the completion of this thesis, theoretical ideas about the role of striatal dopamine release in motivating and invigorating behavior have received significant updates. I started out this work with the assumption that transitioning from a neutral state (e.g., an inter-trial interval) to a state in which either rewards (Win cues) or punishments (Avoid cues) are available would induce a ***state prediction error***, leading to a phasic burst/ pause in the firing of dopaminergic neurons in the midbrain (ventral tegmental area; VTA). This reward prediction signal would then be broadcasted to the striatum and shift the activation balance between direct ("Go") and indirect ("NoGo") pathways (Frank 2005; Collins and Frank 2014). The BOLD recordings presented in Chapter 2 are not consistent with such state prediction errors: striatal BOLD signal did not (consistently) distinguish Win and Avoid cues. Hence, this data does not support the notion that dopaminergic state prediction errors from the VTA sent to the striatum give rise to Pavlovian biases in behavior.

In addition to the work presented in this thesis, recent animal studies have called the idea of dopaminergic state prediction errors induced by reward-predictive cues into question. Studies using microdialysis or fast-scan cyclic voltammetry to measure dopamine concentration in the

striatum have observed **"ramps" in dopamine concentration as animals approach a reward** (Phillips et al. 2003; Wassum et al. 2012; Howe et al. 2013; Hamid et al. 2016; Mohebi et al. 2019). However, such a ramping signal has not been observed in the spiking of dopaminergic neurons, neither in the nucleus accumbens (Eshel et al. 2016) nor upstream in the ventral tegmental area (Ikemoto 2007), suggesting that dopamine concentration during the approach of reward-predictive cues does not reflect state prediction errors originating from the VTA. The apparent conflict between results from different measurement techniques (spike counts measured with electrophysiology vs. measures of dopamine concentration) has led to the recent suggestion that **dopamine ramps are not caused by prediction error signals from the midbrain, but arise *locally* by cholinergic interneurons modulating the dopamine release from axonal terminals independently of cell body spiking** (Berke 2018; Liu et al. 2022). One candidate region that might induce such dopamine ramps in the striatum is the ***basolateral amygdala***, which can affect dopamine levels in the nucleus accumbens even under deactivation of the ventral tegmental area (Floresco et al. 1998; Jones et al. 2010). Another candidate region is the ***vmPFC***, which sends information to the striatum both via direct projections (Haber et al. 1995; Ferry et al. 2000; Clarke et al. 2014; Keistler et al. 2015, 2017; Hamel et al. 2022)—potentially mediated by the above mentioned cholinergic interneurons directly onto the axons of striatal neurons (Stalnaker et al. 2016; Liu et al. 2022)—as well as via indirect projections through the amygdala (Ambroggi et al. 2008; Belin et al. 2009). In conclusion, the effects of reward/ punishment availability on invigoration/ behavior might not arise from state prediction errors signaled from the midbrain to the striatum. Instead, inputs from vmPFC and amygdala might convey information about cue valence that is necessary for the emergence of Pavlovian biases.

Predicting whether and how much reward and punishment is available in a given state is a different computational problem than deciding which action maximizes the chance of getting the desired outcome. The former problem, termed ***latent state inference***, has typically been attributed to circuits in vmPFC (Wilson et al. 2014; Schuck et al. 2016; Hunt et al. 2018; Zhou et al. 2021) and in the amygdala (Schoenbaum et al. 1998; Mollick et al. 2020). Latent state inference refers to the use of sensory cues to infer in which categorical state an agent is in, and, based on the state identity, assess what kind of rewards and punishments are available. This putative dependence on amygdalar and vmPFC inputs can explain why, in Chapter 2, I did not observe striatal BOLD to (consistently) distinguish Win and Avoid cues, but instead, cue valence was encoded by vmPFC, ACC, and hippocampus/ amygdala. Relatedly, in Chapter 5, I observed frontotemporal delta power to encode both reward and punishment stakes parametrically with opposite signs. I speculate that these power modulations also reflect signals from vmPFC or amygdala/ hippocampus. Perhaps then, signals from vmPFC and amygdala drive Pavlovian biases, as their computational role is the **inference which latent state an agent is in and which outcomes are available in this state.**

Further evidence that state representations in vmPFC might even incorporate "irrelevant" reward information comes from studies on *decoy* effects, i.e., unavailable reward options distorting the choice between two available reward options. Decoy value has been found reflected in vmPFC BOLD signal (De Martino et al. 2006; Lebreton et al. 2009; Chau et al. 2014), suggesting that the vmPFC does not just encode choice-relevant rewards, but also seemingly task-irrelevant information. Preliminary evidence suggests that patients with vmPFC lesions are less subject to the influence of such irrelevant reward options and make more "rational" decisions (i.e., decisions more in line with expected value, returning higher earnings) (Manohar and Husain 2016; Manohar et al. 2021). These lines of evidence support the idea that the vmPFC even represents

information—such as cue valence in the context of the MGNG Task—that is irrelevant for performing the task and potentially even biases behavior.

The fact that seemingly task-irrelevant information is not filtered out at early processing stages, but retained into later stages—not just in PFC (Mante et al. 2013), but also in higher-order visual cortex (Hong et al. 2016) and even motor cortex (Takagi et al. 2021)—appears to be a general principle of neural processing (Yoo and Hayden 2018). Brains might be shaped to integrate a large range of situational features into an eventual action by not just muting certain features, but adjusting their weights in a graded, continuous fashion. In such an architecture, phenomena like Pavlovian biases can occur when features are rendered irrelevant by experimental design, but still considered by PFC inference processes. In more naturalistic conditions, it might however be adaptive to consider such seemingly irrelevant features in one's decisions—not least to be able to spontaneously adjust one's behavior once these features (e.g., the presence of a potential predator) are suddenly rendered more important. In conclusion, irrelevant aspects of the state an agent is in should reasonably be represented (e.g., in vmPFC) in case they suddenly become relevant.

In sum, the findings presented in Chapter 2 of this thesis are consistent with vmPFC, ACC, amygdala, and hippocampus encoding cue valence and then forwarding this information to the striatum where it biases action selection and gives rise to Pavlovian biases in behavior. Notably, this information might not arrive at the striatum via state prediction errors arising from midbrain inputs, but instead via separate inputs from vmPFC and/ or amygdala that project directly onto the axons of striatal neurons.

6.3.1.2 The role of the striatum in Pavlovian biases

If the vmPFC and amygdala are responsible for inferring the latent task state—including reward/ punishment availability—and in this way causally contribute to the emergence of Pavlovian biases, the specific contribution of the striatum to these biases is unclear. In Chapter 2, I replicate previous studies (Guitart-Masip, Fuentemilla, et al. 2011; Guitart-Masip, Chowdhury, et al. 2012; Guitart-Masip, Huys, et al. 2012; Moutoussis, Rutledge, et al. 2018) that striatal BOLD signal primarily reflects the action that is performed. The considerable contribution of (motor) actions to the BOLD signal might be overlooked in designs that exclusively feature Go responses (e.g., button presses). In contrast, in the MGNG Task used in Chapter 2, Go and NoGo cues are equally prevalent and orthogonal to cue valence, providing unique data for uncovering the strong effects of actions on the BOLD signal. Such findings align well with recent evidence that large parts of the brain (Musall, Kaufman, et al. 2019; Steinmetz et al. 2019; Stringer et al. 2019), including even visual cortex (Gutteling et al. 2015; Gallivan et al. 2019), are strongly dominated by action-related signals—which might not be surprising given theoretical models emphasizing that the primary purpose of the brain is action selection and the control of muscles (Yoo and Hayden 2018; Cisek 2019, 2020; Fine and Hayden 2021).

Apart from a role in learning from *obtained* rewards, the role of striatal dopamine levels in motivating and invigorating behavior as a function of *expected* rewards is well established. There is strong evidence for the involvement of the striatum—especially its ventral part, the nucleus accumbens—in Pavlovian effects in Pavlovian-to-Instrumental Transfer (PIT), both in animals (Corbit and Balleine 2011; Flagel and Robinson 2017) and humans (Bray et al. 2008; Talmi et al. 2008; Geurts et al. 2013b; Pool et al. 2022). Similarly, effects of dopamine agonists on behavioral activation and motivation (Taylor and Robbins 1984; Wyvell and Berridge 2000; Peciña and Berridge 2013; Halbout et al. 2019) and effects of dopamine antagonists on behavioral inhibition (Dickinson et al. 2000; Corbit et al. 2007; Lex and Hauber 2008; Wassum et al. 2011; Ostlund and

Maidment 2012) are well established. These findings support the notion that the striatum is involved in mediating the effect of reward prospect on action invigoration. However, when information about reward and punishment availability arrives from vmPFC and amygdala rather than from the dopaminergic midbrain, the exact role of dopamine in the invigoration vs. suppression of behavior remains unclear. New light on dopamine dynamics during action selection has come from animal studies showing that, during reward pursuit, striatal dopamine levels do not reflect reward prediction errors, but instead ramp up until a course of action is completed (Howe et al. 2013; Hamid et al. 2016; Mohebi et al. 2019). These ramps likely do not arise through inputs from the midbrain, but through cholinergic interneurons transmitting information from vmPFC and amygdala (Stalnaker et al. 2016; Liu et al. 2022). These ramping dopamine levels have been interpreted as reflecting the "***value of work***", i.e., the marginal increase in expected rewards when investing an additional unit of effort (Hamid et al. 2016; Berke 2018; Walton and Bouret 2018).

This "value of work" perspective on striatal dopamine levels during action selection can potentially explain why I observed slightly higher striatal BOLD signal on Go2Avoid than Go2Win trials. On Go2Avoid trials, an extra "unit" of effort recruited will more likely make a difference in selecting the correct response than it will on Go2Win trials, on which reward prospects already ensure Go action selection through Pavlovian mechanisms. Whether the striatum encodes value per se or the value of effort in particular is hard to disentangle with standard paradigms that typically couple higher effort investment to higher obtained rewards (Knutson et al. 2001). Some attempts to de-confound expected rewards and effort investment (i.e., task difficulty) have found the dorsal striatum (caudate and putamen) in particular to encode effort levels rather than reward levels (Miller et al. 2014). The MGNG Task used in Chapter 2 provides another avenue to de-confound reward and effort levels because highest effort is required by active punishment avoidance. Given that PIT paradigms typically do not include aversive outcomes (Bray et al. 2008; Talmi et al. 2008; Pool et al. 2022), the MGNG Task used in this thesis provides a unique opportunity to test the value of work hypothesis of the striatum. Indeed, the findings presented in Chapter 2 can provide evidence for the idea of the striatum specifically encoding the value of work.

In conclusion, it seems **more likely that Pavlovian biases in the context of the MGNG Task arise from vmPFC or amygdala inputs biasing striatal action selection processes rather than from state prediction errors transmitted from the midbrain**. This conclusion does not imply the stronger statement that midbrain or striatum would never encode state prediction errors. In fact, there is recent evidence from direct recordings in animals demonstrating that midbrain and striatal dopamine encode state prediction errors in the context of sensory preconditioning (Cerri et al. 2014; Sharpe et al. 2017) and state transitions under uncertainty (Starkweather et al. 2017; Mikhael et al. 2022). In fact, it is currently still under debate whether dopamine ramps might not eventually be reducible to prediction errors signals arising from the midbrain (Kim et al. 2020; Mikhael et al. 2022). However, the BOLD recordings described in Chapter 2 do not provide evidence for such state prediction errors. Instead, they support a more dominant role of vmPFC and amygdala. Furthermore, the presented findings are more in line with striatal dopamine levels encoding the value of work instead of value per se, further refining theories on the exact computation the striatum is performing during action selection.

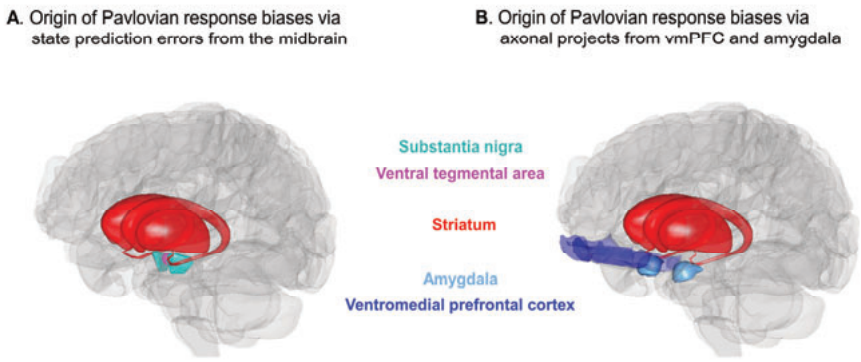


Figure 6.1. Core regions involved in the initial and revised model of the neural origin of Pavlovian response biases.
A. Model used when planning the research presented in Chapter 2. Pavlovian response biases arise though dopaminergic signals from the midbrain (ventral tegmental area, VTA, and substantia nigra pars compacta, SNc) signaling reward/ punishment availability (in the MGNG Task: cue valence) in the form of state prediction errors to the striatum. **B.** Revised model in light of the findings presented in Chapter 2 as well as recent animal findings. Ventromedial prefrontal cortex (vmPFC) and amygdala are involved in inferring the latent state an agent is in as well as reward/ punishment availability in this state. This information is that transmitted to the striatum via cholinergic interneurons that circumvent the striatal cell bodies and directly project on the cell axons. Images created using Anatomography, a website maintained by the Life Science Databases (LSDB), under CC-BY-SA-2.1-jp.

6.3.1.3 Midfrontal theta power as a reflection of evolving action selection and response conflict

In Chapter 2, beyond investigating possible neural origins of Pavlovian response biases, I also tested how these biases are suppressed in situations in which they are maladaptive (i.e., incongruent trials). Previous research has observed increases in midfrontal theta power when participants successfully suppressed Pavlovian biases (Cavanagh et al. 2013; Swart et al. 2018). I did not replicate these previous findings; instead, I found power in lower frequency ranges (1–15 Hz, with a peak in the theta range) to be much higher when participants performed a Go compared to a NoGo response. Notably, I replicated this Go/ NoGo difference in midfrontal theta power in the MEG data presented in Chapter 5. These findings stand in conflict with previous theories suggesting a role of midfrontal theta power increases in response conflict, error, surprise, and other aversive events (Cavanagh, Zambrano-Vazquez, et al. 2012; Cavanagh and Frank 2014; Cohen 2014). By linking theta power to the BOLD signal in different regions, I gained new insights into the potential sources of theta and the kind of processes it might reflect.

Similar to striatal BOLD signal, also midfrontal (and parieto-central) theta power was strongly dominated by the action (Go vs. NoGo) that participants performed. This strong modulation by the performed action is reminiscent of other recent findings which showed that signals across the brain are dominated by action-related information (Musall, Kaufman, et al. 2019; Steinmetz et al. 2019; Stringer et al. 2019). The fact that most task designs in cognitive neuroscience research only feature active Go responses (e.g., button presses) might have occluded these large action-related signal increases. Curiously, however, such a Go/ NoGo difference was not observed in previous EEG studies that used a highly similar task (Cavanagh et al. 2013; Swart et al. 2018). Instead, Swart and colleagues even reported particularly high theta power for NoGo responses to Win cues. One potential explanation for resolving the apparent discrepancy between these and my findings is that, in this previous study, participants by default prepared a Go response on every trial and inhibited

this response once they realized that NoGo was the more appropriate response. Such a strategy is typical for more classical Go/ NoGo task designs in which Go cues are much more frequent (e.g., 80%) than NoGo cues (20%) (Wessel 2018a). In such unbalanced designs, infrequent NoGo trials indeed induce increased theta power (Huster et al. 2013). Although previous research on the MGNG Task used equiprobable design with equal numbers of Go and NoGo cues (Cavanagh et al. 2013; Swart et al. 2018), it is possible that (at least some) participants adopted a "Go default" response strategy. An important discrepancy between the implementations of the (otherwise identical) MGNG Task in the EEG study by Swart and colleagues and the EEG-fMRI study presented in Chapters 2 and 3 is that the latter featured considerably longer trial durations to allow for a better dissociation of different BOLD signals. Under such a slower task design with more "dead time", participants might be less likely to adopt a default Go strategy, leading to a cleaner dissociation between Go and NoGo responses in midfrontal theta signal in the current study.

Still, one might wonder how the findings in Chapter 2 can be reconciled with previous studies finding elevated midfrontal theta signal under response conflict. While earlier research assumed midfrontal theta to be one domain-general signal reflecting response adjustments after conflict, errors, or negative feedback (Cavanagh, Zambrano-Vazquez, et al. 2012; Cavanagh and Frank 2014; Cohen 2014), more recent research has suggested the existence of different, spatially overlapping midfrontal theta signals (Töllner et al. 2017; Zuure et al. 2020). It might be the case that some (centro-parietal) theta signals reflect Go vs. NoGo actions while other (more midfrontal) signals reflect cognitive conflict. In Chapter 2, two findings indeed revealed signatures of conflict (or cognitive effort recruitment) in theta power: Firstly, theta power was higher for responses of participants' non-dominant (left) hand compared to the dominant (right) hand, and secondly, theta power was higher for correct compared to incorrect Go responses. Hence, also the data presented in Chapter 2 appears to provide some support for theta reflecting processes related to response conflict resolution.

The traditional interpretation of midfrontal theta power increases in situations of response conflict is that processes in midfrontal cortex detect a conflict between two active Go responses (e.g., left vs. right in the Simon task), visible in increased theta power over the scalp, and then broadcast a conflict signal via the "hyperdirect" pathway to the subthalamic nucleus (Frank 2006; Wiecki and Frank 2013). The subthalamic nucleus then elevates response thresholds—leading to higher "response caution"—which yields additional time for the more appropriate response to take over (Frank 2006; Cavanagh et al. 2011; Wiecki and Frank 2013). Although this idea is well supported by studies linking scalp EEG recordings to intracranial recordings of subthalamic nucleus activity (Cavanagh et al. 2011; Zavala et al. 2014; Herz et al. 2016), it seemingly only allows for the arbitration between two Go responses. Such a mechanism could boost a Go response on Go2Avoid trials in the MGNG Task, but it is unclear how the same mechanism would allow for boosting an appropriate NoGo response to take over from an inappropriate Go response on NoGo2Win trials. In particular, the question arises whether the brain represents NoGo response options equivalent to Go response options—i.e., as a signal racing to a given response threshold in competition with alternative Go responses (Gomez et al. 2007)—or whether the brain only represents Go responses and performs a NoGo action in case no Go signal reaches the response threshold until a given deadline (Ratcliff 2006). Future research on the neural representations of NoGo responses is needed to adjudicate between these two possibilities, potentially using paradigms that feature Go vs. NoGo decisions in different effector modalities (e.g., eye, hand, and foot responses) (Aron 2011).

So far, I have suggested that action-related theta and response conflict-related theta might reflect two different sources and neural processes. However, results in Chapter 2 are also consistent with the possibility that both phenomena reflect one unified source and process. Theoretical accounts have previously assumed that increases in midfrontal theta power reflect a neural signal elevating the response thresholds in basal ganglia pathways. In Chapter 2, I put forward the alternative interpretation that phenomena of increased theta power under conflict reflect the extra units of accumulated evidence that follow from elevated thresholds—rather than a signal driving such elevated thresholds. Under this interpretation, action-related theta and conflict-related theta could in fact arise from the same underlying process. Previous research has shown that typically, midfrontal theta power increases strongly for any response on any trial, with conflict-related modulations constituting a rather minor increase top of a much larger "Go" signal (Cohen and Cavanagh 2011). It is thus **plausible that theta rises during preparation for a Go response and that this signal rises even further when the detection of a response conflict leads to elevated response thresholds**. Under this assumption, **midfrontal theta power reflects evidence accumulation for a Go response itself**, an explanation that can unify both action-induced and response conflict-induced increases in midfrontal theta power.

Other recent studies have attempted to identify EEG correlates of evidence accumulation in perceptual or value-based decision-making, reporting signatures akin to theta power increases described in Chapter 2. For example, one study using canonical correlation analysis on various frequency bands identified centroparietal theta power as the best candidate for reflecting evidence accumulation (van Vugt et al. 2012). Other research has identified signatures of perceptual and value-based evidence accumulation in the gamma range (Polanía et al. 2014) with centroparietal and frontopolar peaks in the topography, highly similar to the topographies reported in Chapter 2. The noise arising from the MRI environment precluded sufficient-quality gamma signal in our data in Chapter 2, but MEG data in Chapter 5 suggested a similar signal increases around responses in both the theta and the gamma band. Coupling of gamma power the phase of theta oscillations is well established phenomenon, bridging both frequency bands (Lisman and Jensen 2013). Lastly, topography and time course of the rising theta power signal in my data bore resemblance to the P300 potential, which is typically observed over centroparietal electrodes. An emerging line of research proposes the centro-parietal positivity (CPP) (O'Connell et al. 2012), a slow waveform similar to the P300 (Twomey et al. 2015), to reflect perceptual evidence accumulation. Future research will need to address to what extent centroparietal theta and gamma power as well as the CPP potential can give insights into ongoing value-based evidence accumulation in the brain.

Strikingly, trial-by-trial theta power over midfrontal electrodes was best predicted by trial-by-trial BOLD signal from the striatum. This finding contrasts with previous EEG-fMRI studies of related signals—such as the error-related negativity (Debener et al. 2005) and the feedback-related negativity (Hauser et al. 2014)—observing correlations of trial-by-trial EEG amplitude with BOLD signal in midfrontal cortex (i.e., anterior and mid-cingulate cortex). Similarly, source reconstruction studies have identified sources of theta power in midfrontal cortex (Hanslmayr et al. 2008; Cohen and Ridderinkhof 2013). Indeed, it is unlikely that oscillations arising from striatal activity are directly visible over the scalp. Instead, "antennae" in midfrontal, motor, or sensory cortices that receive inputs from the basal ganglia might give rise to theta power modulations over the scalp. Still, it is **worth considering the possibility that midfrontal theta power might give insights into subcortical sensorimotor integration and action selection and processes** (DeCoteau et al. 2007a, 2007b; Womelsdorf et al. 2010).

While I expected theta power to reflect conflict processing in the MGNG Task, I instead observed a short increase in midfrontal alpha power on incongruent trials. This alpha signal appears to be novel and not yet described by previous literature. Research on the role of frontal alpha oscillations is relatively scarce. Some literature points at a role of frontal alpha oscillations in controlling posterior alpha oscillations involved in the control of spatial attention (Bressler et al. 2008; Gregoriou et al. 2009; Wokke and Ro 2019). Potentially, frontal alpha could also be involved in the control of feature-based attention, weighting up certain stimulus features (i.e., the required action) over others (the cue valence). This hypothesis, although highly speculative, aligns with a previous study finding a transient increase in phase-locking between frontal and posterior alpha power—interpreted as reflecting attentional filtering—when participants exerted self-control in dietary choices between healthy and unhealthy food items (Harris et al. 2013). Future research is needed to understand the exact role of frontal alpha signals occurring under response conflict.

In sum, in chapter 2, I do not find support for the notion that midfrontal theta power indexes conflict between Pavlovian biases and instrumental task requirements. Instead, I interpret theta power as reflecting a latent evidence accumulation process for performing a Go action (i.e., recruit effort) or not. Curiously, I do find that striatal BOLD signal and midfrontal theta power are linked on a trial-by-trial basis, though in an unexpected manner: Both signals primarily reflect the eventually performed action. If this association is true, then midfrontal scalp theta power holds potential for insights into subcortical action selection processes. Indeed, the theta evidence accumulation process bore striking resemblance to striatal dopamine "ramps" observed in mice that recruiting effort to approach a rewarding goal. Most likely, this theta power signal I observed is distinct from other theta sources associated with errors, negative feedback, or surprise, adding to the growing literature of multiple independent theta sources. Future research using source reconstruction or intracranial recordings from the striatum might be able to further corroborate this midfrontal theta power-striatum link.

6.3.2 Pavlovian biases in reinforcement learning

Recent research has established that, beyond humans tending towards different actions when trying to win a reward vs. avoid a punishment, action values are also differentially updated after reward vs. punishment receipt (Swart et al. 2017, 2018): Rewards are preferentially attributed to one's own actions, while punishments are only reluctantly attributed to one's inaction. These learning biases are an alternative computational mechanism that can give rise to Pavlovian biases in behavior. In the animal realm, it might explain phenomena such as auto-shaping and negative maintenance in which animals consistently attribute rewards to their own actions without considering the possibility that not acting might in fact increase returned rewards. In Chapter 3, I combined EEG and fMRI recordings to investigate the neural processes underlying such biases in humans. Specifically, I investigated whether biased learning signals were first visible in the striatum—in line with the idea that the asymmetric nature of the direct/ indirect basal ganglia pathways might be sufficient to give rise to these biases—or whether, alternatively, biased learning was first visible in prefrontal circuits, suggestive of more sophisticated mechanisms driving these biases.

Biased learning signals (i.e., biased prediction errors) were present in several cortical and subcortical regions, including vmPFC/ pgACC, dACC, PCC, and striatum. Importantly, EEG correlates of dACC signal and PCC BOLD signal preceded correlates of striatal BOLD signal. These findings suggest that cortical rather than subcortical signals might start the neural cascade that eventually leads to biased updating of action values in the striatum. On the one hand, one

might find this observation quite noteworthy given that prefrontal regions have typically been interpreted as subserving flexible, "un-biased" computations associated with counterfactual reasoning (Boorman et al. 2009; Kolling et al. 2018; Fouragnan et al. 2019)—processes that should prevent phenomena like negative auto-maintenance in which agents get "stuck" in a certain behavioral pattern without considering alternatives. Hence, these results might appear quite peculiar in that they challenge the old notion of behavioral inflexibility arising from subcortical circuits. On the other hand, these results relate to the findings on the neural origin of response biases reported in Chapter 2, which show vmPFC BOLD signal to encode cue valence and exert markedly early EEG correlates—similar to vmPFC/ pgACC and dACC BOLD signals during biased learning.

The involvement of cortical regions in credit assignment might not be surprising given how difficult it can be to infer the latent causes of rewarding outcomes—processes that appear to involve prefrontal circuits (Jocham et al. 2016; Noonan et al. 2017; Monosov and Rushworth 2022). These circuits have to closely monitor the sequence of an agent's actions in relation to obtained outcomes. One mechanism that allows agents to keep track of their own recent actions might be traces of elevated dopamine that remain in the striatum after dopamine has ramped up to an action. These dopamine "residuals" will lead to elevated predictions errors for rewards that follow soon after an action (Cockburn et al. 2014), resulting in stronger credit assignment to the respective action. Notably, dopamine ramps appear to be sensitive to whether an agent contributed actively towards reaching the goal or whether external factors led to a lucky goal attainment (Hamid et al. 2021). To serve such a function, dopaminergic ramps would need to be informed by estimates of how more likely a goal will be reached when investing an extra unit of effort—again consistent with the "value of work" hypothesis of striatal dopamine during action selection. Inferences about the controllability of the task are needed.

Prefrontal circuits have been suggested as candidate regions for estimating task controllability —particular the ACC (Ligneul et al. 2022). This is also the region in which I observed the earliest biased learning signals in Chapter 3. Previous work reported that dACC BOLD signal encodes when and how fast an environment changes, flexibly up-regulating the learning rates in the case that old action value estimates need updating (Behrens et al. 2007). Furthermore, controllability over a stressor has been found to be encoded in the ACC (Amat et al. 2005), which shuts down serotonergic stress responses when stress is inescapable (i.e., uncontrollable). Finally, another line of research has traced projections of pgACC to the striosomes in the striatum which can mute reward prediction errors in the midbrain (Crittenden et al. 2016; Evans et al. 2020). Tonic signals from pgACC can shunt reward prediction errors and prevent credit assignment of outcomes to actions, leading to effort withdrawal and apathy (Amemori and Graybiel 2012). Notably, patients suffering from depression, which is often characterized by apathy, have been found the exhibit reduced BOLD signal in pgACC (Pizzagalli 2011; Ironside et al. 2020), potentially reflecting altered estimates of environmental controllability. Taken together, these results hint at ACC mechanisms inferring environmental controllability and controlling dopamine ramps in the striatum, which are necessary to preferentially attribute outcomes to recent Go actions. Although this mechanism is speculative and needs corroboration by further research, it could potentially explain via prefrontal circuits—notably in ACC—could lead to preferential attribution of rewards to self-initiated Go actions.

In contrast to the boosting of reward encoding by previous actions, it is less clear why participants have problems attributing punishments to refraining from action. A potential explanation might be the low baseline firing rates (i.e., 3–5 Hz) of striatal neurons, which imply

that negative prediction errors need to be encoded via pauses (rather than dips) in dopaminergic cell firing (Schultz et al. 1997). Some theories have questioned whether such pauses might be detectable by downstream regions, at all, and instead postulated an additional neuronal system encoding punishments that opposes the striatum (Daw et al. 2002). This concern might be less severe when actions trigger dopamine ramps against which pauses in firing—as induced by punishments—should be easily detectable. However, in absence of such ramps, baseline firing rates are almost at floor, and pauses induced by punishments hardly noticeable. Hence, controllability estimates from ACC could be crucial for keeping baseline firing in a range appropriate for both boosting reward prediction errors after actions as well as keeping punishment prediction errors noticeable against background activity. This explanation is highly speculative, though in line with findings that movement inhibition reduces striatal firing rates to almost zero (Coddington and Dudman 2018). In sum, **the observation that biased learning signals are visible in prefrontal before striatal circuits might be explained by a mechanism in which prefrontal circuits modulate dopamine baseline levels in a way that facilitates reward encoding after actions, but makes it hard to encode punishment signals after inactions**.

From a functional perspective, learning biases might constitute adaptive "priors" on which action-outcome relationships in the environment should be learned. Selective attribution of rewards to self-initiated actions might help prune the set of all possible actions towards a limited set of actions that are in fact conducive to reward attainment (Cazé and van der Meer 2013; Chambon et al. 2020; Lefebvre et al. 2022). Giving credit to an action that preceded a reward implies that it will be performed more often in the future, resulting in more instances in which it will be evaluated on its ability to increase the chance for rewards (Sepulveda et al. 2020). If an action does not yield rewards in future instances, it will be abandoned again. Sampling action-outcome relationships that appear promising—but might turn out to be spurious—is arguably a better strategy than sampling actions at random—or even remaining passive without taking active means to increase reward rates. Under this perspective, the transient adoption of spurious action-outcome relationships is unlikely to hamper reward pursuit on longer time scales.

In contrast, in the domain of punishments, such an exploration and maximization strategy might be ill-advised. Chances to try out different actions that could potentially avoid a threat are limited—once the agent is caught and eaten by a predator, exploration stops. Hence, in the domain of punishment avoidance, NoGo might be the preferred "sticky" prior given that exploration of different actions could have lethal consequences (Nesse 2001; Haselton and Nettle 2006). When most threats can be avoided by staying still, then a reduced ability to become active under punishment avoidance appears to be a rather minor cost. The net benefit is a powerful, but inflexible Pavlovian mechanism that saves the agent's life in a majority of dangerous situations.

In sum, in chapter 3, I report evidence for biased learning from rewards/ punishments after Go/ NoGo actions. I postulate that **this phenomenon arises through prefrontal circuits—most notably the ACC—that bias striatal dopamine levels and in this way boost reward prediction errors after self-initiated Go actions, while making it hard to detect punishment prediction errors after previous NoGo actions**. I propose that these biases are **likely adaptive by directing learning towards few good candidate actions that maximize rewards, while being cautious about trying out Go actions to avoid punishments**. Together with Pavlovian response biases, also learning biases might be conducive to promoting Go actions under opportunities to win rewards, while promoting NoGo actions under the risk of punishments.

6.3.3 Recruiting Pavlovian biases to support action implementation

Computational models of behavioral control have proposed that the mind features different behavioral control systems because different strategies are adaptive in different environments (Daw et al. 2005; Milli et al. 2021). However, the multiplicity of systems induces the meta-problem of deciding which system to rely on in a given situation (Boureau et al. 2015). The Pavlovian system has been interpreted as providing **"priors" on which actions to perform in novel, uncertain, or uncontrollable environments** (Dorfman and Gershman 2019). This idea implies that the impact of Pavlovian control should vanish once an agent gets familiar with a certain environment and acquires knowledge about more specific action-outcome contingencies. In reality, however, **Pavlovian biases appear to be pervasive even in well-known environments**, inducing maladaptive action slips (Cavanagh et al. 2013; Watson et al. 2014; van Steenbergen et al. 2017; Swart et al. 2018). Hence, I asked whether, beyond providing action priors, Pavlovian control might have additional purposes, rendering a strong Pavlovian system adaptive even in well-known environments.

Situations in which Pavlovian biases clash with instrumental action requirements can be solved in different ways. Recruiting top-down inhibition to suppress biases (Cavanagh et al. 2013) is only one possibility. As an alternative, early theories of Pavlovian control suggested that Go2Avoid and NoGo2Win conflicts could be resolved by mentally re-interpretating these situations. For example, in Go2Avoid situations, agents could imagine the moment they will have escaped a threat ("***safety signaling***")(Mowrer 1947) and in this way create an "imaginary" reward cue that helps them invigorate behavior and escape the situation (Boureau and Dayan 2011). Likewise, in NoGo2Win situations, an agent could imagine the situation of missing out on the available reward ("***frustration signaling***") and in this way create an "imaginary" punishment cue that helps them suppress impulsive actions. Under this perspective, no inhibition is needed, but biases can be altered by creating internal, imaginary reward or punishment cues.

The idea of resisting temptation by reinterpreting a situation—termed "***stimulus control***"—also resonates with the original intention of research on the Marshmallow Test. This research was initially not targeted at the role of individual differences in self-control, but instead had the interventional aim of improving children's patience. Children's ability to delay gratification was successfully boosted by teaching them cognitive strategies to re-interpret rewards in an abstract manner or to come up with distractions (Mischel and Moore 1973; Mischel and Baker 1975). In these trainings, avoiding direct attention the available rewards (e.g., covering one's eyes with one's hands) was crucial (Mischel and Ebbesen 1970). Beyond symbolic imagery, also regulating the direct physical contact to reward cues can be a successful way to reduce (or invigorate) their motivational power (Bushong et al. 2010).

Imagery of cues can be supported by eye movements in physical space. Humans do not only use their gaze to sample information in the external environment, but also to retrieve items from internal working memory. For example, when humans are shown a picture and subsequently asked to retrieve a certain detail about it, such as whether a car parked at the side of the road was red or blue, they will automatically move their gaze to the position on the screen where they have previously seen the car—even though the picture of the scene has already disappeared. This so-called "***looking at nothing***" phenomenon (Johansson and Johansson 2014; Laeng et al. 2014) demonstrates the power of eye movements in directing the internal focus of attention. Recent evidence suggests that even microsaccades—very subtle saccades observed when participants are instructed to keep central fixation—are biased towards a spatial location participants hold in

working memory (van Ede, Chekroud, and Nobre 2019; Van Ede et al. 2021). In sum, re-interpretating a given environment as a "Win" or "Avoid" one can be supplemented by eye movements that attend to (or retrieve memories of) reward and punishment cues. Such a strategy constitutes a powerful alternative to overcoming the influence of Pavlovian biases when these are at odds with action plans.

So far, I have described how imagery—supplemented by eye movements—could help overcome the maladaptive impact of Pavlovian biases. These considerations reveal another strategy of downregulating biases, but they do not explain yet why a strong Pavlovian system might be adaptive in the first place. Strong Pavlovian control might be warranted when attention is not only used to *downregulate* Pavlovian biases, but on the contrary to even *invigorate* them when adaptive. Potentially, humans could speed up their responses by focusing on rewards cues or vice versa suppress impulsive actions by focusing on punishments. In such cases, action execution is seemingly "outsourced" to cues in the environment that automatically trigger an intended action. Pavlovian biases can take the role of an "auto-pilot" or "training wheels" that do not determine the direction of movements but, once an action plan is formed, aid its smooth implementation.

Outsourcing control to the Pavlovian system could spare costs involved in implementing an action (e.g., the active maintenance of the plan in working memory or effortful movement invigoration) and effectively shield the action plan against interference. Such a strategy might be akin to the concept of "***mental offloading***", which describes the phenomenon that humans "*outsource*" memories and intentions to cues in the environment which, upon encountering them, serve as reminders for planned activities (Gilbert 2015; Risko and Gilbert 2016). For example, humans can strategically place an umbrella next to the front door so as to not forget it the next day, or set a timer for when to take the cake out of the oven. In a similar way, they could potentially surround themselves with positive cues (e.g., pictures of positive memories) to keep themselves engaged with the task at hand, or by negative cues that keep them vigilant and open to sudden changes in the environment. Crucially, both the environmental setup and the focus of attention are not fixed. Instead, humans can actively stock their environment with cues as well as decide which cues to attend to. In this thesis, I suggest that ideas about avoiding the maladaptive impact of biases—using imagery and attention—can also be used to invigorate these biases and use them as an auto-pilot for action execution.

Apart from saving potential control costs, one might ask whether "transferring" action control from the instrumental to the Pavlovian control system confers any specific benefits. There are a few possible advantages of the Pavlovian system. First, the inflexibility of Pavlovian biases could be an advantage: Once fixated, reward and punishment cues set agents on a "ballistic" track towards action invigoration/ inhibition in a way that might be **robust to interference by other cues**. Second, the presence of Pavlovian cues has been found to **speed up learning**. The presence of a Pavlovian cue at which behavior can be directed ("sign tracking") can lead to more reliable acquisition of behavior. Cues that, in some circumstances, are the target of sign-tracking behavior can in other circumstances lead to higher performance levels than a mere 100% reinforcement schedule without such a cue (Brown and Jenkins 1968; Dayan et al. 2006). Third, Pavlovian cues appear to be able to induce instant outcome re-evaluation. Unlike instrumental conditioning that results in learning of abstract reward values, Pavlovian cues seem to also confer **information about the reward identity** (Robinson and Berridge 2013; Dayan and Berridge 2014). This knowledge is relevant if a stimulus is aversive in one motivational state, but becomes appetitive in another. For example, if animals have learned that a certain response yields an unpleasant outcome (e.g., a salty liquid), they will typically fail to re-evaluate this response in a different motivational

state (e.g., under salt deprivation) in which the outcome has become pleasant and would satisfy a newly arisen need (Dickinson 1986). Instead, animals typically have to first re-sample the outcome to recognize its changed value. In contrast, when an outcome has been conditioned to a Pavlovian cue, resampling is not needed: Animals will instantly approach the cue and start showing the response that leads to the re-valued outcome (Robinson and Berridge 2013). In sum, Pavlovian control can even account for changes in motivational states in a way that is not available to instrumental learning systems.

In sum, humans may not always need to recruit top-down inhibition systems to downregulate biases. Instead, they can symbolically reinterpret a given situation as a "Win" or "Avoid" situation such that it is in line with ongoing Go/ NoGo action plans. Selective attention to certain cues in the environment might be an important contributor to such a reinterpretation. By outsourcing control to Pavlovian biases, action plans might be more smoothly and robustly implemented. Notably, participants could even actively shape their environment to contain cues that help them reinterpret situations in line with action plans. Taken together, **under the perspective that Pavlovian control can be used as a "training wheel" that support action implementation, a strong Pavlovian control might in fact be conducive to goal-directed action**. Pavlovian and instrumental control should not be seen as competitors, but as systems that interact and can jointly achieve goals in a more reliable way (Gershman et al. 2014; Moran et al. 2019).

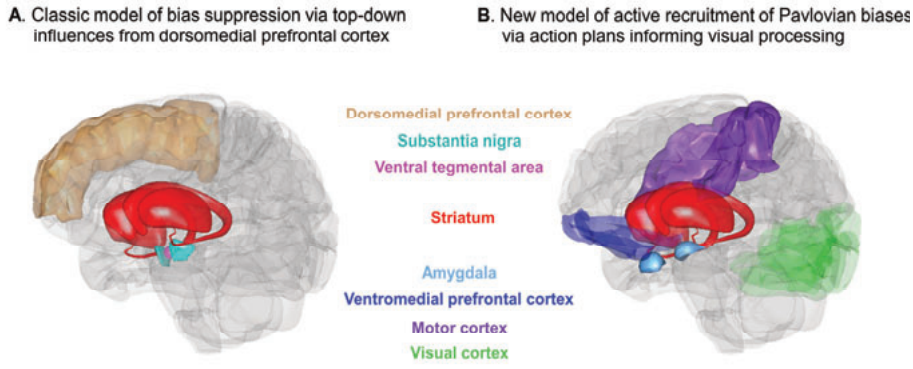


Figure 6.2. Core regions involved in the classic model of top-down suppression Pavlovian biases as well as the new model of active recruitment of biases via selective visual attention informed by action plans.
A. Model used when planning the research presented in Chapter 2. The dorsomedial prefrontal cortex (dmPFC) detects conflict between response tendencies triggered by Pavlovian biases and more contextually appropriate response tendencies that lead to desired outcomes. A top-down signal from dmPFC is sent to the striatum to downregulate cue valence signals that arise from the midbrain (ventral tegmental area, VTA, and substantia nigra pars compacta, SNc). **B.** Revised model in light of the findings presented in Chapter 4 and 5. Go/ NoGo action plans are formed in motor cortex and then inform visual regions, which then selectively (attend to and) process reward/ punishment cues that are in line with the action plan. Reward/ punishment information reaches ventromedial prefrontal cortex (vmPFC) and amygdala, which transmit it to the striatum and in this way give rise to Pavlovian response biases. Images created using Anatomography, a website maintained by the Life Science Databases (LSDB), under CC-BY-SA-2.1-jp.

6.3.4 Using strategic attention to invigorate Pavlovian biases

In Chapters 4 and 5, I tested the hypothesis that participants seek out cues that triggers Pavlovian biases in a way that is in line with their goal-directed action plans. Many action plans are

extended in time and prone to interference by suddenly occurring stimuli in the environment (Wolfe 2021). Hence, participants might have to actively direct their attention to seek out cues that invigorate their action plans as well as avoid cues that could disrupt action plans. In Chapter 4, I present evidence that eye-movements to reward and punishment information are indeed informed by agents' action plans, with higher attention to reward (compared to punishment) cues when agents plan to perform a Go (compared to a NoGo) action. Vice versa, more attention to rewards (compared to punishments) strongly drove participants to perform a Go (compared to a NoGo) response, corroborating the notion that attention can be used to invigorate Pavlovian response biases. Crucially, participants with a stronger "outsourcing" of actions to their fixation pattern tended to show higher performance, suggesting that aligning attention to reward and punishment cues to one's action plans can be an adaptive strategy. In Chapter 5, I replicated these findings in eye-movements, but did not find evidence for a similar modulation of covert attention as indexed by posterior alpha power lateralization in the MEG signal. Similarly, I did not find evidence for beta power desynchronization as an index of ongoing latent action preparation to affect alpha power lateralization. Taken together, I found strong evidence that eye-movements are used to invigorate biases in line with action plans, but no evidence for covert attention playing a similar role. If anything, modulations of covert attention by task factors occurred rather late and in a merely reactive fashion.

In real life, action plans and memory representation are prone to be interrupted by sudden events (Wessel, Jenkinson, et al. 2016; Wessel 2018b). If agents do not steer their attention actively, it will likely be caught by irrelevant stimuli that potentially distract from their goals. Hence, attention in real life is highly structured, incorporating prior knowledge on where relevant information will appear (Peelen and Kastner 2014; Wolfe 2021). Such prior knowledge might also comprise implicit knowledge about the fact that action invigoration/ inhibition can profit from attention to reward/ punishment information. Even if benefits of attending to action-matched information might be negligible in some situations, in any case, it will be important to not attend to action-mismatching information that has the potential to disrupt action plans (Verbruggen and De Houwer 2007; Pessoa et al. 2012). Hence, attention should always take action plans into account.

In both Chapters 4 and 5, I observed consistent effects of action plans on eye-movements as well as, vice versa, effects of eye movements on eventual responses. However, these effects in overt attention were not paralleled by covert attention as indexed by posterior alpha power lateralization. Possibly, action plans play a (stronger) role in situations in which cues are distributed across the environment and have to be actively sampled via eye movements. Such constraints introduce a bottleneck on the order in which cues can be processed, incentivizing agents to actively plan the trajectory of their gaze pattern. In contrast, when cues are in close proximity and within the range of covert attention, participants might not engage a proactive attention strategy, but instead explore space in a merely reactive, "anarchic" manner (Wolfe et al. 2000). Such a reactive strategy might be warranted by the fact that proactive, voluntarily steered attention is slower and less efficient than reactive attention. The fact that proactive response strategies are effortful and computationally costly implies that effort might be wasted if it turns out that a task can be successfully mastered without an explicit attentional strategy (Braver 2012). In the adapted MGNG Task used in the Chapters 4 and 5, I included catch trials, which required participants to always attend to both reward and punishment cues and compare them. In Chapter 4, I even employed a gaze-contingent design requiring participants to saccade to stakes in order to render them visible. Though rather artificial for a lab experiment, this setup might be closer to real life situations in

which cues are scattered across the environment and have to be actively sampled. In conclusion, the gaze contingent paradigm used in Chapter 4 might have been more suitable to induce proactive attentional plans in participants compared to the paradigm in Chapter 5 that instead relied on covert attention.

Nonetheless, the absence of any modulations of posterior alpha power lateralization in Chapter 5—either by action plan or other factors manipulated within the task—might appear surprising given that previous studies have consistently found alpha power to change with lateralized attention (Thut et al. 2006; Rihs et al. 2007; Bonnefond and Jensen 2012). Only recently, studies have questioned whether alpha power lateralization is in fact necessary and/ or sufficient for lateralized attention (Antonov et al. 2020; Gundlach et al. 2020). In the data presented in this thesis, there was no evidence for alpha power lateralization being modulated by participants' action plans. The only modulation of alpha power lateralization in these data was a relative decrease contralateral to the side of the screen at which the respectively higher stake appeared, suggesting a higher focus on higher magnitude stakes. This modulation was only visible towards the end of the stakes presentation phase. Notably, in this time window, signals reflecting stakes valence and magnitude occurred also in other frequency bands, i.e., in occipital gamma power and frontotemporal delta power. It is possible that any reactive attentional orientation, whether to action-matching stakes or simply to high-magnitude stakes, was slow and occurred rather late. Notably, modulation of central beta power by the net stakes difference occurred *before* these delta/ alpha/ gamma power modulations, questioning the causal role of these late signals in biasing behavior. The modulation of central beta power implies that stakes must have been processed already at earlier time points—potentially visible in occipital gamma power. In sum, in the data presented in Chapter 5, alpha power lateralization occurred late and only after the influence of stakes had become visible in action preparation signals, questioning its causal role in attending to cues that would trigger Pavlovian biases.

The fact that saccades tended to be targeted at stakes that matched participants' action plans implies that, instead of posterior alpha power, frontal signals reflecting oculomotor control might show the expected lateralization. Hence, future analyses of these data could focus on frontal signals that mediate the effect of action plans on attention allocation. Given that fronto-temporal signals encoded the overall valence and magnitude of stakes—at least towards the end of the stakes presentation phase—these signals are also likely candidates for triggering eye-movements towards rewards and punishments in a reactive manner. Future analyses of this data could explore whether these signals might arise from (left) amygdala and hippocampus, regions in which I also observed effects of cue valence on BOLD signal in the data reported in Chapter 2. Taken together, apart from alpha power modulations, it is possible that other, frontal signals that are informed by action plans contribute to lateralized attention to rewards and punishments, instead.

In sum, in Chapters 4 and 5, I present evidence that **participants indeed used overt attention (i.e., eye movements) to reward/ punishment information to trigger Pavlovian biases in a way that was in line with their action plans.** These results support the notion that humans can **actively use attention to rewards or punishment to turn a situation into a "Win" or "Avoid" one, facilitating invigoration or inhibition depending on their action plans**. There was no complementary signal in posterior alpha power lateralization, questioning its involvement in attention allocation in the context of the particular task used in these chapters. Future research might thus explore whether frontal signals directly involved in oculomotor control might be informed by Go/ NoGo action plans.

6.4 CAVEATS AND LIMITATIONS

6.4.1 The causal roles of involved brain regions

The research presented in this thesis relies heavily on observing rather than manipulating behavior (e.g., button presses, reaction times, eye-tracking) and neural responses (EEG, MEG, fMRI). This correlational evidence does not speak to the causal roles of the striatum or PFC subregions in the emergence of Pavlovian biases. Although the role of the direct/ indirect pathways in action invigoration/ suppression appears well-established by animal studies employing optogenetics (Kravitz et al. 2012; Lammel et al. 2012), it is unclear whether the same principles apply to humans. There are important differences between the human, macaque, and mouse striatum (Balsters et al. 2020), with motor circuits being relatively conserved, but other striatal circuits substantially expanded in humans and primates. Similarly, depending on its definition, (certain regions of) prefrontal cortex might be unique to primates (Carlén 2017; Cisek 2020; Mars et al. 2021). The research presented in Chapters 2 and 3 is in line with an important role of dACC and pgACC in inducing both Pavlovian response and learning biases, which appear to exhibit signatures of biased learning prior to the striatum. However, whether these regions are indeed integral to such biases needs to be studied via causal interventions. For deep neural sources such as ACC and striatum, transcranial ultrasonic stimulation (TUS) constitutes an exciting new tool (Fomenko et al. 2018; Darmani et al. 2022). TUS has been successfully used to stimulate deep neural regions and, in this way, alter choice behavior and learning in macaques (Fouragnan et al. 2019; Bongioanni et al. 2021; Folloni et al. 2021). Using TUS could help to disentangle the different roles of vmPFC, ACC, amygdala, and striatum in Pavlovian response and learning biases.

Furthermore, in Chapters 4 and 5, I advance the notion that humans can use selective visual attention to seek out reward and punishment information and that attention to such cues causally triggers Pavlovian biases in their behavior. In Chapter 4, I present results from a small online study manipulating attention by placing Go/ NoGo stimuli closer to either the reward or the punishment stake. Results tentatively suggest a causal effect, as reward proximity boosted Go responding while punishment proximity suppressed Go responding. Further studies have to corroborate this causal effect of attention on choices. Adapted designs might present reward and punishment cues not simultaneously, but sequentially, potentially manipulating the duration of how long the cues are visible (Pärnamets et al. 2015; Reeck et al. 2017). Such designs might be particularly interesting from an interventional perspective of training people to use their biases in an adaptive way.

6.4.2 Individual differences in biases

Though Pavlovian biases appear to be omnipresent in all species studied, their magnitude varies across individuals and some humans even appear unaffected by Win vs. Avoid cues in the MGNG Task. Animal research has suggested that sign-tracking vs. goal-tracking, i.e., the extent to which behavior is oriented towards a Pavlovian cue predicting reward availability rather than towards the instrumental goal, might be a stable individual trait (Flagel et al. 2010). Ongoing research aims to extend this idea to humans (Garofalo and di Pellegrino 2015; Colaizzi et al. 2020; Schad et al. 2020). The correlational results that I present in Chapter 4 tentatively suggest that incorporating stake magnitudes into behavior hurts performance in the context of the adapted MGNG Task, while relying on attention to the stakes helps performance. However, till today, there are no validated computational measurement tools to reliably quantify individual differences in Pavlovian biases.

In Chapters 2 and 3, I investigated the neural mechanisms underlying both Pavlovian response and learning biases. A computational model that incorporated a combination of both biases explained behavioral patterns better than each bias in isolation. Both phenomena appeared to involve similar regions in PFC and striatum. Still, on an individual person or even trial-level, it might be difficult to tell apart whether biased behavior arose from an impulse-like response bias or from biased inferences on past outcomes as embodied in learning biases. Particularly the size of the latter bias might be substantial in some individuals. Some of the participants featured in Chapters 2 and 3 (almost) never performed a NoGo action to a Win cue, suggesting that they did not even consider the possibility that NoGo actions could ever increase the rate of obtained rewards. Such phenomena of negative auto-maintenance might be pervasive in everyday life and constitute a different behavioral phenotype than impulsive action slips attributed to response biases. Although both biases can potentially be disentangled by the computational model that I used in Chapter 3, test-retest data is needed to assess the reliability of parameter estimates. An initial study using a simplified MGNG Task reported rather low reliability for the response bias parameter (Moutoussis, Bullmore, et al. 2018). Alternatively, model families such as drift-diffusion models, which also incorporate reaction times on top of responses, have been found to yield more reliable estimates for other choice tasks (Shahar et al. 2019; Brown et al. 2020) and might constitute an interesting possibility for future research.

6.4.3 Expanding the range of tools for measuring Pavlovian biases in behavior

Finally, I used two rather simplistic and abstract tasks to measure Pavlovian biases. These tasks might have limited ability to discriminate response and learning biases and did not induce biases in every participant. New tools will be needed to test for the presence of such biases in a broader range of behaviors and situations, including non-laboratory contexts.

Effects of rewards and punishments on motor behavior might not just affect Go/ NoGo decisions or their response vigor/ speed, but more subtle aspects of movement kinematics. For example, measuring subthreshold muscle twitches with electromyography could reveal when participants almost performed an impulsive Go action, but inhibited it at the last moment (Cohen and van Gaal 2014). Similarly, tracking mouse movements (Dignath et al. 2014) or touch screen use (Meule et al. 2019) could reveal the influence of Pavlovian biases. Finally, beyond finger movements, also visceral responses such as body sway (Ly et al. 2014), the post-auricular reflex (Johnson et al. 2012; Stussi et al. 2018), or the startle reflex (Kuhn et al. 2020) could reveal implicit "liking" or "freezing" reactions.

Another approach could be to quantify spontaneous, ongoing behaviors, e.g. whisking and facial movements in rodent research, via the use of video monitoring (Musall, Urai, et al. 2019; Roy et al. 2021), which has the potential to yield new, unprecedented insights into how behavior is affected by task manipulations. Automatic quantification of affective facial responses from video recordings has just begun in human research (Chang et al. 2021) and might yield novel insights into humans' affective reactions during goal pursuit. Taken together, Pavlovian biases likely influence actions beyond simple Go/ NoGo decisions and reaction times, warranting the use of novel tools to further scrutinize them.

In sum, future research would need to use causal interventions to corroborate the findings presented in this thesis; it should develop new tools to quantify individual differences in both response and learning biases; and it could use new measures to further explore the reach of Pavlovian biases into everyday life.

6.5 FUTURE DIRECTIONS AND APPLICATIONS

6.5.1 Pavlovian biases may explain framing effects in decision making

The focus of this thesis has been on the intrinsic link between Go/ NoGo actions and reward/ punishment availability, termed Pavlovian biases. When this link is reframed and broadened, it might have the potential to explain various other phenomena in the decision-making literature. The categories "Go" and "NoGo"—often attached to the direct and indirect basal ganglia pathways—hardly reflect decision dilemmas that agents face in everyday life. These labels derive from early basal ganglia models that focused on its role in motor control. More recent models have reframed the role of the basal ganglia not just in motor control, but decision making and (cognitive) action selection, more generally (Frank et al. 2001; Dayan 2012). Instead of "Go" and "NoGo", the implications of direct/ indirect pathway activation might be more appropriately characterized as *acceptance* and *rejection* of a certain choice option that is currently considered (Amita and Hikosaka 2019; Hikosaka et al. 2019). Recent evolutionarily and ecologically inspired decision-making models have highlighted that multi-alternative choices might often be performed in a sequential, *foraging*-like manner (Kacelnik et al. 2011; Hayden and Moreno-Bote 2018; Hunt et al. 2018; Cisek 2020). In such a strategy, one option is considered at a time, accepted if it surpasses a certain aspiration level, and rejected otherwise, in which case the next option is considered. Basal ganglia pathways might play opponent roles in such sequential decision processes by accepting vs. rejecting the option that is currently considered (Hikosaka et al. 2019).

Furthermore, a focus on "rewards" and "punishments" might be too limited—instead, the pathways might be more appropriately characterized as encoding the *benefits* and *costs* of a particular action under consideration (Collins and Frank 2014; Westbrook et al. 2020, 2021). Specifically, the indirect pathway should plausibly not only consider explicitly negative outcomes such as punishments, but also the (physical or cognitive) effort that comes with completing a certain action. Benefits lead to increases in striatal dopamine levels, promoting acceptance of the considered action, while costs decrease dopamine levels and provide evidence for its rejection. ***Acceptance* and *rejection* promoted by *benefits* and *costs* of a choice option** might be more meaningful, ecologically valid categories. In fact, research using accept/ reject framing manipulations has observed phenomena akin to Pavlovian biases, with higher sensitivity to positive information in accept frames and higher sensitivity to negative information in reject frames (Meloy and Russo 2004; Glickman et al. 2018; Frömer et al. 2019; Sepulveda et al. 2020). In such a new framework, **Pavlovian biases can explain how contextual factors modulate reward and punishment sensitivity, which then gives rise to framing effects** (Glickman et al. 2018). Future research should more systematically explore the explanatory power of Pavlovian biases in explaining various contextual decision anomalies.

6.5.2 The automatic effects of rewards on cognitive control recruitment

Suboptimal performance in certain cognitive tasks (e.g., on incongruent trials in Stroop or Flanker tasks) has long been attributed to fixed limits in cognitive capacity that cannot be exceeded (Cohen 2017). In contrast, in recent years, studies have demonstrated that humans can seemingly transcend those limits when they are offered sufficient incentives. For example offering higher rewards can reduce congruency effects in the Stroop task (Krebs et al. 2010; Dixon and Christoff 2012). Such findings have been interpreted in the sense that humans may not lack the ability, but rather the motivation to recruit effortful control to boost behavior (Cools 2016; Westbrook and Braver 2016). Recruiting cognitive control has thus been cast as a decision process itself which

trades off changes in rewards deriving from higher control investment against the costs of investing control (Shenhav et al. 2013; Westbrook et al. 2021).

The exact conditions under which rewards motivate additional cognitive control recruitment remain unclear. In the classic decision-making literature, cost-benefit tradeoffs are often conceptualized as slow, deliberate choice processes involving propositional reasoning, e.g., in buying a car or picking a holiday destination. In contrast, cognitive tasks typically work with time pressure and trial-by-trial rewards continuously changing. Decisions about whether and how much control to invest must be much faster than any conscious deliberation process. Hence, it is dubious whether the process that mediates the effects of rewards on cognitive control recruitment is anything like a propositional cost-benefit tradeoff. Instead, I suggest that such effects might be Pavlovian in nature, with reward cues automatically triggering control irrespective of whether the cued reward is contingent on task performance or not. Research has systematically compared the effects of performance-contingent and performance-non-contingent rewards on behavioral invigoration. One prominent set of findings has shown increased saccade vigor even for non-performance-contingent rewards (Manohar et al. 2017; Grogan et al. 2020) or rewards whose attainment is contingent on accuracy, but not on speed (Kawagoe et al. 1998; Milstein and Dorris 2007; Reppert et al. 2015). Recently, it has been suggested that saccade vigor might even allow for indirect inferences about the subjective value that humans assign to cues (Shadmehr et al. 2019). Similar modulations of movement vigor by rewards have been observed for other body movements, e.g. reaching movements targeted at abstract locations (Summerside et al. 2018) or at candy bars (Sackaloo et al. 2015). A modulation of cognitive control investment by performance-non-contingent rewards is still under debate (Chiew 2021), with some studies finding non-contingent rewards to boost control (van Steenbergen et al. 2009, 2012; Chiew and Braver 2014), while other studies have found such rewards to undermine (Dreisbach and Goschke 2004; Fröber and Dreisbach 2014) or to not affect control (Fröber and Dreisbach 2016). Taken together, **these findings suggest very fast and "involuntary" effects of rewards on movement vigor and speed, which might reflect automatic Pavlovian processes rather than deliberate cost-benefit tradeoffs.**

In sum, in this thesis, I advance the position that **many "motivational" effects of reward cues on cognitive control recruitment might not arise from a strategic, deliberate choice processes, but instead reflect "Pavlovian biases" of such cues automatically increasing response speed and vigor**. Future research is needed to disentangle to what extent the contingency between good task performance and reward delivery is relevant for motivational effects to arise.

6.5.3 Reward and punishment processing via unified vs. segregated systems

Much literature has assumed that distinct brain regions are involved in processing positive and negative information: regions such as vmPFC, PCC, or striatum have been deemed "reward regions" as they show higher BOLD signal for rewards, while other regions such as ACC, insula, and amygdala have been deemed "punishment regions" with higher BOLD signal for punishments (Pessiglione and Delgado 2015). In contrast, the direct/ indirect pathway model of the basal ganglia proposes a single system that is sensitive to both reward and punishment cues, with opposite effects on behavior. Understanding the neural basis of Pavlovian biases could thus shed further light on the nature of rewards and punishments as polar opposites on a unified dimension or as qualitatively distinct sets of stimuli recruiting different brain circuits.

The strongest evidence for a unified system representing rewards and punishments as poles of a single continuum comes from experiments on conditioned inhibition (Rescorla 1969; Tobler et al. 2003). In such experiments, a Cue A+ predicts the delivery of a reward; however, occurrence of A+ together with another cue X- means that no reward will be delivered. Presenting X- alone will subsequently reduce behavior or even induce fear responses, suggesting that X- has obtained a negative value by predicting the absence of an otherwise expected reward. Presenting both A+ and X- will induce a net expectation of zero reward as indexed by the absence of dopaminergic prediction errors (Tobler et al. 2003). These experiments suggest that both rewards and punishments might be presented on a unified scale by potentially one underlying system, the striatum.

Recent animal studies provided further evidence for the opposing nature of direct and indirect pathways in increasing and decreasing value. These studies have found optogenetic stimulation of D1 receptor-expressing direct pathway neurons to induce conditioned place preference, while equivalent stimulation of D2 receptor-expressing indirect pathway neurons induced conditioned place aversion (Kravitz et al. 2012; Danjo et al. 2014), with similar effects observed when stimulating (Lammel et al. 2012) vs. suppressing (Danjo et al. 2014) upstream projections from the VTA. Optogenetic stimulation vs. suppression of basal ganglia pathways can also modulate movement velocity (Yttri and Dudman 2016) and licking behavior (Bakhurin et al. 2020). In sum, these studies provide strong causal evidence for basal ganglia pathways giving rise to bidirectional, segregated value learning with consequences for activating vs. suppressing behavior.

Compared to ample research on appetitive motivation by reward cues, there is considerably less research (in humans) on how aversive cues can lead to behavioral inhibition. Some studies have employed PIT paradigms with both positive and negative cues (Huys et al. 2011; Geurts et al. 2013a, 2013b), finding that aversive cues tend to suppress choice behavior. However, other studies have found that, when an aversive cue such as a loud, unpleasant noise is already present, it actually invigorates behavior that serves the "escape" from this cue (Millner et al. 2017). Also in the animal realm, there are observations that aversive cues can invigorate hiding or escape behavior (Pinel and Treit 1979; Dayan et al. 2006). The exact conditions under which aversive cues suppress vs. invigorate behavior need to be explored by future research.

Although the direct/ indirect pathway model of a basal ganglia gives an elegant explanation for the opponent control of behavior by rewards and punishments, the idea that the whole striatum (and the upstream VTA) uniformly code for rewards (minus punishments) is likely not correct. Instead, dopamine bursts (interpreted as positive prediction errors) have been observed to positive cues, but also negative ones (Roitman et al. 2005; Matsumoto and Hikosaka 2009; Menegas et al. 2015, 2017). Potentially, separate subclasses of dopamine cells code for aversive events or, more generally, any form of unexpected or surprising event that deserves attention (Horvitz et al. 1997; Schultz 2016). The processing of fearful stimuli has been traditionally ascribed to nearby circuits in the amygdala (Herry et al. 2008). However, to complicate things further, the amygdala has also been found to be crucial for appetitive PIT effects (Hall et al. 2001; Corbit and Balleine 2005; Lichtenberg and Wassum 2017; Lichtenberg et al. 2021), likely because the amygdala can modulate dopamine release in the ventral striatum (Floresco et al. 1998; Ambroggi et al. 2008; Jones et al. 2010; Stuber et al. 2011). Hence, the idea of a uniform reward system in the striatum or a uniform punishment system in the amygdala is unlikely to hold, as both neural structures encode both forms of cues and likely interact in modulating behavior.

Taken together, **the direct/ indirect pathway model of the basal ganglia gives a unifying account of reward and punishment processing by the same structures.** While this model might be useful to explain many experimental findings, it does not account for the fact that reward/ punishment coding is heterogeneous and influenced by inputs from the amygdala and prefrontal cortex. Future models need to incorporate these additional findings for a more complete picture (Mollick et al. 2020).

6.5.4 Using Pavlovian biases to regulate other behaviors

What makes Pavlovian cues particularly interesting for practical applications is their ability to invigorate goal-directed actions targeted at other objects, which has been found both in animals (Estes 1943, 1948; Rescorla and Soloman 1967; LoLordo et al. 1974; Schwartz 1976; Lovibond 1983) and humans (Knutson et al. 2001; Beierholm et al. 2013). Even reward cues that are currently irrelevant or unattainable have the power to motivate agents in their ongoing behavior, including saccades (Manohar et al. 2017) and cognitive control exertion (Chiew and Braver 2014). This property makes Pavlovian cues a promising interventional tool—in clinical settings as well as in everyday life. Humans often have deliberate control about how they arrange their environment (e.g., their home office), and strategically decorating it with certain cues (e.g., holiday pictures) could boost their productivity in motivational dips (Risko and Gilbert 2016).

However, animal studies have shown that Pavlovian cues do not always modify the kinematics of ongoing goal-directed actions, but instead can induce alternative, competing actions (Breland and Breland 1961) such as pecking behavior towards the cues, termed sign-tracking (Flagel et al. 2007). The ability of cues to boost a goal-directed action or to interfere with it likely depends on physical properties of the cue (e.g., visual vs. auditory), a phenomenon that has not yet been investigated systematically (Rescorla 1988). The capture of attention by rewarding cues is a commonly observed phenomenon that disrupts ongoing task performance (Failing et al. 2015; Le Pelley et al. 2015; Watson et al. 2020) and might constitute a human version of sign-tracking behavior (Garofalo and di Pellegrino 2015; Schad et al. 2020). More research is needed to clarify when it is safe to use Pavlovian cues to invigorate goal-directed behavior.

While past research has mostly focused on the ability of rewards to invigorate behavior, it could be interesting to further consider the ability of punishment or threat cues to slow down or suppress behavior (Huys et al. 2011; Geurts et al. 2013a, 2013b). The presentation of incidental negative cues has been found to slow down and even disrupt performance on a focal task (Padmala et al. 2011; Pessoa et al. 2012). Another well-known phenomenon is the slowing-down of behavior after previous errors, called post-error slowing (PES), conflict adaptation, or "Gratton effect" (Rabbitt and Rodgers 1977; Gratton et al. 1992). While previously interpreted as a strategic response adjustment, post-error slowing has been revealed to occur with little volitional control (Notebaert et al. 2009; Wessel 2018c). Errors might thus constitute yet another class of aversive events that induce slowing-down of behavior, a phenomenon that, in many situations, is likely conducive to regaining control over an interrupted action sequence. More generally, unexpected surprising events, e.g., loud tones, can induce involuntary motor stopping (Wessel 2017; Dutra et al. 2018). Going beyond simple Go/ NoGo actions, emphasizing the downsides of a course of action—e.g., its effort costs—by presenting them first increases the chances that the action will be abandoned (Vassena et al. 2019; Westbrook et al. 2020; Müller et al. 2022). Taken together, the impact of aversive cues and events on motor slowing is well established and could potentially be used strategically. Slowing down behavior could be particular adaptive in high-stakes situations in which slightly mis-calibrated actions lead to the miss of a large reward. Indeed, slowing-down of

behavior has been observed when humans are aware of a large reward opportunity coming up (den Ouden et al. 2015; Shevlin et al. 2022). Especially such high-stakes situations that require response caution, careful deliberation, and/ or fine motor precision could benefit from the presence of aversive cues.

Finally, while Pavlovian biases describe how the valence of cues affects Go/ NoGo actions, the inverse link also holds up: Repeated Go actions towards a cue (e.g., healthy food) have been found to increase its liking (and consumption; so-called "cued-approach training") (Schonberg et al. 2014; Veling et al. 2017), while repeated NoGo actions towards a cue have been found to decrease its liking (and consumption) (Chen et al. 2016). Hence, links between valence and action are bidirectional, implying that that the performance of (in)actions towards a cue affords a tool to alter its affective value. This notion is in line with older ideas that the appetitive/ aversive valence of a cue depends on its ability to facilitate/ impair certain pre-programmed, species-specific consummatory behaviors (Glickman and Schiff 1967). Studies on fish and roosters have observed that cues commonly classified as aversive—e.g., the sight of a rival that induces anger—do not function as penalties, but as reinforcers, putatively because they induce an active fighting response (Thompson 1963, 1964). Although this theory has not been systematically investigated, it opens the interesting perspective that the valence of a cue is in fact rooted in its association with motor invigoration versus suppression.

Taken together, **Pavlovian biases might be at stake in many real-life situations in which irrelevant or unattainable positive/ negative cues invigorate/ slow down behavior**. Vice versa, **inducing action invigoration vs. inhibition could potentially alter the valence of cues**. As previously mentioned, new tasks and tracking methods, such as computer mouse data, touch screen, or automatic video capture could gain new insights into the variety of behavior that is affected by incidental positive and negative cues.

6.5.5 Balancing different behavioral control systems

Pavlovian biases constitute one of several behavioral control systems, inducing the meta-decision problem of which system to rely on (Daw et al. 2005; Boureau et al. 2015). Normative accounts have suggested that Pavlovian biases might constitute computationally cheap action "defaults" in environments in which control over the attainment of outcomes is low (Dorfman and Gershman 2019). Hence, a low sense of agency and environmental controllability might be crucial for participants' decision to rely on Pavlovian defaults instead of recruiting more effortful strategies. Inferences about task controllability might also be at the root of Pavlovian learning biases, but with *opposite* implications, i.e., agents giving *higher* credit to their own actions (compared to inactions) for rewards, which assumes *high* task controllability. Such beliefs about the impact of one's own actions on the environmental reward return rate might be highly relevant for arbitrating the impact of Pavlovian biases against other decision-making strategies and deserve further scrutiny.

While incremental action invigoration through reward cues might reliably increase the total number of accrued rewards only when considered on global, temporally extended time scales, the impact of threat cues on motor inhibition can be life-saving in every single moment in which it prevents the detection by a predator. Hence, even in well-known and seemingly controllable environments, Pavlovian biases need to have the power to disrupt ongoing instrumental behavior and prioritize survival via freezing (Nesse 2001; Haselton and Nettle 2006; Roelofs 2017). Under this perspective, Pavlovian biases might be recruited not just in uncontrollable environments, but also in "extreme modes" of large threats and/or large rewards. Stress manipulations have been

found to increase aversive Pavlovian biases (Mkrtchian, Roiser, et al. 2017), particularly in individuals with anxiety disorders (Mkrtchian, Aylward, et al. 2017). Vice versa, increased reward availability in a (task) environment has been found to decouple behavior from previously learned action values (Beeler 2012; Wittmann et al. 2020), with agents exerting higher physical vigor at the cost of accuracy (Niv et al. 2007; Guitart-Masip, Beierholm, et al. 2011; Beierholm et al. 2013) and similar effects on response speed in the domain of cognitive effort recruitment (Otto and Daw 2019). Taken together, Pavlovian cues seem to strongly affect behavior both in uncontrollable environments, but also in situations with high reward opportunity or high risk of punishments.

One crucial determinant of how much reward and punishment cues affect behavior might be their physical proximity. One study showed that participants' willingness to pay for a food item was 40–61 percent higher when the item was physically present (compared to a picture of it or a mere textual description) (Bushong et al. 2010). In contrast, placing a fully transparent plexiglass wall between a participant and a motivating piece of food undermined this effect. Beyond physical obstacles, also temporal delay can attenuate the motivational effects of rewards. Participants respond faster to cues signaling immediate rewards than to cues signaling subjective value-matched rewards available in the future (Luo et al. 2009). Research on temporal discounting has suggested that any temporal delay (compared to immediacy) in reward availability adds an intercept-like malus to the subjective value that an agent assigns to it, a phenomenon called "present bias" (McClure et al. 2004; Benhabib et al. 2010; Figner et al. 2010). Individual differences in this bias have been linked to credit card debts (Meier and Sprenger 2010) and suboptimal mortgage choices (Atlas et al. 2017), supporting the behavioral relevance of this potentially Pavlovian bias for everyday decision-making.

Finally, in Chapters 4 and 5, I present evidence that humans can actively shape their environment and seek out positive/ negative cues to strategically invigorate/ suppress their actions. Shaping one's environment and using it as a tool (i.e., as an "extended mind") (Clarke 2010; Risko and Gilbert 2016) means that agents are not merely passive subjects to environmental constraints, but have an active role in triggering Pavlovian biases by stocking their environment with respective cues and/ or via selectively attending to such cues. In the end, not self-control, but "stimulus control" may be a more effective strategy for "resisting temptations" evoked by Pavlovian cues (Duckworth et al. 2016, 2018), a perspective that is hopefully advanced through the work presented in this thesis.

In sum, **Pavlovian biases might be given priority over other decision systems both in uncontrollable environments as well as in situations characterized by particularly high rewards or threats**. Both spatial and temporal proximity might increase the impact of reward and punishment cues on behavior. Beyond these global links, individuals could actively shape their exposure to cues triggering Pavlovian biases, e.g., by stocking their environment with such cues or steering attention to such cues in a strategic manner.

6.6 Implications

The effects of rewards and punishments on behavior are two-fold: One the one hand, they shape future actions based on past behavior through *reinforcement learning*. On the other hand, their expectation energizes current behavior, leading to various phenomena typically subsumed under the term "*motivation*". In this thesis, I have advanced that viewpoint that **many motivational phenomena do not derive from deliberate, propositional reasoning processes that integrate costs and benefits, but rather from automatic phenomena that operate on fast time scales**

and integrate these factors irrespective of whether behavior is able to modify reward/ punishment delivery. This perspective is important given recent positions questioning the usefulness of terms such as "impulsive" vs. "self-controlled" choices, arguing that all choices are essentially just "value-based" (Berkman, Hutcherson, et al. 2017; Berkman, Livingston, et al. 2017). Differences in response speed clearly indicate that some responses are more "automatic" than others (Schneider and Shiffrin 1977). Although (almost) all parts of the brain are involved in value-based decision making and likely perform similar computations (Cisek and Kalaska 2010; Yoo and Hayden 2018), there appears to be functional differentiation amongst prefrontal subregions (Walton et al. 2010; Kennerley et al. 2011; Jocham et al. 2016; Noonan et al. 2017; Hunt et al. 2018; Klein-Flügge et al. 2019). Hence, it is important to distinguish different behavioral strategies—likely implemented by different brain regions—that are not directly visible in behavior. **Only by considering such different strategies of decision-making—some of them likely "automatic" and not under deliberate control—one can start understanding which factors might make humans perform "suboptimal" decisions.**

Pavlovian biases can lead to impulsive approach or behavioral inhibition, two phenomena that are likely involved in the etiology and maintenance of various psychiatric disorders. A growing line of research proposes and reports evidence for altered PIT effects in individuals prone to or suffering from alcohol addiction (Garbusow et al. 2014, 2016, 2019; Sekutowicz et al. 2019; Schad et al. 2020; Sommer et al. 2020; Doñamayor et al. 2021; Sebold et al. 2021; Chen et al. 2023). These PIT effects might correspond to the similar phenomenon of sign-tracking in animals, which constitutes a risk factor for developing addictions (Saunders and Robinson 2013; Flagel and Robinson 2017). Conversely, increased aversive inhibition has been reported in patients suffering from anxiety disorders (Mkrtchian, Aylward, et al. 2017). Finally, reduced biases have been reported in patients suffering from depression, with relatively maintained biases predicting recovery (Huys et al. 2016). This last finding in particular highlights that strong Pavlovian biases might not necessarily be maladaptive, but important contributors to healthy human functioning. In the context of the results that I present in Chapters 4 and 5, it is particularly interesting to quantify individual differences in the extent to which people use selective attention to reward and punishment cues to selectively invigorate Pavlovian biases in some situations, but not others. In sum, there is initial evidence that Pavlovian biases play an important role in psychiatric disorders as well as in healthy functioning. Future studies should make use of better measurement and computational tools to quantify individual differences in these biases, while at the same time considering training regimes that teach how to up- or downregulate their influence.

Lastly, given the pervasive impact of reward and punishment cues on choice, Pavlovian biases should be taken into account when designing (choice) environments (Johnson 2022). Instructing participants to select the better vs. reject the worse of two choice options can markedly alter their sensitivity to positive vs. negative information (Meloy and Russo 2004). In different countries, signups as organ donors are drastically affected by whether the country has an active opt-in vs. a passive opt-out default (Johnson and Goldstein 2003), highlighting the fact that choosing accept vs. reject frames and presenting information as positive vs. negative can have marked effects on citizens' choices. Care needs to be taken in designing respective environments, taking the effects of Pavlovian biases into account.

In sum, I believe that the findings of this thesis have implications for (at least) three distinct fields of society: they should inform future research on the fact that many motivational phenomena are likely automatic and arising from Pavlovian biases; they should inform clinical research in better understanding the etiology and maintenance of psychiatric disorders, with the eventual goal of

designing programs to treat those disorders; and finally, to inform policymakers on the effects of incidental reward and punishment cues as well as accept vs. reject frames on citizens' choice behavior.

6.7 Summary

At the beginning of the General Introduction of this thesis, I have highlighted the millennia-old search for explanations for why humans make seemingly irrational choices. I have pursued the perspective that the mind comprises several "behavioral control" systems that rely on different strategies of information integration and compete for control over behavior. In this thesis, I have focused on one particular system that appears to be especially fast and automatic, but at the same time highly inflexible: the Pavlovian control system. Pavlovian influences have first been described systematically in the animal realm, where incidental reward and punishment cues have been observed to invigorate or suppress ongoing behavior. Over the last years, a similar impact of Pavlovian control on human choice behavior has been mapped. This thesis hopefully contributes to the better understanding of the origin and control over these biases. Using EEG and fMRI, I have observed evidence that **biases arise from—or are at least are shaped by—influences of prefrontal cortical circuits on subcortical striatal circuits**. Furthermore, I have proposed and obtained evidence for the idea that the influence of Pavlovian biases on behavior does not need to be top-down suppressed in situations in which it is maladaptive, but that, instead, **humans can deliberate steer their attention towards reward or punishment cues in the environment and in this way trigger responses that align with their ongoing action plan**. This novel perspective expands our view on the potential adaptiveness of Pavlovian biases: Beyond providing useful action priors in novel or uncontrollable environments, they can aid instrumental goal pursuit in the form of an "auto-pilot" or "training wheels" that do not determine the direction of movement but, once movement has been initiated, support its safe completion and shield it off potential distractors. Under this perspective, Pavlovian control does not oppose instrumental control systems, but instead often acts in concert with them to support goal pursuit.

Previous perspectives have often emphasized the rigid nature of Pavlovian biases, seemingly limiting our range of choice. Beyond this limiting nature, I hope that this thesis helps pointing out a complementary role of these biases in extending our freedom of choice: By strategically focusing attention on reward and punishment cues, humans can recruit an extra "auto-pilot" or "training wheel" that helps enact their action plans even in face of interference. Freedom of will might thus be better characterized as the ability to successfully enact one's goals rather than the complete absence of limiting factors, a perspective advanced by Kant and Hegel (see e.g., Kant's mocking comment on the frictionless "freedom of a turnspit" in his *Critique of Practical Reason*, 5:96–97). In sum, I hope that this thesis conveys that, at times, Pavlovian control appears as a limiting factor leading to impulsive maladaptive choices, while at other times, its existence is crucial to enact our action plans.

Appendices

REFERENCES

Aarts E, Verhage M, Veenvliet JV, Dolan CV, van der Sluis S. 2014. A solution to dependency: Using multilevel analysis to accommodate nested data. Nature Neuroscience. 17:491–496.

Aarts H, Custers R, Wegner DM. 2005. On the inference of personal authorship: Enhancing experienced agency by priming effect information. Consciousness and Cognition. 14:439–458.

Albin RL, Young AB, Penney JB. 1989. The functional anatomy of basal ganglia disorders. Trends in Neurosciences. 12:366–375.

Alexander GE, DeLong MR, Strick PL. 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual Review of Neuroscience. 9:357–381.

Alexander WH, Brown JW. 2011. Medial prefrontal cortex as an action-outcome predictor. Nature Neuroscience. 14:1338–1344.

Alexander WH, Brown JW. 2018. Frontal cortex function as derived from hierarchical predictive coding. Sci Rep. 8:3843.

Algermissen J, den Ouden HEM. 2022. Goal-directed recruitment of Pavlovian biases through selective visual attention. bioRxiv preprint.

Algermissen J, Swart JC, Scheeringa R, Cools R, den Ouden HEM. 2021. Biased credit assignment in motivational learning biases arises through prefrontal influences on striatal learning. bioRxiv preprint.

Algermissen J, Swart JC, Scheeringa R, Cools R, den Ouden HEM. 2022. Striatal BOLD and midfrontal theta power express motivation for action. Cerebral Cortex. 32:2924–2942.

Allen PJ, Josephs O, Turner R. 2000. A method for removing imaging artifact from continuous EEG recorded during functional MRI. NeuroImage. 12:230–239.

Amat J, Baratta MV, Paul E, Bland ST, Watkins LR, Maier SF. 2005. Medial prefrontal cortex determines how stressor controllability affects behavior and dorsal raphe nucleus. Nature Neuroscience. 8:365–371.

Ambroggi F, Ishikawa A, Fields HL, Nicola SM. 2008. Basolateral amygdala neurons facilitate reward-seeking behavior by exciting nucleus accumbens neurons. Neuron. 59:648–661.

Amemori K, Amemori S, Gibson DJ, Graybiel AM. 2018. Striatal microstimulation induces persistent and repetitive negative decision-making predicted by striatal beta-band oscillation. Neuron. 99:829-841.e6.

Amemori K, Amemori S, Gibson DJ, Graybiel AM. 2020. Striatal beta oscillation and neuronal activity in the primate caudate nucleus differentially represent valence and arousal under approach-avoidance conflict. Frontiers in Neuroscience. 14:1–17.

Amemori K, Graybiel AM. 2012. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. Nature Neuroscience. 15:776–785.

Amita H, Hikosaka O. 2019. Indirect pathway from caudate tail mediates rejection of bad objects in periphery. Science Advances. 5:eaaw9297.

Andersson JLR, Jenkinson M, Smith S. 2007. Non-linear registration, aka spatial normalisation. FMRIB Technial Report TR07JA2.

Andreou C, Frielinghaus H, Rauh J, Mußmann M, Vauth S, Braun P, Leicht G, Mulert C. 2017. Theta and high-beta networks for feedback processing: a simultaneous EEG–fMRI study in healthy male subjects. Transl Psychiatry. 7:e1016–e1016.

Antonov PA, Chakravarthi R, Andersen SK. 2020. Too little, too late, and in the wrong place: Alpha band activity does not reflect an active mechanism of selective attention. NeuroImage. 219:117006.

Antzoulatos EG, Miller EK. 2014. Increases in functional connectivity between prefrontal cortex and striatum during category learning. Neuron. 83:216–225.

Anwyl-Irvine A, Dalmaijer ES, Hodges N, Evershed JK. 2021. Realistic precision and accuracy of online experiment platforms, web browsers, and devices. Behav Res. 53:1407–1425.
Armel KC, Beaumel A, Rangel A. 2008. Biasing simple choices by manipulating relative visual attention. Judgment and Decision Making. 3:396–403.
Aron AR. 2011. From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. Biological Psychiatry. 69:e55–e68.
Aron AR, Herz DM, Brown P, Forstmann BU, Zaghloul K. 2016. Frontosubthalamic circuits for control of action and cognition. J Neurosci. 36:11489–11495.
Atlas LY, Doll BB, Li J, Daw ND, Phelps EA. 2016. Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. eLife. 5:e15192.
Atlas SA, Johnson EJ, Payne JW. 2017. Time preferences and mortgage choice. Journal of Marketing Research. 54:415–429.
Aubert I, Ghorayeb I, Normand E, Bloch B. 2000. Phenotypical characterization of the neurons expressing the D1 and D2 dopamine receptors in the monkey striatum. Journal of Comparative Neurology. 418:22–32.
Bakhurin KI, Li X, Friedman AD, Lusk NA, Watson GD, Kim N, Yin HH. 2020. Opponent regulation of action performance and timing by striatonigral and striatopallidal pathways. eLife. 9:e54831.
Balleine BW, Dickinson A. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology. 37:407–419.
Balsters JH, Zerbi V, Sallet J, Wenderoth N, Mars RB. 2020. Primate homologs of mouse cortico-striatal circuits. eLife. 9:e53680.
Baluch F, Itti L. 2011. Mechanisms of top-down attention. Trends in Neurosciences. 34:210–224.
Baron J, Ritov I. 1994. Reference points and omission bias. Organizational Behavior and Human Decision Processes. 59:475–498.
Barr DJ, Levy R, Scheepers C, Tily HJ. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 68:255–278.
Bartra O, McGuire JT, Kable JW. 2013. The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage. 76:412–427.
Bastos AM, Vezoli J, Bosman CA, Schoffelen J-M, Oostenveld R, Dowdall JR, De Weerd P, Kennedy H, Fries P. 2015. Visual areas exert feedforward and feedback influences through distinct frequency channels. Neuron. 85:390–401.
Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. 67:1–48.
Beeler J. 2012. Thorndike's Law 2.0: Dopamine and the regulation of thrift. Frontiers in Neuroscience. 6.
Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. 2007. Learning the value of information in an uncertain world. Nature Neuroscience. 10:1214–1221.
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.
Bekkering H, Neggers SFW. 2002. Visual search is modulated by action intentions. Psychological Science. 13:370–374.
Belin D, Jonkman S, Dickinson A, Robbins TW, Everitt BJ. 2009. Parallel and interactive learning processes within the basal ganglia: Relevance for the understanding of addiction. Behavioural Brain Research. 199:89–102.
Benhabib J, Bisin A, Schotter A. 2010. Present-bias, quasi-hyperbolic discounting, and fixed costs. Games and Economic Behavior. 69:205–223.

Berke JD. 2018. What does dopamine mean? Nature Neuroscience. 21:787–793.
Berkman ET, Hutcherson CA, Livingston JL, Kahn LE, Inzlicht M. 2017. Self-control as value-based choice. Current Directions in Psychological Science. 26:422–428.
Berkman ET, Livingston JL, Kahn LE. 2017. Finding the "self" in self-regulation: The identity-value model. Psychological Inquiry. 28:77–98.
Bernat EM, Nelson LD, Baskin-Sommers AR. 2015. Time-frequency theta and delta measures index separable components of feedback processing in a gambling task. Psychophysiology. 52:626–637.
Bland BH, Oddie SD. 2001. Theta band oscillation and synchrony in the hippocampal formation and associated structures: The case for its role in sensorimotor integration. Behavioural Brain Research. 127:119–136.
Boettcher SEP, Gresch D, Nobre AC, van Ede F. 2021. Output planning at the input stage in visual working memory. Science Advances. 7:eabe8212.
Bongioanni A, Folloni D, Verhagen L, Sallet J, Klein-Flügge MC, Rushworth MFS. 2021. Activation and disruption of a neural mechanism for novel choice in monkeys. Nature. 591:270–274.
Bonnefond M, Jensen O. 2012. Alpha oscillations serve to protect working memory maintenance against anticipated distracters. Current Biology. 22:1969–1974.
Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS. 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.
Botvinick M, Nystrom LE, Fissell K, Carter CS, Cohen JD. 1999. Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature. 402:179–181.
Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. 2001. Conflict monitoring and cognitive control. Psychological Review. 108:624–652.
Botvinick MM, Cohen JD, Carter CS. 2004. Conflict monitoring and anterior cingulate cortex: An update. Trends in Cognitive Sciences. 8:539–546.
Boureau Y-L, Dayan P. 2011. Opponency revisited: Competition and cooperation between dopamine and serotonin. Neuropsychopharmacology. 36:74–97.
Boureau Y-L, Sokol-Hessner P, Daw ND. 2015. Deciding how to decide: Self-control and meta-decision making. Trends in Cognitive Sciences. 19:700–710.
Braem S, King JA, Korb FM, Krebs RM, Notebaert W, Egner T. 2017. The role of anterior cingulate cortex in the affective evaluation of conflict. Journal of Cognitive Neuroscience. 29:137–149.
Braver TS. 2012. The variable nature of cognitive control: A dual mechanisms framework. Trends in Cognitive Sciences. 16:106–113.
Bray S, Rangel A, Shimojo S, Balleine B, O'Doherty JP. 2008. The neural mechanisms underlying the influence of Pavlovian cues on human decision making. Journal of Neuroscience. 28:5861–5866.
Breland K, Breland M. 1961. The misbehavior of organisms. American Psychologist. 16:681–684.
Bressler SL, Tang W, Sylvester CM, Shulman GL, Corbetta M. 2008. Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. Journal of Neuroscience. 28:10056–10061.
Brier MR, Ferree TC, Maguire MJ, Moore P, Spence J, Tillman GD, Hart, Jr. J, Kraut MA. 2010. Frontal theta and alpha power and coherence changes are modulated by semantic complexity in Go/NoGo tasks. International Journal of Psychophysiology. 78:215–224.
Brown PL, Jenkins HM. 1968. Autoshaping of pigeon's key-peck. Journal of the Experimental Analysis of Behavior. 11:1–8.

Brown VM, Chen J, Gillan CM, Price RB. 2020. Improving the reliability of computational analyses: Model-based planning and its relationship with compulsivity. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 5:601–609.
Buschman TJ, Denovellis EL, Diogo C, Bullock D, Miller EK. 2012. Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. Neuron. 76:838–846.
Bushong B, King LM, Camerer CF, Rangel A. 2010. Pavlovian processes in consumer choice: The physical presence of a good increases willingness-to-pay. American Economic Review. 100:1556–1571.
Callaway F, Rangel A, Griffiths TL. 2021. Fixation patterns in simple choice reflect optimal information sampling. PLOS Computational Biology. 17:e1008863.
Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsch HE, Berger MS, Barbaro NM, Knight RT. 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science. 313:1626–1628.
Caplan JB, Madsen JR, Schulze-Bonhage A, Aschenbrenner-Scheibe R, Newman EL, Kahana MJ. 2003. Human θ oscillations related to sensorimotor integration and spatial learning. J Neurosci. 23:4726–4736.
Carlén M. 2017. What constitutes the prefrontal cortex? Science. 358:478–482.
Carter CS, Braver TS, Barch DM, Botvinick MM, Noll D, Cohen JD. 1998. Anterior cingulate cortex, error detection, and the online monitoring of performance. Science. 280:747–749.
Carver CS, White TL. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. Journal of Personality and Social Psychology. 67:319–333.
Cavanagh JF. 2015. Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times. NeuroImage. 110:205–216.
Cavanagh JF, Eisenberg I, Guitart-Masip M, Huys QJM, Frank MJ. 2013. Frontal theta overrides Pavlovian learning biases. Journal of Neuroscience. 33:8541–8548.
Cavanagh JF, Figueroa CM, Cohen MX, Frank MJ. 2012. Frontal theta reflects uncertainty and unexpectedness during exploration and exploitation. Cerebral Cortex. 22:2575–2586.
Cavanagh JF, Frank MJ. 2014. Frontal theta as a mechanism for cognitive control. Trends in Cognitive Sciences. 18:414–421.
Cavanagh JF, Frank MJ, Klein TJ, Allen JJB. 2010. Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. NeuroImage. 49:3198–3209.
Cavanagh JF, Wiecki TV, Cohen MX, Figueroa CM, Samanta J, Sherman SJ, Frank MJ. 2011. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nature Neuroscience. 14:1462–1467.
Cavanagh JF, Wiecki TV, Kochar A, Frank MJ. 2014. Eye tracking and pupillometry are indicators of dissociable latent decision processes. Journal of Experimental Psychology: General. 143:1476–1488.
Cavanagh JF, Zambrano-Vazquez L, Allen JJB. 2012. Theta lingua franca: A common mid-frontal substrate for action monitoring processes. Psychophysiology. 49:220–238.
Cazé RD, van der Meer MAA. 2013. Adaptive properties of differential learning rates for positive and negative outcomes. Biological Cybernetics. 107:711–719.
Cerri DH, Saddoris MP, Carelli RM. 2014. Nucleus accumbens core neurons encode value-independent associations necessary for sensory preconditioning. Behavioral Neuroscience. 128:567–578.
Cervera RL, Wang MZ, Hayden BY. 2020. Systems neuroscience of curiosity. Current Opinion in Behavioral Sciences. 35:48–55.
Chambon V, Théro H, Vidal M, Vandendriessche H, Haggard P, Palminteri S. 2020. Information about action outcomes differentially affects learning from self-determined versus imposed choices. Nat Hum Behav. 4:1067–1079.

Chang LJ, Jolly E, Cheong JH, Rapuano KM, Greenstein N, Chen P-HAHA, Manning JR. 2021. Endogenous variation in ventromedial prefrontal cortex state dynamics during naturalistic viewing reflects affective experience. Science Advances. 7:eabf7129.
Chau BKH, Kolling N, Hunt LT, Walton ME, Rushworth MFS. 2014. A neural mechanism underlying failure of optimal choice with multiple alternatives. Nature Neuroscience. 17:463–470.
Chen H, Belanger MJ, Garbusow M, Kuitunen-Paul S, Huys QJM, Heinz A, Rapp MA, Smolka MN. 2023. Susceptibility to interference between Pavlovian and instrumental control predisposes risky alcohol use developmental trajectory from ages 18 to 24. Addiction Biology. 28:e13263.
Chen Z, Veling H, Dijksterhuis A, Holland RW. 2016. How does not responding to appetitive stimuli cause devaluation: Evaluative conditioning or response inhibition? Journal of Experimental Psychology: General. 145:1687–1701.
Chiew KS. 2021. Revisiting positive affect and reward influences on cognitive control. Current Opinion in Behavioral Sciences. 39:27–33.
Chiew KS, Braver TS. 2014. Dissociable influences of reward motivation and positive emotion on cognitive control. Cognitive, Affective, & Behavioral Neuroscience. 14:509–529.
Cisek P. 2019. Resynthesizing behavior through phylogenetic refinement. Atten Percept Psychophys. 81:2265–2287.
Cisek P. 2020. Evolution of behavioural control from chordates to primates. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 377:20200522.
Cisek P, Kalaska JF. 2010. Neural mechanisms for interacting with a world full of action choices. Annual Review of Neuroscience. 33:269–298.
Cisek P, Pastor-Bernier A. 2014. On the challenges and mechanisms of embodied decisions. Philosophical Transactions of the Royal Society B: Biological Sciences. 369:20130479.
Clarke A. 2010. Supersizing the mind: Embodiment, action, and cognitive extension. Oxford, UK: Oxford University Press.
Clarke HF, Cardinal RN, Rygula R, Hong YT, Fryer TD, Sawiak SJ, Ferrari V, Cockcroft G, Aigbirhio FI, Robbins TW, Roberts AC. 2014. Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behavior via changes in reinforcement sensitivity. Journal of Neuroscience. 34:7663–7676.
Cockburn J, Collins AGE, Frank MJ. 2014. A reinforcement learning mechanism responsible for the valuation of free choice. Neuron. 83:551–557.
Coddington LT, Dudman JT. 2018. The timing of action determines reward prediction signals in identified midbrain dopamine neurons. Nature Neuroscience. 21:1563–1573.
Coddington LT, Dudman JT. 2019. Learning from action: Reconsidering movement signaling in midbrain dopamine neuron activity. Neuron. 104:63–77.
Cohen JD. 2017. Cognitive control. Core constructs and current considerations. In: Egner T, editor. The Wiley Handbook of Cognitive Control. Chichester, UK: John Wiley & Sons, Ltd. p. 1–28.
Cohen JD, Dunbar K, McClelland JL. 1990. On the control of automatic processes: A parallel distributed processing account of the Stroop effect. Psychological Review. 97:332–361.
Cohen M, Cavanagh JF. 2011. Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. Frontiers in Psychology. 2.
Cohen MX. 2014. A neural microcircuit for cognitive conflict detection and signaling. Trends in Neurosciences. 37:480–490.
Cohen MX, Axmacher N, Lenartz D, Elger CE, Sturm V, Schlaepfer TE. 2009. Neuroelectric signatures of reward learning and decision-making in the human nucleus accumbens. Neuropsychopharmacology. 34:1649–1658.

Cohen MX, Cavanagh JF, Slagter HA. 2011. Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity: Commentary. Human Brain Mapping. 32:2270–2271.

Cohen MX, Donner TH. 2013. Midfrontal conflict-related theta-band power reflects neural oscillations that predict behavior. Journal of Neurophysiology. 110:2752–2763.

Cohen MX, Ridderinkhof KR. 2013. EEG source reconstruction reveals frontal-parietal dynamics of spatial conflict processing. PLOS ONE. 8:e57293.

Cohen MX, van Gaal S. 2014. Subthreshold muscle twitches dissociate oscillatory neural signatures of conflicts from errors. NeuroImage. 86:503–513.

Cohen MX, Wilmes KA, van de Vijver I. 2011. Cortical electrophysiological network dynamics of feedback learning. Trends in Cognitive Sciences. 15:558–566.

Colaizzi JM, Flagel SB, Joyner MA, Gearhardt AN, Stewart JL, Paulus MP. 2020. Mapping sign-tracking and goal-tracking onto human behaviors. Neuroscience & Biobehavioral Reviews. 111:84–94.

Collins AGE, Frank MJ. 2014. Opponent actor learning (OpAL): Modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychological Review. 121:337–366.

Cools R. 2016. The costs and benefits of brain dopamine for cognitive control. Wiley Interdisciplinary Reviews: Cognitive Science. 7:317–329.

Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. Nature Reviews Neuroscience. 3:201–215.

Corbit LH, Balleine BW. 2005. Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-Instrumental Transfer. Journal of Neuroscience. 25:962–970.

Corbit LH, Balleine BW. 2011. The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell. Journal of Neuroscience. 31:11786–11794.

Corbit LH, Janak PH. 2007. Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. Journal of Neuroscience. 27:13977–13981.

Corbit LH, Janak PH, Balleine BW. 2007. General and outcome-specific forms of Pavlovian-instrumental transfer: the effect of shifts in motivational state and inactivation of the ventral tegmental area. European Journal of Neuroscience. 26:3141–3149.

Craighero L, Fadiga L, Rizzolatti G, Umiltà C. 1999. Action for perception: A motor-visual attentional effect. Journal of Experimental Psychology: Human Perception and Performance. 25:1673–1692.

Crittenden JR, Tillberg PW, Riad MH, Shima Y, Gerfen CR, Curry J, Housman DE, Nelson SB, Boyden ES, Graybiel AM. 2016. Striosome–dendron bouquets highlight a unique striatonigral circuit targeting dopamine-containing neurons. Proceedings of the National Academy of Sciences. 113:11318–11323.

Csifcsák G, Melsæter E, Mittner M. 2020. Intermittent absence of control during reinforcement learning interferes with Pavlovian bias in action selection. Journal of Cognitive Neuroscience. 32:646–663.

da Silva JA, Tecuapetla F, Paixão V, Costa RM. 2018. Dopamine neuron activity before action initiation gates and invigorates future movements. Nature. 554:244–248.

Dalal SS, Jerbi K, Bertrand O, Adam C, Ducorps A, Denis S, Martinerie J, Lachaux J-P. 2013. Simultaneous MEG-intracranial EEG: New insights into the ability of MEG to capture oscillatory modulations in the neocortex and the hippocampus. Epilepsy and Behavior. 28:283–302.

Danjo T, Yoshimi K, Funabiki K, Yawata S, Nakanishi S. 2014. Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. Proceedings of the National Academy of Sciences. 111:6455–6460.

Darmani G, Bergmann TO, Butts Pauly K, Caskey CF, de Lecea L, Fomenko A, Fouragnan E, Legon W, Murphy KR, Nandi T, Phipps MA, Pinton G, Ramezanpour H, Sallet J, Yaakub SN, Yoo SS, Chen R. 2022. Non-invasive transcranial ultrasound stimulation for neuromodulation. Clinical Neurophysiology. 135:51–73.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron. 69:1204–1215.

Daw ND, Kakade S, Dayan P. 2002. Opponent interactions between serotonin and dopamine. Neural Networks. 15:603–616.

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nature Neuroscience. 8:1704–1711.

Dayan P. 2012. How to set the switches on this thing. Current Opinion in Neurobiology. 22:1068–1074.

Dayan P, Berridge KC. 2014. Model-based and model-free Pavlovian reward learning: Revaluation, revision, and revelation. Cognitive, Affective, & Behavioral Neuroscience. 14:473–492.

Dayan P, Daw ND. 2008. Decision theory, reinforcement learning, and the brain. Cognitive, Affective, & Behavioral Neuroscience. 8:429–453.

Dayan P, Niv Y, Seymour B, Daw N. 2006. The misbehavior of value and the discipline of the will. Neural Networks. 19:1153–1160.

de Boer L, Axelsson J, Chowdhury R, Riklund K, Dolan RJ, Nyberg L, Bäckman L, Guitart-Masip M. 2019. Dorsal striatal dopamine D1 receptor availability predicts an instrumental bias in action learning. Proceedings of the National Academy of Sciences. 116:261–270.

De Martino B, Kumaran D, Seymour B, Dolan RJ. 2006. Frames, biases, and rational decision-making in the human brain. Science. 313:684–687.

Debener S, Ullsperger M, Siegel M, Fiehler K, von Cramon DY, Engel AK. 2005. Trial-by-trial coupling of concurrent electroencephalogram and functional magnetic resonance imaging identifies the dynamics of performance monitoring. Journal of Neuroscience. 25:11730–11737.

DeCoteau WE, Thorn C, Gibson DJ, Courtemanche R, Mitra P, Kubota Y, Graybiel AM. 2007a. Learning-related coordination of striatal and hippocampal theta rhythms during acquisition of a procedural maze task. Proceedings of the National Academy of Sciences of the United States of America. 104:5644–5649.

DeCoteau WE, Thorn C, Gibson DJ, Courtemanche R, Mitra P, Kubota Y, Graybiel AM. 2007b. Oscillations of local field potentials in the rat dorsal striatum during spontaneous and instructed behaviors. Journal of Neurophysiology. 97:3800–3805.

DeLong MR. 1990. Primate models of movement disorders of basal ganglia origin. Trends in Neurosciences. 13:281–285.

den Ouden HEM, Swart JC, Schmidt K, Fekkes D, Geurts DEM, Cools R. 2015. Acute serotonin depletion releases motivated inhibition of response vigour. Psychopharmacology. 232:1303–1312.

Dickinson A. 1986. Re-examination of the role of the instrumental contingency in the sodium-appetite irrelevant incentive effect. The Quarterly Journal of Experimental Psychology Section B. 38:161–172.

Dickinson A, Balleine B. 1994. Motivational control of goal-directed action. Animal Learning & Behavior. 22:1–18.

Dickinson A, Smith J, Mirenowicz J. 2000. Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behavioral Neuroscience. 114:468–483.

Dignath D, Pfister R, Eder AB, Kiesel A, Kunde W. 2014. Something in the way she moves—movement trajectories reveal dynamics of self-control. Psychonomic Bulletin & Review. 21:809–816.

Dixon ML, Christoff K. 2012. The decision to engage cognitive control is driven by expected reward-value: Neural and behavioral evidence. PLoS ONE. 7:e51637.

Doll BB, Hutchison KE, Frank MJ. 2011. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. Journal of Neuroscience. 31:6188–6198.

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ. 2009. Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. Brain Research. 1299:74–94.

Domenech P, Rheims S, Koechlin E. 2020. Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. Science. 369:eabb0184.

Doñamayor N, Ebrahimi C, Garbusow M, Wedemeyer F, Schlagenhauf F, Heinz A. 2021. Instrumental and Pavlovian mechanisms in alcohol use disorder. Current Addiction Reports. 8:156–180.

Donner TH, Siegel M, Fries P, Engel AK. 2009. Buildup of choice-predictive activity in human motor cortex during perceptual decision making. Current Biology. 19:1581–1585.

Dorfman HM, Gershman SJ. 2019. Controllability governs the balance between Pavlovian and instrumental action selection. Nat Commun. 10:5826.

Doya K. 2000. Complementary roles of basal ganglia and cerebellum in learning and motor control. Current Opinion in Neurobiology. 10:732–739.

Dreisbach G, Goschke T. 2004. How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. Journal of Experimental Psychology: Learning, Memory, and Cognition. 30:343–353.

Duckworth AL, Gendler TS, Gross JJ. 2016. Situational strategies for self-control. Perspectives on Psychological Science. 11:35–55.

Duckworth AL, Milkman KL, Laibson D. 2018. Beyond willpower: Strategies for reducing failures of self-control. Psychological Science in the Public Interest. 19:102–129.

Dutra IC, Waller DA, Wessel JR. 2018. Perceptual surprise improves action stopping by nonselectively suppressing motor activity via a neural mechanism for motor inhibition. J Neurosci. 38:1482–1492.

Eimer M, Kiss M. 2008. Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. Journal of Cognitive Neuroscience. 20:1423–1433.

Enel P, Wallis JD, Rich EL. 2020. Stable and dynamic representations of value in the prefrontal cortex. eLife. 9:e54313.

Engel AK, Fries P. 2010. Beta-band oscillations—signalling the status quo? Current Opinion in Neurobiology. 20:156–165.

Engelhard B, Finkelstein J, Cox J, Fleming W, Jang HJ, Ornelas S, Koay SA, Thiberge SY, Daw ND, Tank DW, Witten IB. 2019. Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. Nature. 570:509–513.

Eshel N, Tian J, Bukwich M, Uchida N. 2016. Dopamine neurons share common response function for reward prediction error. Nature Neuroscience. 19:479–486.

Estes WK. 1943. Discriminative conditioning. I. A discriminative property of conditioned anticipation. Journal of Experimental Psychology. 32:150–155.

Estes WK. 1948. Discriminative conditioning. II. Effects of a Pavlovian conditioned stimulus upon a subsequently established operant response. Journal of Experimental Psychology. 38:173–177.

Evans RC, Twedell EL, Zhu M, Ascencio J, Zhang R, Khaliq ZM. 2020. Functional dissection of basal ganglia inhibitory inputs onto substantia nigra dopaminergic neurons. Cell Reports. 32.

Fagioli S, Hommel B, Schubotz RI. 2007. Intentional control of attention: action planning primes action-related stimulus dimensions. Psychological Research. 71:22–29.
Failing M, Nissens T, Pearson D, Le Pelley M, Theeuwes J. 2015. Oculomotor capture by stimuli that signal the availability of reward. Journal of Neurophysiology. 114:2316–2327.
Fawcett TW, Fallenstein B, Higginson AD, Houston AI, Mallpress DEW, Trimmer PC, McNamara JM. 2014. The evolution of decision rules in complex environments. Trends in Cognitive Sciences. 18:153–161.
Feingold J, Gibson DJ, Depasquale B, Graybiel AM. 2015. Bursts of beta oscillation differentiate postperformance activity in the striatum and motor cortex of monkeys performing movement tasks. Proceedings of the National Academy of Sciences of the United States of America. 112:13687–13692.
Fellner M-C, Volberg G, Mullinger KJ, Goldhacker M, Wimber M, Greenlee MW, Hanslmayr S. 2016. Spurious correlations in simultaneous EEG-fMRI driven by in-scanner movement. NeuroImage. 133:354–366.
Ferry AT, Öngür D, An X, Price JL. 2000. Prefrontal cortical projections to the striatum in macaque monkeys: Evidence for an organization related to prefrontal networks. Journal of Comparative Neurology. 425:447–470.
Fiedler S, Glöckner A. 2012. The dynamics of decision making in risky choice: An eye-tracking analysis. Frontiers in Psychology. 3.
Fiedler S, Schulte-Mecklenbeck M, Renkewitz F, Orquin JL. 2020. Guideline for reporting standards of eye-tracking research in decision sciences. PsyArXiv.
Figner B, Knoch D, Johnson EJ, Krosch AR, Lisanby SH, Fehr E, Weber EU. 2010. Lateral prefrontal cortex and self-control in intertemporal choice. Nature Neuroscience. 13:538–539.
Fine JM, Hayden BY. 2021. The whole prefrontal cortex is premotor cortex. Philosophical Transactions of the Royal Society B: Biological Sciences. 377:20200524.
Flagel SB, Akil H, Robinson TE. 2009. Individual differences in the attribution of incentive salience to reward-related cues: Implications for addiction. Neuropharmacology. 56:139–148.
Flagel SB, Robinson TE. 2017. Neurobiological basis of individual variation in stimulus-reward learning. Current Opinion in Behavioral Sciences. 13:178–185.
Flagel SB, Robinson TE, Clark JJ, Clinton SM, Watson SJ, Seeman P, Phillips PEMM, Akil H. 2010. An animal model of genetic vulnerability to behavioral disinhibition and responsiveness to reward-related cues: Implications for addiction. Neuropsychopharmacology. 35:388–400.
Flagel SB, Watson SJ, Robinson TE, Akil H. 2007. Individual differences in the propensity to approach signals vs goals promote different adaptations in the dopamine system of rats. Psychopharmacology. 191:599–607.
Floresco SB, Yang CR, Phillips AG, Blaha CD. 1998. Basolateral amygdala stimulation evokes glutamate receptor-dependent dopamine efflux in the nucleus accumbens of the anaesthetized rat. European Journal of Neuroscience. 10:1241–1251.
Folk CL, Remington RW, Johnston JC. 1992. Involuntary covert orienting is contingent on attentional control settings. Journal of Experimental Psychology: Human Perception and Performance. 18:1030–1044.
Folloni D, Fouragnan E, Wittmann MK, Roumazeilles L, Tankelevitch L, Verhagen L, Attali D, Aubry J-F, Sallet J, Rushworth MFS. 2021. Ultrasound modulation of macaque prefrontal cortex selectively alters credit assignment–related activity and behavior. Science Advances. 7:eabg7700.
Fomenko A, Neudorfer C, Dallapiazza RF, Kalia SK, Lozano AM. 2018. Low-intensity ultrasound neuromodulation: An overview of mechanisms and emerging human applications. Brain Stimulation. 11:1209–1217.

Foti D, Weinberg A, Dien J, Hajcak G. 2011. Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. Human Brain Mapping. 32:2207–2216.
Fouragnan E, Retzler C, Mullinger K, Philiastides MG. 2015. Two spatiotemporally distinct value systems shape reward-based learning in the human brain. Nat Commun. 6:8107.
Fouragnan E, Retzler C, Philiastides MG. 2018. Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. Human Brain Mapping. 39:2887–2906.
Fouragnan EF, Chau BKH, Folloni D, Kolling N, Verhagen L, Klein-Flügge M, Tankelevitch L, Papageorgiou GK, Aubry J-F, Sallet J, Rushworth MFS. 2019. The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. Nat Neurosci. 22:797–808.
Frank MJ. 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.
Frank MJ. 2006. Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. Neural Networks. 19:1120–1136.
Frank MJ, Gagne C, Nyhus E, Masters S, Wiecki TV, Cavanagh JF, Badre D. 2015. fMRI and EEG predictors of dynamic decision parameters during human reinforcement learning. Journal of Neuroscience. 35:485–494.
Frank MJ, Loughry B, O'Reilly RC. 2001. Interactions between frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, & Behavioral Neuroscience. 1:137–160.
Frank MJ, Rudy JW, Levy WB, O'Reilly RC. 2005. When logic fails: Implicit transitive inference in humans. Memory & Cognition. 33:742–750.
Frank MJ, Woroch BS, Curran T. 2005. Error-related negativity predicts reinforcement learning and conflict biases. Neuron. 47:495–501.
Friston KJ, Daunizeau J, Kilner J, Kiebel SJ. 2010. Action and behavior: a free-energy formulation. Biological Cybernetics. 102:227–260.
Fröber K, Dreisbach G. 2014. The differential influences of positive affect, random reward, and performance-contingent reward on cognitive control. Cognitive, Affective, & Behavioral Neuroscience. 14:530–547.
Fröber K, Dreisbach G. 2016. How performance (non-)contingent reward modulates cognitive control. Acta Psychologica. 168:65–77.
Frömer R, Dean Wolf CK, Shenhav A. 2019. Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. Nat Commun. 10:4926.
Gallivan JP, Chapman CS, Gale DJ, Flanagan JR, Culham JC. 2019. Selective modulation of early visual cortical activity by movement intention. Cerebral Cortex. 29:4662–4678.
Garbusow M, Nebe S, Sommer C, Kuitunen-Paul S, Sebold M, Schad DJ, Friedel E, Veer IM, Wittchen H-U, Rapp MA, Ripke S, Walter H, Huys QJM, Schlagenhauf F, Smolka MN, Heinz A. 2019. Pavlovian-to-instrumental transfer and alcohol consumption in young male social drinkers: Behavioral, neural and polygenic correlates. Journal of Clinical Medicine. 8:1188.
Garbusow M, Schad DJ, Sebold M, Friedel E, Bernhardt N, Koch SP, Steinacher B, Kathmann N, Geurts DEM, Sommer C, Müller DK, Nebe S, Paul S, Wittchen H-U, Zimmermann US, Walter H, Smolka MN, Sterzer P, Rapp MA, Huys QJM, Schlagenhauf F, Heinz A. 2016. Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. Addiction Biology. 21:719–731.
Garbusow M, Schad DJ, Sommer C, Jünger E, Sebold M, Friedel E, Wendt J, Kathmann N, Schlagenhauf F, Zimmermann US, Heinz A, Huys QJM, Rapp MA. 2014. Pavlovian-to-

instrumental transfer in alcohol dependence: A pilot study. Neuropsychobiology. 70:111–121.
Garofalo S, di Pellegrino G. 2015. Individual differences in the influence of task-irrelevant Pavlovian cues on human behavior. Frontiers in Behavioral Neuroscience. 9.
Gawronski B, Bodenhausen GV. 2006. Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. Psychological Bulletin. 132:692–731.
Geng JJ. 2014. Attentional mechanisms of distractor suppression. Current Directions in Psychological Science. 23:147–153.
Gerfen CR. 1992. The neostriatal mosaic: Multiple levels of compartmental organization in the basal ganglia. Annual Review of Neuroscience. 15:285–320.
Gershman SJ. 2014. Dopamine ramps are a consequence of reward prediction errors. Neural Computation. 26:467–471.
Gershman SJ, Markman AB, Otto AR. 2014. Retrospective revaluation in sequential decision making: A tale of two systems. Journal of Experimental Psychology: General. 143:182–194.
Geurts DEM, Huys QJM, den Ouden HEM, Cools R. 2013a. Serotonin and aversive Pavlovian control of instrumental behavior in humans. Journal of Neuroscience. 33:18932–18939.
Geurts DEM, Huys QJM, den Ouden HEM, Cools R. 2013b. Aversive Pavlovian control of instrumental behavior in humans. Journal of Cognitive Neuroscience. 25:1428–1441.
Gilbert SJ. 2015. Strategic offloading of delayed intentions into the external environment. Quarterly Journal of Experimental Psychology. 68:971–992.
Glickman M, Tsetsos K, Usher M. 2018. Attentional selection mediates framing and risk-bias effects. Psychological Science. 29:2010–2019.
Glickman SE, Schiff BB. 1967. A biological theory of reinforcement. Psychological Review. 74:81–109.
Gluth S, Rieskamp J, Büchel C. 2013. Deciding not to decide: Computational and neural evidence for hidden behavior in sequential choice. PLOS Computational Biology. 9:e1003309.
Gold JI, Shadlen MN. 2007. The neural basis of decision making. Annual Review of Neuroscience. 30:535–574.
Gomez P, Ratcliff R, Perea M. 2007. A model of the go/no-go task. Journal of Experimental Psychology: General. 136:389–413.
Gottlieb J. 2018. Understanding active sampling strategies: Empirical approaches and implications for attention and decision research. Cortex. 102:150–160.
Gottlieb J, Oudeyer P-Y. 2018. Towards a neuroscience of active sampling and curiosity. Nature Reviews Neuroscience. 22:541–548.
Gratton G, Coles MGH, Donchin E. 1992. Optimizing the use of information: Strategic control of activation of responses. Journal of Experimental Psychology: General. 121:480–506.
Gregoriou GG, Gotts SJ, Zhou H, Desimone R. 2009. High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science. 324:1207–1210.
Grogan JP, Sandhu TR, Hu MT, Manohar SG. 2020. Dopamine promotes instrumental motivation, but reduces reward-related vigour. eLife. 9:e58321.
Gueguen MCM, Lopez-Persem A, Billeke P, Lachaux J, Rheims S, Kahane P, Minotti L, David O, Pessiglione M, Bastin J. 2021. Anatomical dissociation of intracerebral signals for reward and punishment prediction errors in humans. Nature Communications. 12:3344.
Guitart-Masip M, Beierholm UR, Dolan R, Duzel E, Dayan P. 2011. Vigor in the face of fluctuating rates of reward: An experimental examination. Journal of Cognitive Neuroscience. 23:3933–3938.

Guitart-Masip M, Chowdhury R, Sharot T, Dayan P, Duzel E, Dolan RJ. 2012. Action controls dopaminergic enhancement of reward representations. Proceedings of the National Academy of Sciences. 109:7511–7516.
Guitart-Masip M, Duzel E, Dolan R, Dayan P. 2014. Action versus valence in decision making. Trends in Cognitive Sciences. 18:194–202.
Guitart-Masip M, Economides M, Huys QJM, Frank MJ, Chowdhury R, Duzel E, Dayan P, Dolan RJ. 2014. Differential, but not opponent, effects of l-DOPA and citalopram on action learning with reward and punishment. Psychopharmacology. 231:955–966.
Guitart-Masip M, Fuentemilla L, Bach DR, Huys QJM, Dayan P, Dolan RJ, Duzel E. 2011. Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. Journal of Neuroscience. 31:7867–7875.
Guitart-Masip M, Huys QJM, Fuentemilla L, Dayan P, Duzel E, Dolan RJ. 2012. Go and no-go learning in reward and punishment: Interactions between affect and effect. NeuroImage. 62:154–166.
Gundlach C, Moratti S, Forschack N, Müller MM. 2020. Spatial attentional selection modulates early visual stimulus processing independently of visual alpha modulations. Cerebral Cortex. 30:3686–3703.
Gurney K, Prescott TJ, Redgrave P. 2001. A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biological Cybernetics. 84:401–410.
Gutteling TP, Petridou N, Dumoulin SO, Harvey BM, Aarnoutse EJ, Kenemans JL, Neggers SFW. 2015. Action preparation shapes processing in early visual cortex. Journal of Neuroscience. 35:6472–6480.
Haber SN. 2003. The primate basal ganglia: Parallel and integrative networks. Journal of Chemical Neuroanatomy. 26:317–330.
Haber SN, Behrens TEJ. 2014. The neural network underlying incentive-based learning: Implications for interpreting circuit disruptions in psychiatric disorders. Neuron. 83:1019–1039.
Haber SN, Fudge JL, McFarland NR. 2000. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. Journal of Neuroscience. 20:2369–2382.
Haber SN, Knutson B. 2010. The reward circuit: Linking primate anatomy and human imaging. Neuropsychopharmacology. 35:4–26.
Haber SN, Kunishio K, Mizobuchi M, Lynd-Balta E. 1995. The orbital and medial prefrontal circuit through the primate basal ganglia. J Neurosci. 15:4851–4867.
Haegens S, Cousijn H, Wallis G, Harrison PJ, Nobre AC. 2014. Inter- and intra-individual variability in alpha peak frequency. NeuroImage. 92:46–55.
Halbout B, Marshall AT, Azimi A, Liljeholm M, Mahler SV, Wassum KM, Ostlund SB. 2019. Mesolimbic dopamine projections mediate cue-motivated reward seeking but not reward retrieval in rats. eLife. 8:e43551.
Hall J, Parkinson JA, Connor TM, Dickinson A, Everitt BJ. 2001. Involvement of the central nucleus of the amygdala and nucleus accumbens core in mediating Pavlovian influences on instrumental behaviour. European Journal of Neuroscience. 13:1984–1992.
Hamel L, Cavdaroglu B, Yeates D, Nguyen D, Riaz S, Patterson D, Khan N, Kirolos N, Roper K, Ha QA, Ito R. 2022. Cortico-striatal control over adaptive goal-directed responding elicited by cues signaling sucrose reward or punishment. J Neurosci. 42:3811–3822.
Hamid AA, Frank MJ, Moore CI. 2021. Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment. Cell. 184:2733-2749.e16.
Hamid AA, Pettibone JR, Mabrouk OS, Hetrick VL, Schmidt R, Vander Weele CM, Kennedy RT, Aragona BJ, Berke JD. 2016. Mesolimbic dopamine signals the value of work. Nature Neuroscience. 19:117–126.

Hanslmayr S, Pastötter B, Bäuml K-H, Gruber S, Wimber M, Klimesch W. 2008. The electrophysiological dynamics of interference during the Stroop task. Journal of Cognitive Neuroscience. 20:215–225.
Harris A, Adolphs R, Camerer C, Rangel A. 2011. Dynamic construction of stimulus values in the ventromedial prefrontal cortex. PLoS ONE. 6:e21074.
Harris A, Hare T, Rangel A. 2013. Temporally dissociable mechanisms of self-control: Early attentional filtering versus late value modulation. Journal of Neuroscience. 33:18917–18931.
Haselton MG, Bryant GA, Wilke A, Frederick DA, Galperin A, Frankenhuis WE, Moore T. 2009. Adaptive rationality: An evolutionary perspective on cognitive bias. Social Cognition. 27:733–763.
Haselton MG, Nettle D. 2006. The paranoid optimist: An integrative evolutionary model of cognitive biases. Pers Soc Psychol Rev. 10:47–66.
Hauser TU, Hunt LT, Iannaccone R, Walitza S, Brandeis D, Brem S, Dolan RJ. 2015. Temporally dissociable contributions of human medial prefrontal subregions to reward-guided learning. Journal of Neuroscience. 35:11209–11220.
Hauser TU, Iannaccone R, Stämpfli P, Drechsler R, Brandeis D, Walitza S, Brem S. 2014. The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization. NeuroImage. 84:159–168.
Hayden BY, Moreno-Bote R. 2018. A neuronal theory of sequential economic choice. Brain and Neuroscience Advances. 2:2398212818766675.
Hebart MN, Gläscher J. 2015. Serotonin and dopamine differentially affect appetitive and aversive general Pavlovian-to-instrumental transfer. Psychopharmacology. 232:437–451.
Helfrich RF, Fiebelkorn IC, Szczepanski SM, Lin JJ, Parvizi J, Knight RT, Kastner S. 2018. Neural mechanisms of sustained attention are rhythmic. Neuron. 99:854-865.e5.
Hernández-López S, Bargas J, Surmeier DJ, Reyes A, Galarraga E. 1997. D 1 receptor activation enhances evoked discharge in neostriatal medium spiny neurons by modulating an L-type Ca 2+ conductance. J Neurosci. 17:3334–3342.
Hernandez-Lopez S, Tkatch T, Perez-Garci E, Galarraga E, Bargas J, Hamm H, Surmeier DJ, Hernádez-López S, Tkatch T, Perez-Garci E, Galarraga E, Bargas J, Hamm H, Surmeier DJ. 2000. D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca2+ currents and excitability via a novel PLCβ1-IP3-Calcineurin-signaling cascade. Journal of Neuroscience. 20:8987–8995.
Herry C, Ciocchi S, Senn V, Demmou L, Müller C, Lüthi A. 2008. Switching on and off fear by distinct neuronal circuits. Nature. 454:600–606.
Hershberger WA. 1986. An approach through the looking-glass. Animal Learning & Behavior. 14:443–451.
Herz DM, Zavala BA, Bogacz R, Brown P. 2016. Neural correlates of decision thresholds in the human subthalamic nucleus. Current Biology. 26:916–920.
Heuer A, Crawford JD, Schubö A. 2017. Action relevance induces an attentional weighting of representations in visual working memory. Memory & Cognition. 45:413–427.
Heuer A, Ohl S, Rolfs M. 2020. Memory for action: a functional view of selection in visual working memory. Visual Cognition. 28:388–400.
Heuer A, Schubö A. 2017. Selective weighting of action-related feature dimensions in visual working memory. Psychonomic Bulletin and Review. 24:1129–1134.
Hickey C, Chelazzi L, Theeuwes J. 2010. Reward changes salience in human vision via the anterior cingulate. Journal of Neuroscience. 30:11096–11103.
Hikosaka O. 1998. Neural systems for control of voluntary action--A hypothesis. Advances in Biophysics. 35:81–102.

Hikosaka O, Kim HF, Amita H, Yasuda M, Isoda M, Tachibana Y, Yoshida A. 2019. Direct and indirect pathways for choosing objects and actions. European Journal of Neuroscience. 49:637–645.

Hillebrand A, Barnes GR. 2002. A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. NeuroImage. 16:638–650.

Holmes NM, Marchand AR, Coutureau E. 2010. Pavlovian to instrumental transfer: A neurobehavioural perspective. Neuroscience & Biobehavioral Reviews. 34:1277–1295.

Hong H, Yamins DLK, Majaj NJ, DiCarlo JJ. 2016. Explicit information for category-orthogonal object properties increases along the ventral stream. Nature Neuroscience. 19:613–622.

Horvitz JC, Stewart T, Jacobs BL. 1997. Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. Brain Research. 759:251–258.

Howard JD, Reynolds R, Smith DE, Voss JL, Schoenbaum G, Kahnt T. 2020. Targeted stimulation of human orbitofrontal networks disrupts outcome-guided behavior. Current Biology. 30:490-498.e4.

Howe MW, Dombeck DA. 2016. Rapid signalling in distinct dopaminergic axons during locomotion and reward. Nature. 535:505–510.

Howe MW, Tierney PL, Sandberg SG, Phillips PEM, Graybiel AM. 2013. Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. Nature. 500:575–579.

Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MFS, Behrens TEJ. 2012. Mechanisms underlying cortical activity during value-guided choice. Nature Neuroscience. 15:470–476.

Hunt LT, Malalasekera WMN, de Berker AO, Miranda B, Farmer SF, Behrens TEJ, Kennerley SW. 2018. Triple dissociation of attention and decision computations across prefrontal cortex. Nature Neuroscience. 21:1471–1481.

Hunt LT, Rutledge RB, Malalasekera WMN, Kennerley SW, Dolan RJ. 2016. Approach-induced biases in human information sampling. PLOS Biology. 14:e2000638.

Hunt LT, Woolrich MW, Rushworth MFS, Behrens TEJ. 2013. Trial-type dependent frames of reference for value comparison. PLoS Computational Biology. 9:e1003225.

Huster RJ, Enriquez-Geppert S, Lavallee CF, Falkenstein M, Herrmann CS. 2013. Electroencephalography of response inhibition tasks: Functional networks and cognitive contributions. International Journal of Psychophysiology. 87:217–233.

Huys QJM, Cools R, Gölzer M, Friedel E, Heinz A, Dolan RJ, Dayan P. 2011. Disentangling the roles of approach, activation and valence in instrumental and Pavlovian responding. PLoS Computational Biology. 7:e1002028.

Huys QJM, Gölzer M, Friedel E, Heinz A, Cools R, Dayan P, Dolan RJ. 2016. The specificity of Pavlovian regulation is associated with recovery from depression. Psychological Medicine. 46:1027–1035.

Ikemoto S. 2007. Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. Brain Research Reviews. 56:27–78.

Ironside M, Amemori K, McGrath CL, Pedersen ML, Kang MS, Amemori S, Frank MJ, Graybiel AM, Pizzagalli DA. 2020. Approach-avoidance conflict in major depressive disorder: Congruent neural findings in humans and nonhuman primates. Biological Psychiatry. 87:399–408.

Itti L, Koch C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research. 40:1489–1506.

Jang AI, Sharma R, Drugowitsch J. 2021. Optimal policy for attention-modulated decisions explains human fixation behavior. eLife. 10:e63436.

Jenkinson M, Bannister P, Brady M, Smith S. 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. NeuroImage. 17:825–841.

Jensen O, Mazaheri A. 2010. Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. Frontiers in Human Neuroscience. 4.

Jocham G, Brodersen KH, Constantinescu AO, Kahn MC, Ianni AM, Walton M, Rushworth MFS, Behrens TEJ. 2016. Reward-guided learning with and without causal attribution. Neuron. 90:177–190.

Jocham G, Klein TA, Ullsperger M. 2011. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. Journal of Neuroscience. 31:1606–1613.

Joel D, Weiner I. 2000. The connections of the dopaminergic system with the striatum in rats and primates: an analysis with respect to the functional and compartmental organization of the striatum. Neuroscience. 96:451–474.

Johansson R, Johansson M. 2014. Look here, eye movements play a functional role in memory retrieval. Psychological Science. 25:236–242.

Johnson EJ. 2022. The elements of choice: Why the way we decide matters. Penguin Publishing Group.

Johnson EJ, Goldstein D. 2003. Do defaults save lives? Science. 302:1338–1339.

Johnson GM, Valle-Inclán F, Geary DC, Hackley SA. 2012. The nursing hypothesis: An evolutionary account of emotional modulation of the postauricular reflex. Psychophysiology. 49:178–185.

Jones JL, Day JJ, Aragona BJ, Wheeler RA, Wightman RM, Carelli RM. 2010. Basolateral amygdala modulates terminal dopamine release in the Nucleus Accumbens and conditioned responding. Biological Psychiatry. 67:737–744.

Jurkiewicz MT, Gaetz WC, Bostan AC, Cheyne D. 2006. Post-movement beta rebound is generated in motor cortex: Evidence from neuromagnetic recordings. NeuroImage. 32:1281–1289.

Kaanders P, Nili H, O'Reilly JX, Hunt L. 2021. Medial frontal cortex activity predicts information sampling in economic choice. J Neurosci. 41:8403–8413.

Kable JW, Glimcher PW. 2009. The neurobiology of decision: Consensus and controversy. Neuron. 63:733–745.

Kacelnik A, Vasconcelos M, Monteiro T, Aw J. 2011. Darwin's "tug-of-war" vs. starlings' "horse-racing": How adaptations for sequential encounters drive simultaneous choice. Behavioral Ecology and Sociobiology. 65:547–558.

Kahneman D. 2011. Thinking, fast and slow. New York, NY: Farrar, Strauss, and Giroux.

Karalis N, Dejean C, Chaudun F, Khoder S, Rozeske RR, Wurtz H, Bagur S, Benchenane K, Sirota A, Courtin J, Herry C. 2016. 4-Hz oscillations synchronize prefrontal–amygdala circuits during fear behavior. Nature Neuroscience. 19:605–612.

Kawagoe R, Takikawa Y, Hikosaka O. 1998. Expectation of reward modulates cognitive signals in the basal ganglia. Nature Neuroscience. 1:411–416.

Keistler C, Barker JM, Taylor JR. 2015. Infralimbic prefrontal cortex interacts with nucleus accumbens shell to unmask expression of outcome-selective Pavlovian-to-instrumental transfer. Learning & Memory. 22:509–513.

Keistler CR, Hammarlund E, Barker JM, Bond CW, DiLeone RJ, Pittenger C, Taylor JR. 2017. Regulation of alcohol extinction and cue-induced reinstatement by specific projections among medial prefrontal cortex, nucleus accumbens, and basolateral amygdala. J Neurosci. 37:4462–4471.

Kelly SP, O'Connell RG. 2013. Internal and external influences on the rate of sensory evidence accumulation in the human brain. Journal of Neuroscience. 33:19434–19441.

Kennerley SW, Behrens TEJ, Wallis JD. 2011. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nature Neuroscience. 14:1581–1589.
Keramati M, Dezfouli A, Piray P. 2011. Speed/accuracy trade-off between the habitual and the goal-directed processes. PLoS Computational Biology. 7:e1002055.
Kim HR, Malik AN, Mikhael JG, Bech P, Tsutsui-Kimura I, Sun F, Zhang Y, Li Y, Watabe-Uchida M, Gershman SJ, Uchida N. 2020. A unified framework for dopamine signals across timescales. Cell. 183:1600-1616.e25.
King J-R, Dehaene S. 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends in Cognitive Sciences. 18:203–210.
Klein-Flügge MC, Wittmann MK, Shpektor A, Jensen DEA, Rushworth MFS. 2019. Multiple associative structures created by reinforcement and incidental statistical learning mechanisms. Nature Communications. 10:4835.
Klimesch W. 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. Brain Research Reviews. 29:169–195.
Klimesch W, Sauseng P, Hanslmayr S. 2007. EEG alpha oscillations: The inhibition–timing hypothesis. Brain Research Reviews. 53:63–88.
Knudsen EB, Wallis JD. 2020. Closed-loop theta stimulation in the orbitofrontal cortex prevents reward-based learning. Neuron. 106:537-547.e4.
Knutson B, Adams CM, Fong GW, Hommer D. 2001. Anticipation of increasing monetary reward selectively recruits Nucleus Accumbens. J Neurosci. 21:RC159–RC159.
Kolling N, Scholl J, Chekroud A, Trier HA, Rushworth MFS. 2018. Prospection, perseverance, and insight in sequential behavior. Neuron. 99:1069-1082.e7.
Konovalov A, Krajbich I. 2016. Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. Nature Communications. 7:12438.
Krajbich I, Armel C, Rangel A. 2010. Visual fixations and the computation and comparison of value in simple choice. Nature Neuroscience. 13:1292–1298.
Krause F, Benjamins C, Eck J, Lührs M, Hoof R, Goebel R. 2019. Active head motion reduction in magnetic resonance imaging using tactile feedback. Human Brain Mapping. 40:4026–4037.
Kravitz AV, Tye LD, Kreitzer AC. 2012. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nature Neuroscience. 15:816–818.
Krebs RM, Boehler CN, Woldorff MG. 2010. The influence of reward associations on conflict processing in the Stroop task. Cognition. 117:341–347.
Kreussel L, Hewig J, Kretschmer N, Hecht H, Coles MGH, Miltner WHR. 2012. The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. Psychophysiology. 49:207–219.
Kuhn M, Wendt J, Sjouwerman R, Büchel C, Hamm A, Lonsdorf TB. 2020. The neurofunctional basis of affective startle modulation in humans – evidence from combined facial electromyography and functional magnetic resonance imaging. Biological Psychiatry. 87:548–558.
Laeng B, Bloem IM, D'Ascenzo S, Tommasi L. 2014. Scrutinizing visual images: The role of gaze in mental imagery and memory. Cognition. 131:263–283.
Lammel S, Lim BK, Ran C, Huang KW, Betley MJ, Tye KM, Deisseroth K, Malenka RC. 2012. Input-specific control of reward and aversion in the ventral tegmental area. Nature. 491:212–217.
Landau AN, Schreyer HM, van Pelt S, Fries P. 2015. Distributed attention is implemented through theta-rhythmic gamma modulation. Current Biology. 25:2332–2337.
Le Pelley ME, Pearson D, Griffiths O, Beesley T. 2015. When goals conflict with values: Counterproductive attentional and oculomotor capture by reward-related stimuli. Journal of Experimental Psychology: General. 144:158–171.

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.
Lee SW, Shimojo S, O'Doherty JP. 2014. Neural computations underlying arbitration between model-based and model-free learning. Neuron. 81:687–699.
Lefebvre G, Summerfield C, Bogacz R. 2022. A normative account of confirmation bias during reinforcement learning. Neural Computation. 34:307–337.
Lex A, Hauber W. 2008. Dopamine D1 and D2 receptors in the nucleus accumbens core and shell mediate Pavlovian-instrumental transfer. Learning & Memory. 15:483–491.
Lichtenberg NT, Sepe-Forrest L, Pennington ZT, Lamparelli AC, Greenfield VY, Wassum KM. 2021. The medial orbitofrontal cortex–basolateral amygdala circuit regulates the influence of reward cues on adaptive behavior and choice. J Neurosci. 41:7267–7277.
Lichtenberg NT, Wassum KM. 2017. Amygdala mu-opioid receptors mediate the motivating influence of cue-triggered reward expectations. European Journal of Neuroscience. 45:381–387.
Ligneul R, Mainen ZF, Ly V, Cools R. 2022. Stress-sensitive inference of task controllability. Nat Hum Behav. 6:812–822.
Lisman JE, Jensen O. 2013. The theta-gamma neural code. Neuron. 77:1002–1016.
Liu C, Cai X, Ritzau-Jost A, Kramer PF, Li Y, Khaliq ZM, Hallermann S, Kaeser PS. 2022. An action potential initiation mechanism in distal axons for the control of dopamine release. Science. 375:1378–1385.
Loewenstein GF, O'Donoghue T. 2004. Animal spirits: Affective and deliberative processes in economic behavior. SSRN Electronic Journal.
LoLordo VM, McMillan JC, Riley AL. 1974. The effects upon food-reinforced pecking and treadle-pressing of auditory and visual signals for response-independent food. Learning and Motivation. 5:24–41.
Lovibond PF. 1983. Facilitation of instrumental behavior by a Pavlovian appetitive conditioned stimulus. Journal of Experimental Psychology: Animal Behavior Processes. 9:225–247.
Luo S, Ainslie G, Giragosian L, Monterosso JR. 2009. Behavioral and neural evidence of incentive bias for immediate rewards relative to preference-matched delayed rewards. Journal of Neuroscience. 29:14820–14827.
Luque D, Beesley T, Morris RW, Jack BN, Griffiths O, Whitford TJ, Pelley MEL. 2017. Goal-directed and habit-like modulations of stimulus processing during reinforcement learning. J Neurosci. 37:3009–3017.
Ly V, Huys QJM, Stins JF, Roelofs K, Cools R. 2014. Individual differences in bodily freezing predict emotional biases in decision making. Frontiers in Behavioral Neuroscience. 8.
Mahlberg J, Seabrooke T, Weidemann G, Hogarth L, Mitchell CJ, Moustafa AA. 2021. Human appetitive Pavlovian-to-instrumental transfer: A goal-directed account. Psychological Research. 85:449–463.
Manohar S, Husain M. 2013. Attention as foraging for information and value. Frontiers in Human Neuroscience. 7.
Manohar S, Lockwood P, Drew D, Fallon SJ, Chong TT-J, Jeyaretna DS, Baker I, Husain M. 2021. Reduced decision bias and more rational decision making following ventromedial prefrontal cortex damage. Cortex. 138:24–37.
Manohar SG, Finzi RD, Drew D, Husain M. 2017. Distinct motivational effects of contingent and noncontingent rewards. Psychological Science. 28:1016–1026.
Manohar SG, Husain M. 2016. Human ventromedial prefrontal lesions alter incentivisation by reward. Cortex. 76:104–120.
Manssuer L, Qiong D, Wei L, Yang R, Zhang C, Zhao Y, Sun B, Zhan S, Voon V. 2022. Integrated amygdala, orbitofrontal and hippocampal contributions to reward and loss coding revealed with human intracranial EEG. J Neurosci. 42:2756–2771.

Mante V, Sussillo D, Shenoy KV, Newsome WT. 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature. 503:78–84.
Marco-Pallarés J, Cucurell D, Cunillera T, García R, Andrés-Pueyo A, Münte TF, Rodríguez-Fornells A. 2008. Human oscillatory activity associated to reward processing in a gambling task. Neuropsychologia. 46:241–248.
Marco-Pallarés J, Münte TF, Rodríguez-Fornells A. 2015. The role of high-frequency oscillatory activity in reward processing and learning. Neuroscience and Biobehavioral Reviews. 49:1–7.
Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. Journal of Neuroscience Methods. 164:177–190.
Maris E, van Vugt M, Kahana M. 2011. Spatially distributed patterns of oscillatory coupling between high-frequency amplitudes and low-frequency phases in human iEEG. NeuroImage. 54:836–850.
Mars RB, Jbabdi S, Rushworth MFS. 2021. A common space approach to comparative neuroscience. Annual Review of Neuroscience. 44:69–86.
Marshall TR, den Boer S, Cools R, Jensen O, Fallon SJ, Zumer JM. 2018. Occipital alpha and gamma oscillations support complementary mechanisms for processing stimulus value associations. Journal of Cognitive Neuroscience. 30:119–129.
Matsumoto M, Hikosaka O. 2009. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 459:837–841.
Mattar MG, Daw ND. 2018. Prioritized memory access explains planning and hippocampal replay. Nat Neurosci. 21:1609–1617.
McClure SM, Laibson DI, Loewenstein G, Cohen JD. 2004. Separate neural systems value immediate and delayed monetary rewards. Science. 306:503–507.
Meder D, Kolling N, Verhagen L, Wittmann MK, Scholl J, Madsen KH, Hulme OJ, Behrens TEJ, Rushworth MFS. 2017. Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. Nature Communications. 8:1942.
Meier S, Sprenger C. 2010. Present-biased preferences and credit card borrowing. American Economic Journal: Applied Economics. 193–210.
Meloy MG, Russo JE. 2004. Binary choice under instructions to select versus reject. Organizational Behavior and Human Decision Processes. 93:114–128.
Menegas W, Babayan BM, Uchida N, Watabe-Uchida M. 2017. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. eLife. 6:e21886.
Menegas W, Bergan JF, Ogawa SK, Isogai Y, Umadevi Venkataraju K, Osten P, Uchida N, Watabe-Uchida M. 2015. Dopamine neurons projecting to the posterior striatum form an anatomically distinct subclass. eLife. 4:e10032.
Metcalfe J, Mischel W. 1999. A hot/cool-system analysis of delay of gratification: Dynamics of willpower. Psychological Review. 106:3–19.
Meule A, Lender A, Richard A, Dinic R, Blechert J. 2019. Approach–avoidance tendencies towards food: Measurement on a touchscreen and the role of attention and food craving. Appetite. 137:145–151.
Middleton FA, Strick PL. 2000. Basal ganglia and cerebellar loops: Motor and cognitive circuits. Brain Research Reviews. 31:236–250.
Mikhael JG, Kim HR, Uchida N, Gershman SJ. 2022. The role of state uncertainty in the dynamics of dopamine. Current Biology. 32:1077-1087.e9.
Miller EM, Shankar MU, Knutson B, McClure SM. 2014. Dissociating motivation from reward in human striatal activity. Journal of Cognitive Neuroscience. 26:1075–1084.
Milli S, Lieder F, Griffiths TL. 2021. A rational reinterpretation of dual-process theories. Cognition. 217:104881.
Millner AJ, Gershman SJ, Nock MK, den Ouden HEM. 2017. Pavlovian control of escape and avoidance. Journal of Cognitive Neuroscience. 26:1–12.

Milstein DM, Dorris MC. 2007. The influence of expected value on saccadic preparation. Journal of Neuroscience. 27:4810–4818.
Mink JW. 1996. The basal ganglia: Focused selection and inhibition of competing motor programs. Progress in Neurobiology. 50:381–425.
Mischel W, Baker N. 1975. Cognitive appraisals and transformations in delay behavior. Journal of Personality and Social Psychology. 31:254–261.
Mischel W, Ebbesen EB. 1970. Attention in delay of gratification. Journal of Personality and Social Psychology. 16:329–337.
Mischel W, Moore B. 1973. Effects of attention to symbolically presented rewards on self-control. Journal of Personality and Social Psychology. 28:172–179.
Mkrtchian A, Aylward J, Dayan P, Roiser JP, Robinson OJ. 2017. Modeling avoidance in mood and anxiety disorders using reinforcement learning. Biological Psychiatry. 82:532–539.
Mkrtchian A, Roiser JP, Robinson OJ. 2017. Threat of shock and aversive inhibition: Induced anxiety modulates Pavlovian-instrumental interactions. Journal of Experimental Psychology: General. 146:1694–1704.
Mohebi A, Pettibone JR, Hamid AA, Wong J-MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD. 2019. Dissociable dopamine dynamics for learning and motivation. Nature. 570:65–70.
Mollick JA, Hazy TE, Krueger KA, Nair A, Mackie P, Herd SA, O'Reilly RC. 2020. A systems-neuroscience model of phasic dopamine. Psychological Review. 127:972–1021.
Monosov IE, Rushworth MFS. 2022. Interactions between ventrolateral prefrontal and anterior cingulate cortex during learning and behavioural change. Neuropsychopharmacol. 47:196–210.
Moran R, Keramati M, Dayan P, Dolan RJ. 2019. Retrospective model-based inference guides model-free credit assignment. Nature Communications. 10:750.
Morey RD. 2008. Confidence intervals from normalized data: A correction to Cousineau (2005). Tutorials in Quantitative Methods for Psychology. 4:61–64.
Morse WH, Skinner BF. 1957. A second type of superstition in the pigeon. The American Journal of Psychology. 70:308–311.
Moutoussis M, Bullmore ET, Goodyer IM, Fonagy P, Jones PB, Dolan RJ, Dayan P. 2018. Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. PLOS Computational Biology. 14:e1006679.
Moutoussis M, Rutledge RB, Prabhu G, Hrynkiewicz L, Lam J, Ousdal O-T, Guitart-Masip M, Fonagy P, Dolan RJ. 2018. Neural activity and fundamental learning, motivated by monetary loss and reward, are intact in mild to moderate major depressive disorder. PLOS ONE. 13:e0201451.
Mowrer OH. 1947. On the dual nature of learning--A re-interpretation of "conditioning" and "problem -solving." Harvard Educational Review. 17:102–148.
Müller T, Husain M, Apps MAJ. 2022. Preferences for seeking effort or reward information bias the willingness to work. Sci Rep. 12:19486.
Murayama K, Usami S, Sakaki M. 2022. Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. Psychological Methods. 27:1014–1038.
Murphy PR, Robertson IH, Harty S, O'Connell RG. 2015. Neural evidence accumulation persists after choice to inform metacognitive judgments. eLife. 4:e11946.
Musall S, Kaufman MT, Juavinett AL, Gluf S, Churchland AK. 2019. Single-trial neural dynamics are dominated by richly varied movements. Nature Neuroscience. 22:1677–1686.
Musall S, Urai AE, Sussillo D, Churchland AK. 2019. Harnessing behavioral diversity to understand neural computations for cognition. Current Opinion in Neurobiology. 58:229–238.

Nassar MR, Frank MJ. 2016. Taming the beast: Extracting generalizable knowledge from computational models of cognition. Current Opinion in Behavioral Sciences. 11:49–54.
Nesse RM. 2001. The smoke detector principle: Natural selection and the regulation of defensive responses. Annals of the New York Academy of Sciences. 935:75–85.
Neuper C, Wörtz M, Pfurtscheller G. 2006. ERD/ERS patterns reflecting sensorimotor activation and deactivation. In: Neuper C,, Klimesch W, editors. Progress in Brain Research. Event-Related Dynamics of Brain Oscillations. Elsevier. p. 211–222.
Niv Y, Daw ND, Joel D, Dayan P. 2007. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology. 191:507–520.
Niv Y, Edlund JA, Dayan P, O'Doherty JP. 2012. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. Journal of Neuroscience. 32:551–562.
Noonan MP, Chau BKH, Rushworth MFS, Fellows LK. 2017. Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. Journal of Neuroscience. 37:7023–7035.
Noonan MP, Crittenden BM, Jensen O, Stokes MG. 2018. Selective inhibition of distracting input. Behavioural Brain Research. 355:36–47.
Notebaert W, Houtman F, Opstal FV, Gevers W, Fias W, Verguts T. 2009. Post-error slowing: An orienting account. Cognition. 111:275–279.
O'Connell RG, Dockree PM, Kelly SP. 2012. A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. Nature Neuroscience. 15:1729–1735.
O'Doherty JP, Cockburn J, Pauli WM. 2017. Learning, reward, and decision making. Annual Review of Psychology. 68:73–100.
O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. 2003. Temporal difference models and reward-related learning in the human brain. Neuron. 38:329–337.
O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ. 2002. Neural responses during anticipation of a primary taste reward. Neuron. 33:815–826.
Olivers CNL, Roelfsema PR. 2020. Attention for action in visual working memory. Cortex. 131:179–194.
Oostenveld R, Fries P, Maris E, Schoffelen J-M. 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience. 2011:1–9.
O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS. 2013. Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. Proceedings of the National Academy of Sciences. 110:E3660–E3669.
O'Reilly JX, Woolrich MW, Behrens TEJ, Smith SM, Johansen-Berg H. 2012. Tools of the trade: Psychophysiological interactions and functional connectivity. Social Cognitive and Affective Neuroscience. 7:604–609.
Ostlund SB, Maidment NT. 2012. Dopamine receptor blockade attenuates the general incentive motivational effects of noncontingently delivered rewards and reward-paired cues without affecting their ability to bias action selection. Neuropsychopharmacology. 37:508–519.
Otto AR, Daw ND. 2019. The opportunity cost of time modulates cognitive effort. Neuropsychologia, Cognitive Effort. 123:92–105.
Ousdal OT, Huys QJ, Milde AM, Craven AR, Ersland L, Endestad T, Melinder A, Hugdahl K, Dolan RJ. 2018. The impact of traumatic stress on Pavlovian biases. Psychological Medicine. 48:327–336.
Pachur T, Schulte-Mecklenbeck M, Murphy RO, Hertwig R. 2018. Prospect theory reflects selective allocation of attention. Journal of Experimental Psychology: General. 147:147–169.

Padmala S, Bauer A, Pessoa L. 2011. Negative emotion impairs conflict-driven executive control. Frontiers in Psychology. 2:1–5.
Palminteri S, Wyart V, Koechlin E. 2017. The importance of falsification in computational cognitive modeling. Trends in Cognitive Sciences. 21:425–433.
Pärnamets P, Johansson P, Hall L, Balkenius C, Spivey MJ, Richardson DC. 2015. Biasing moral decisions by exploiting the dynamics of eye gaze. Proceedings of the National Academy of Sciences. 112:4170–4175.
Parpart P, Jones M, Love BC. 2018. Heuristics as Bayesian inference under extreme priors. Cognitive Psychology. 102:127–144.
Pasupathy A, Miller EK. 2005. Different time courses of learning-related activity in the prefrontal cortex and striatum. Nature. 433:873–876.
Paul K, Vassena E, Severo MC, Pourtois G. 2020. Dissociable effects of reward magnitude on fronto-medial theta and FRN during performance monitoring. Psychophysiology. 57:e13481.
Peciña S, Berridge KC. 2013. Dopamine or opioid stimulation of nucleus accumbens similarly amplify cue-triggered 'wanting' for reward: entire core and medial shell mapped as substrates for PIT enhancement. European Journal of Neuroscience. 37:1529–1540.
Peelen MV, Kastner S. 2014. Attention in the real world: toward understanding its neural basis. Trends in Cognitive Sciences. 18:242–250.
Pessiglione M, Delgado MR. 2015. The good, the bad and the brain: Neural correlates of appetitive and aversive values underlying decision making. Current Opinion in Behavioral Sciences. 5:78–84.
Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 442:1042–1045.
Pessoa L, Padmala S, Kenzer A, Bauer A. 2012. Interactions between cognition and emotion during response inhibition. Emotion. 12:192–197.
Peters AJ, Fabre JMJ, Steinmetz NA, Harris KD, Carandini M. 2021. Striatal activity topographically reflects cortical activity. Nature. 591:420–425.
Pfurtscheller G, Graimann B, Huggins JE, Levine SP, Schuh LA. 2003. Spatiotemporal patterns of beta desynchronization and gamma synchronization in corticographic data during self-paced movement. Clinical Neurophysiology. 114:1226–1236.
Philiastides MG, Heekeren HR, Sajda P. 2014. Human scalp potentials reflect a mixture of decision-related signals during perceptual choices. Journal of Neuroscience. 34:16877–16889.
Phillips PEM, Stuber GD, Heien MLAV, Mark Wightman R, Carelli RM, Wightman RM, Carelli RM. 2003. Subsecond dopamine release promotes cocaine seeking. Nature. 422:614–618.
Pinel JPJ, Treit D. 1979. Conditioned defensive burying in rats: Availability of burying materials. Animal Learning & Behavior. 7:392–396.
Piray P, Dezfouli A, Heskes T, Frank MJ, Daw ND. 2019. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. PLOS Computational Biology. 15:e1007043.
Piray P, Toni I, Cools R. 2016. Human choice strategy varies with anatomical projections from ventromedial prefrontal cortex to medial striatum. Journal of Neuroscience. 36:2857–2867.
Pizzagalli DA. 2011. Frontocingulate dysfunction in depression: Toward biomarkers of treatment response. Neuropsychopharmacology. 36:183–206.
Pizzo F, Roehri N, Medina Villalon S, Trébuchon A, Chen S, Lagarde S, Carron R, Gavaret M, Giusiano B, McGonigal A, Bartolomei F, Badier JM, Bénar CG. 2019. Deep brain activities can be detected with magnetoencephalography. Nature Communications. 10:971.

Polanía R, Krajbich I, Grueschow M, Ruff CC. 2014. Neural oscillations and synchronization differentially support evidence accumulation in perceptual and value-based decision making. Neuron. 82:709–720.
Pool ER, Tord DM, Delplanque S, Stussi Y, Cereghetti D, Vuilleumier P, Sander D. 2022. Differential contributions of ventral striatum subregions to the motivational and hedonic components of the affective processing of reward. J Neurosci. 42:2716–2728.
Proudfit GH. 2015. The reward positivity: From basic research on reward to a biomarker for depression. Psychophysiology. 52:449–459.
Pruim RHR, Mennes M, van Rooij D, Llera A, Buitelaar JK, Beckmann CF. 2015. ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. NeuroImage. 112:267–277.
Rabbitt P, Rodgers B. 1977. What does a man do after he makes an error? An analysis of response programming. Quarterly Journal of Experimental Psychology. 29:727–743.
Ratcliff R. 2006. Modeling response signal and response time data. Cognitive Psychology. 53:195–237.
Reeck C, Wall D, Johnson EJ. 2017. Search predicts and changes patience in intertemporal choice. Proceedings of the National Academy of Sciences. 114:11890–11895.
Reppert TR, Lempert KM, Glimcher PW, Shadmehr R. 2015. Modulation of saccade vigor during value-based decision making. Journal of Neuroscience. 35:15369–15378.
Rescorla RA. 1969. Conditioned inhibition of fear resulting from negative CS-US contingencies. Journal of Comparative and Physiological Psychology. 67:504–509.
Rescorla RA. 1988. Pavlovian conditioning: It's not what you think it is. American Psychologist. 43:151–160.
Rescorla RA, Soloman RL. 1967. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychological Review. 74:151–182.
Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH,, Prokasy WF, editors. Classical Conditioning II: Current Research and Theory. New York, NY: Appleton Century Crofts. p. 64–99.
Rihs TA, Michel CM, Thut G. 2007. Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. European Journal of Neuroscience. 25:603–610.
Risko EF, Gilbert SJ. 2016. Cognitive offloading. Trends in Cognitive Sciences. 20:676–688.
Ritov I, Baron J. 1990. Reluctance to vaccinate: Omission bias and ambiguity. Journal of Behavioral Decision Making. 3:263–277.
Ritov I, Baron J. 1995. Outcome knowledge, regret, and omission bias. Organizational Behavior and Human Decision Processes. 64:119–127.
Ritter P, Moosmann M, Villringer A. 2009. Rolandic alpha and beta EEG rhythms' strengths are inversely related to fMRI-BOLD signal in primary somatosensory and motor cortex. Human Brain Mapping. 30:1168–1187.
Rizzolatti G, Riggio L, Dascola I, Umiltá C. 1987. Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. Neuropsychologia. 25:31–40.
Robbins TW, Everitt BJ. 1992. Functions of dopamine in the dorsal and ventral striatum. Seminars in Neuroscience. 4:119–127.
Robbins TW, Everitt BJ. 2007. A role for mesencephalic dopamine in activation: Commentary on Berridge (2006). Psychopharmacology. 191:433–437.
Robinson MJF, Berridge KC. 2013. Instant transformation of learned repulsion into motivational "wanting." Current Biology. 23:282–289.
Robinson TE, Berridge KC. 1993. The neural basis of drug craving: An incentive-sensitization theory of addiction. Brain Research Reviews. 18:247–291.

Robinson TE, Yager LM, Cogan ES, Saunders BT. 2014. On the motivational properties of reward cues: Individual differences. Neuropharmacology. 76:450–459.
Roelofs K. 2017. Freeze for action: Neurobiological mechanisms in animal and human freezing. Philosophical Transactions of the Royal Society B: Biological Sciences. 372:20160206.
Roitman MF, Stuber GD, Phillips PEM, Wightman RM, Carelli RM. 2004. Dopamine operates as a subsecond modulator of food seeking. Journal of Neuroscience. 24:1265–1271.
Roitman MF, Wheeler RA, Carelli RM. 2005. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. Neuron. 45:587–597.
Roy NA, Bak JH, Akrami A, Brody CD, Pillow JW. 2021. Extracting the dynamics of behavior in sensory decision-making experiments. Neuron. 109:597-610.e6.
Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. 2009. Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. Journal of Neuroscience. 29:15104–15114.
Sackaloo K, Strouse E, Rice MS. 2015. Degree of preference and its influence on motor control when reaching for most preferred, neutrally preferred, and least preferred candy. OTJR: Occupation, Participation and Health. 35:81–88.
Sadaghiani S, Scheeringa R, Lehongre K, Morillon B, Giraud A-L, Kleinschmidt A. 2010. Intrinsic connectivity networks, alpha oscillations, and tonic alertness: A simultaneous electroencephalography/functional magnetic resonance imaging study. Journal of Neuroscience. 30:10243–10250.
Salamone JD, Correa M. 2012. The mysterious motivational functions of mesolimbic dopamine. Neuron. 76:470–485.
Salmelin R, Forss N, Knuutila J, Hari R. 1995. Bilateral activation of the human somatomotor cortex by distal hand movements. Electroencephalography and Clinical Neurophysiology. 95:444–452.
Salmelin R, Hämäläinen M, Kajola M, Hari R. 1995. Functional segregation of movement-related rhythmic activity in the human brain. NeuroImage. 2:237–243.
Salmelin R, Hari R. 1994. Spatiotemporal characteristics of rhythmic neuromagnetic activity related to thumb movement. Neuroscience. 60:537–550.
Sambrook TD, Goslin J. 2016. Principal components analysis of reward prediction errors in a reinforcement learning task. NeuroImage. 124:276–286.
Sanes JN, Donoghue JP. 1993. Oscillations in local field potentials of the primate motor cortex during voluntary movement. Proceedings of the National Academy of Sciences. 90:4470–4474.
Sassenhagen J, Draschkow D. 2019. Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology. 56:e13335.
Sato A, Yasuda A, Ohira H, Miyawaki K, Nishikawa M, Kumano H, Kuboki T. 2005. Effects of value and reward magnitude on feedback negativity and P300. NeuroReport. 16:407–411.
Saunders BT, Robinson TE. 2013. Individual variation in resisting temptation: Implications for addiction. Neuroscience and Biobehavioral Reviews. 37:1955–1975.
Schad DJ, Rapp MA, Garbusow M, Nebe S, Sebold M, Obst E, Sommer C, Deserno L, Rabovsky M, Friedel E, Romanczuk-Seiferth N, Wittchen H-U, Zimmermann US, Walter H, Sterzer P, Smolka MN, Schlagenhauf F, Heinz A, Dayan P, Huys QJM. 2020. Dissociating neural learning signals in human sign- and goal-trackers. Nat Hum Behav. 4:201–214.
Scheeringa R, Bastiaansen MCM, Petersson KM, Oostenveld R, Norris DG, Hagoort P. 2008. Frontal theta EEG activity correlates negatively with the default mode network in resting state. International Journal of Psychophysiology. 67:242–251.

Scheeringa R, Fries P, Petersson K-M, Oostenveld R, Grothe I, Norris DG, Hagoort P, Bastiaansen MCM. 2011. Neuronal dynamics underlying high-and low-frequency EEG oscillations contribute independently to the human BOLD signal. Neuron. 69:572–583.

Scheeringa R, Petersson KM, Oostenveld R, Norris DG, Hagoort P, Bastiaansen MCM. 2009. Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. NeuroImage. 44:1224–1238.

Schneider W, Shiffrin RM. 1977. Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review. 84:1–66.

Schoenbaum G, Chiba AA, Gallagher M. 1998. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nature Neuroscience. 1:155–159.

Schonberg T, Bakkour A, Hover AM, Mumford JA, Nagar L, Perez J, Poldrack RA. 2014. Changing value through cued approach: An automatic mechanism of behavior change. Nature Neuroscience. 17:625–630.

Schuck NW, Cai MB, Wilson RC, Niv Y. 2016. Human orbitofrontal cortex represents a cognitive map of state space. Neuron. 91:1402–1412.

Schultz W. 2016. Dopamine reward prediction-error signalling: A two-component response. Nature Reviews Neuroscience. 17:183–195.

Schultz W. 2019. Recent advances in understanding the role of phasic dopamine activity. F1000Research. 8:1680.

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. Science. 275:1593–1599.

Schwartz B. 1976. Positive and negative conditioned suppression in the pigeon: Effects of the locus and modality of the CS. Learning and Motivation. 7:86–100.

Sebold M, Garbusow M, Cerci D, Chen K, Sommer C, Huys QJ, Nebe S, Rapp M, Veer IM, Zimmermann US, Smolka MN, Walter H, Heinz A, Friedel E. 2021. Association of the OPRM1 A118G polymorphism and Pavlovian-to-instrumental transfer: Clinical relevance for alcohol dependence. J Psychopharmacol. 35:566–578.

Sekutowicz M, Guggenmos M, Kuitunen-Paul S, Garbusow M, Sebold M, Pelz P, Priller J, Wittchen H-U, Smolka MN, Zimmermann US, Heinz A, Sterzer P, Schmack K. 2019. Neural response patterns during Pavlovian-to-Instrumental Transfer predict alcohol relapse and young adult drinking. Biological Psychiatry. 86:857–863.

Seo M, Lee E, Averbeck BB. 2012. Action selection and action value in frontal-striatal circuits. Neuron. 74:947–960.

Sepulveda P, Usher M, Davies N, Benson AA, Ortoleva P, De Martino B. 2020. Visual attention modulates the integration of goal-relevant evidence and not value. eLife. 9:e60705.

Shackman AJ, Salomons TV, Slagter HA, Fox AS, Winter JJ, Davidson RJ. 2011. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nature Reviews Neuroscience. 12:154–167.

Shadlen MN, Shohamy D. 2016. Decision making and sequential sampling from memory. Neuron. 90:927–939.

Shadmehr R, Reppert TR, Summerside EM, Yoon T, Ahmed AA. 2019. Movement vigor as a reflection of subjective economic utility. Trends in Neurosciences. 42:323–336.

Shadmehr R, Smith MA, Krakauer JW. 2010. Error correction, sensory prediction, and adaptation in motor control. Annual Review of Neuroscience. 33:89–108.

Shahar N, Hauser TU, Moutoussis M, Moran R, Keramati M, Dolan RJ. 2019. Improving the reliability of model-based decision-making estimates in the two-stage decision task with reaction-times and drift-diffusion modeling. PLOS Computational Biology. 15:e1006803.

Sharpe MJ, Chang CY, Liu MA, Batchelor HM, Mueller LE, Jones JL, Niv Y, Schoenbaum G. 2017. Dopamine transients are sufficient and necessary for acquisition of model-based associations. Nature Neuroscience. 20:735–742.

Sharpe MJ, Stalnaker T, Schuck NW, Killcross S, Schoenbaum G, Niv Y. 2019. An integrated model of action selection: Distinct modes of cortical control of striatal decision making. Annual Review of Psychology. 70:53–76.
Sheliga BM, Craighero L, Riggio L, Rizzolatti G. 1997. Effects of spatial attention on directional manual and ocular responses. Experimental Brain Research. 114:339–351.
Shenhav A, Botvinick MM, Cohen JD. 2013. The expected value of control: An integrative theory of anterior cingulate cortex function. Neuron. 79:217–240.
Shenhav A, Cohen JD, Botvinick MM. 2016. Dorsal anterior cingulate cortex and the value of control. Nature Neuroscience. 19:1286–1291.
Shevlin BRK, Smith SM, Hausfeld J, Krajbich I. 2022. High-value decisions are fast and accurate, inconsistent with diminishing value sensitivity. Proceedings of the National Academy of Sciences. 119:e2101508119.
Shiffrin RM, Schneider W. 1977. Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review. 84:127–190.
Singmann H, Bolker B, Westfall J, Aust F. 2018. afex: Analysis of factorial experiments.
Skinner BF. 1938. The behavior of organisms. New York, NY: Appleton-Century-Crofts.
Smith SM. 2002. Fast robust automated brain extraction. Human Brain Mapping. 17:143–155.
Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM. 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage, Mathematics in Brain Imaging. 23:S208–S219.
Sommer C, Birkenstock J, Garbusow M, Obst E, Schad DJ, Bernhardt N, Huys QM, Wurst FM, Weinmann W, Heinz A, Smolka MN, Zimmermann US. 2020. Dysfunctional approach behavior triggered by alcohol-unrelated Pavlovian cues predicts long-term relapse in alcohol dependence. Addiction Biology. 25:1–10.
Sommer C, Garbusow M, Jünger E, Pooseh S, Bernhardt N, Birkenstock J, Schad DJ, Jabs B, Glöckler T, Huys QM, Heinz A, Smolka MN, Zimmermann US. 2017. Strong seduction: impulsivity and the impact of contextual cues on instrumental behavior in alcohol dependence. Translational Psychiatry. 7:e1183–e1183.
Squire RF, Noudoost B, Schafer RJ, Moore T. 2013. Prefrontal contributions to visual selective attention. Annual Review of Neuroscience. 36:451–466.
Stalnaker TA, Berg B, Aujla N, Schoenbaum G. 2016. Cholinergic interneurons use orbitofrontal input to track beliefs about current state. Journal of Neuroscience. 36:6242–6257.
Starkweather CK, Babayan BM, Uchida N, Gershman SJ. 2017. Dopamine reward prediction errors reflect hidden-state inference across time. Nature Neuroscience. 20:581–589.
Steingroever H, Wetzels R, Wagenmakers E-J. 2014. Absolute performance of reinforcement-learning models for the Iowa Gambling Task. Decision. 1:161–183.
Steinmetz NA, Zatka-Haas P, Carandini M, Harris KD. 2019. Distributed coding of choice, action and engagement across the mouse brain. Nature. 576:266–273.
Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009. Bayesian model selection for group studies. NeuroImage. 46:1004–1017.
Stolk A, Brinkman L, Vansteensel MJ, Aarnoutse E, Leijten FS, Dijkerman CH, Knight RT, de Lange FP, Toni I. 2019. Electrocorticographic dissociation of alpha and beta rhythmic activity in the human sensorimotor system. eLife. 8:e48065.
Stolk A, Todorovic A, Schoffelen J-M, Oostenveld R. 2013. Online and offline tools for head movement compensation in MEG. NeuroImage. 68:39–48.
Strack F, Deutsch R. 2004. Reflective and impulsive determinants of social behavior. Personality and Social Psychology Review. 8:220–247.

Stringer C, Pachitariu M, Steinmetz N, Reddy CB, Carandini M, Harris KD. 2019. Spontaneous behaviors drive multidimensional, brainwide activity. Science. 364:eaav7893.
Strube A, Rose M, Fazeli S, Büchel C. 2021. The temporal and spectral characteristics of expectations and prediction errors in pain and thermoception. eLife. 10:e62809.
Stuber GD, Sparta DR, Stamatakis AM, van Leeuwen WA, Hardjoprajitno JE, Cho S, Tye KM, Kempadoo KA, Zhang F, Deisseroth K, Bonci A. 2011. Excitatory transmission from the amygdala to nucleus accumbens facilitates reward seeking. Nature. 475:377–380.
Stujenske JM, Likhtik E, Topiwala MA, Gordon JA. 2014. Fear and safety engage competing patterns of theta-gamma coupling in the basolateral amygdala. Neuron. 83:919–933.
Stussi Y, Delplanque S, Coraj S, Pourtois G, Sander D. 2018. Measuring Pavlovian appetitive conditioning in humans with the postauricular reflex. Psychophysiology. 55:e13073.
Summerside EM, Shadmehr R, Ahmed AA. 2018. Vigor of reaching movements: reward discounts the cost of effort. Journal of Neurophysiology. 119:2347–2357.
Swart JC, Frank MJ, Määttä JI, Jensen O, Cools R, den Ouden HEM. 2018. Frontal network dynamics reflect neurocomputational mechanisms for reducing maladaptive biases in motivated action. PLOS Biology. 16:e2005979.
Swart JC, Froböse MI, Cook JL, Geurts DE, Frank MJ, Cools R, den Ouden HE. 2017. Catecholaminergic challenge uncovers distinct Pavlovian and instrumental mechanisms of motivated (in)action. eLife. 6:e22169.
Syed ECJ, Grima LL, Magill PJ, Bogacz R, Brown P, Walton ME. 2016. Action initiation shapes mesolimbic dopamine encoding of future rewards. Nature Neuroscience. 19:34–36.
Takagi Y, Hunt LT, Woolrich MW, Behrens TE, Klein-Flügge MC. 2021. Adapting non-invasive human recordings along multiple task-axes shows unfolding of spontaneous and over-trained choice. eLife. 10:e60988.
Talmi D, Atkinson R, El-Deredy W. 2013. The feedback-related negativity signals salience prediction errors, not reward prediction errors. Journal of Neuroscience. 33:8264–8269.
Talmi D, Seymour B, Dayan P, Dolan RJ. 2008. Human pavlovian-instrumental transfer. Journal of Neuroscience. 28:360–368.
Tangney JP, Baumeister RF, Boone AL. 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality. 72:271–324.
Tanner D, Morgan-Short K, Luck SJ. 2015. How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. Psychophysiology. 52:997–1009.
Taylor JR, Robbins TW. 1984. Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. Psychopharmacology. 84:405–412.
Taylor JR, Robbins TW. 1986. 6-Hydroxydopamine lesions of the nucleus accumbens, but not of the caudate nucleus, attenuate enhanced responding with reward-related stimuli produced by intra-accumbens d-amphetamine. Psychopharmacology. 90:390–397.
Thompson TI. 1963. Visual reinforcement in Siamese fighting fish. Science. 141:55–57.
Thompson TI. 1964. Visual reinforcement in fighting cocks. Journal of the Experimental Analysis of Behavior. 7:45–49.
Thut G, Nietzel A, Brandt SA, Pascual-Leone A. 2006. Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. Journal of Neuroscience. 26:9494–9502.
Tobler PN, Dickinson A, Schultz W. 2003. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. J Neurosci. 23:10402–10410.
Tobler PN, O'Doherty JP, Dolan RJ, Schultz W. 2007. Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. Journal of Neurophysiology. 97:1621–1632.

Töllner T, Wang Y, Makeig S, Müller HJ, Jung T-P, Gramann K. 2017. Two independent frontal midline theta oscillations during conflict detection and adaptation in a Simon-type manual reaching task. J Neurosci. 37:2504–2515.
Trudel N, Scholl J, Klein-Flügge MC, Fouragnan E, Tankelevitch L, Wittmann MK, Rushworth MFS. 2021. Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. Nat Hum Behav. 5:83–98.
Twomey DM, Murphy PR, Kelly SP, O'Connell RG. 2015. The classic P300 encodes a build-to-threshold decision variable. European Journal of Neuroscience. 42:1636–1643.
van de Vijver I, Ridderinkhof KR, Cohen MX. 2011. Frontal oscillatory dynamics predict feedback learning and action adjustment. Journal of Cognitive Neuroscience. 23:4106–4121.
van der Meij R, van Ede F, Maris E. 2016. Two independent frontal midline theta oscillations during conflict detection and adaptation in a Simon-type manual reaching task. PLOS ONE. 11:e0154881.
Van der Stigchel S, Hollingworth A. 2018. Visuospatial working memory as a fundamental component of the eye movement system. Current Directions in Psychological Science. 27:136–143.
van Ede F. 2020. Visual working memory and action: Functional links and bi-directional influences. Visual Cognition. 28:401–413.
van Ede F, Chekroud SR, Nobre AC. 2019. Human gaze tracks attentional focusing in memorized visual space. Nature Human Behaviour. 27–29.
van Ede F, Chekroud SR, Stokes MG, Nobre AC. 2019. Concurrent visual and motor selection during visual working memory guided action. Nat Neurosci. 22:477–483.
Van Ede F, Deden J, Nobre AC. 2021. Looking ahead in working memory to guide sequential behaviour. Current Biology. 31:R779–R780.
van Kerkoerle T, Self MW, Dagnino B, Gariel-Mathis M-A, Poort J, van der Togt C, Roelfsema PR. 2014. Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. Proceedings of the National Academy of Sciences. 111:14332–14341.
van Moorselaar D, Slagter HA. 2020. Inhibition in selective attention. Annals of the New York Academy of Sciences. 1464:204–221.
van Nuland AJ, Helmich RC, Dirkx MF, Zach H, Toni I, Cools R, den Ouden HEM. 2020. Effects of dopamine on reinforcement learning in Parkinson's disease depend on motor phenotype. Brain. 143:3422–3434.
van Schouwenburg MR, O'Shea J, Mars RB, Rushworth MFS, Cools R. 2012. Controlling human striatal cognitive function via the frontal cortex. Journal of Neuroscience. 32:5631–5637.
van Steenbergen H, Band GPH, Hommel B. 2009. Reward counteracts conflict adaptation: Evidence for a role of affect in executive control. Psychological Science. 20:1473–1477.
van Steenbergen H, Band GPH, Hommel B. 2012. Reward valence modulates conflict-driven attentional adaptation: Electrophysiological evidence. Biological Psychology. 90:234–241.
van Steenbergen H, Watson P, Wiers RW, Hommel B, de Wit S. 2017. Dissociable corticostriatal circuits underlie goal-directed vs. cue-elicited habitual food seeking after satiation: Evidence from a multimodal MRI study. European Journal of Neuroscience. 46:1815–1827.
van Vugt MK, Simen P, Nystrom LE, Holmes P, Cohen JD. 2012. EEG oscillations reveal neural correlates of evidence accumulation. Frontiers in Neuroscience. 6:1–13.
Van Wingerden M, Vinck M, Lankelma J, Pennartz CMA. 2010. Theta-band phase locking of orbitofrontal neurons during reward expectancy. Journal of Neuroscience. 30:7078–7087.
Vasconcelos M, Monteiro T, Kacelnik A. 2015. Irrational choice and the value of information. Sci Rep. 5:13874.

Vassena E, Deraeve J, Alexander WH. 2019. Task-specific prioritization of reward and effort information: Novel insights from behavior and computational modeling. Cogn Affect Behav Neurosci. 19:619–636.
Veling H, Chen Z, Tombrock MC, Verpaalen IAM, Schmitz LI, Dijksterhuis A, Holland RW. 2017. Training impulsive choices for healthy and sustainable food. Journal of Experimental Psychology: Applied. 23:204–215.
Verbruggen F, De Houwer J. 2007. Do emotional stimuli interfere with response inhibition? Evidence from the stop signal paradigm. Cognition & Emotion. 21:391–403.
Vinck M, van Wingerden M, Womelsdorf T, Fries P, Pennartz CMA. 2010. The pairwise phase consistency: A bias-free measure of rhythmic neuronal synchronization. NeuroImage. 51:112–122.
Vyas S, O'Shea DJ, Ryu SI, Shenoy KV. 2020. Causal role of motor preparation during error-driven learning. Neuron. 106:329-339.e4.
Walsh E, Kühn S, Brass M, Wenke D, Haggard P. 2010. EEG activations during intentional inhibition of voluntary action: An electrophysiological correlate of self-control? Neuropsychologia. 48:619–626.
Walton ME, Behrens TEJ, Buckley MJ, Rudebeck PH, Rushworth MFS. 2010. Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning. Neuron. 65:927–939.
Walton ME, Bouret S. 2018. What Is the relationship between dopamine and effort? Trends in Neurosciences. 42:1–13.
Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Soyer H, Leibo JZ, Hassabis D, Botvinick M. 2018. Prefrontal cortex as a meta-reinforcement learning system. Nat Neurosci. 21:860–868.
Wassum KM, Ostlund SB, Balleine BW, Maidment NT. 2011. Differential dependence of Pavlovian incentive motivation and instrumental incentive learning processes on dopamine signaling. Learning and Memory. 18:475–483.
Wassum KM, Ostlund SB, Maidment NT. 2012. Phasic mesolimbic dopamine signaling precedes and predicts performance of a self-initiated action sequence task. Biological Psychiatry. 71:846–854.
Watson P, Pearson D, Theeuwes J, Most SB, Le Pelley ME. 2020. Delayed disengagement of attention from distractors signalling reward. Cognition. 195:104125.
Watson P, Wiers RW, Hommel B, de Wit S. 2014. Working for food you don't desire. Cues interfere with goal-directed food-seeking. Appetite. 79:139–148.
Wegner DM, Sparrow B, Winerman L. 2004. Vicarious agency: Experiencing control over the movements of others. Journal of Personality and Social Psychology. 86:838–848.
Weilbächer RA, Krajbich I, Rieskamp J, Gluth S. 2021. The influence of visual attention on memory-based preferential choice. Cognition. 215:104804.
Welsh TN, Pratt J. 2008. Actions modulate attentional capture. Quarterly Journal of Experimental Psychology. 61:968–976.
Wessel JR. 2017. Perceptual surprise aides inhibitory motor control. Journal of Experimental Psychology: Human Perception and Performance. 43:1585–1593.
Wessel JR. 2018a. Prepotent motor activity and inhibitory control demands in different variants of the go/no-go paradigm. Psychophysiology. 55:e12871.
Wessel JR. 2018b. A neural mechanism for surprise-related interruptions of visuospatial working memory. Cerebral Cortex. 28:199–212.
Wessel JR. 2018c. An adaptive orienting theory of error processing. Psychophysiology. 55:e13041.
Wessel JR, Aron AR. 2017. On the globality of motor suppression: Unexpected events and their influence on behavior and cognition. Neuron. 93:259–280.

Wessel JR, Ghahremani A, Udupa K, Saha U, Kalia SK, Hodaie M, Lozano AM, Aron AR, Chen R. 2016. Stop-related subthalamic beta activity indexes global motor suppression in Parkinson's disease. Movement Disorders. 31:1846–1853.
Wessel JR, Jenkinson N, Brittain J-S, Voets SHEM, Aziz TZ, Aron AR. 2016. Surprise disrupts cognition via a fronto-basal ganglia suppressive mechanism. Nature Communications. 7:11195.
Wessel JR, Waller DA, Greenlee JD. 2019. Non-selective inhibition of inappropriate motor-tendencies during response-conflict by a fronto-subthalamic mechanism. eLife. 8:e42959.
Westbrook A, Braver TS. 2016. Dopamine does double duty in motivating cognitive effort. Neuron. 91:708.
Westbrook A, Frank MJ, Cools R. 2021. A mosaic of cost–benefit control over cortico-striatal circuitry. Trends in Cognitive Sciences. 25:710–721.
Westbrook A, van den Bosch R, Määttä JI, Hofmans L, Papadopetraki D, Cools R, Frank MJ. 2020. Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. Science. 367:1362–1366.
Wiecki TV, Frank MJ. 2013. A computational model of inhibitory control in frontal cortex and basal ganglia. Psychological Review. 120:329–355.
Williams DR, Williams H. 1969. Auto-maintenance in the pigeon: Sustained pecking despite contingent non-reinforcement. Journal of the Experimental Analysis of Behavior. 12:511–520.
Wilson RC, Niv Y. 2015. Is model fitting necessary for model-based fMRI? PLOS Computational Biology. 11:e1004237.
Wilson RC, Takahashi YK, Schoenbaum G, Niv Y. 2014. Orbitofrontal cortex as a cognitive map of task space. Neuron. 81:267–279.
Wittmann BC, Daw ND, Seymour B, Dolan RJ. 2008. Striatal activity underlies novelty-based choice in humans. Neuron. 58:967–973.
Wittmann MK, Fouragnan E, Folloni D, Klein-Flügge MC, Chau BKH, Khamassi M, Rushworth MFS. 2020. Global reward state affects learning and activity in raphe nucleus and anterior insula in monkeys. Nature Communications. 11:3771.
Wokke ME, Ro T. 2019. Competitive frontoparietal interactions mediate implicit inferences. Journal of Neuroscience. 39:5183–5194.
Wolfe JM. 2021. Guided Search 6.0: An updated model of visual search. Psychonomic Bulletin & Review. 28:1060–1092.
Wolfe JM, Alvarez GA, Horowitz TS. 2000. Attention is fast but volition is slow. Nature. 406:691–691.
Womelsdorf T, Vinck M, Leung S, Everling S. 2010. Selective theta-synchronization of choice-relevant information subserves goal-directed behavior. Frontiers in Human Neuroscience. 4.
Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. 2004. Multilevel linear modelling for FMRI group analysis using Bayesian inference. NeuroImage. 21:1732–1747.
Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM. 2009. Bayesian analysis of neuroimaging data in FSL. NeuroImage. 45:S173–S186.
Worden MS, Foxe JJ, Wang N, Simpson GV. 2000. Anticipatory biasing of visuospatial attention indexed by retinotopically specific alpha-band electroencephalography increases over occipital cortex. J Neurosci. 20:RC63–RC63.
Wu Y, Zhou X. 2009. The P300 and reward valence, magnitude, and expectancy in outcome evaluation. Brain Research. 1286:114–122.

Wykowska A, Schubö A, Hommel B. 2009. How you move is what you see: Action planning biases selection in visual search. Journal of Experimental Psychology: Human Perception and Performance. 35:1755–1769.
Wyvell CL, Berridge KC. 2000. Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: Enhancement of reward "wanting" without enhanced "liking" or response reinforcement. J Neurosci. 20:8122–8130.
Yager LM, Robinson TE. 2013. A classically conditioned cocaine cue acquires greater control over motivated behavior in rats prone to attribute incentive salience to a food cue. Psychopharmacology. 226:217–228.
Yang SC-H, Wolpert DM, Lengyel M. 2016. Theoretical perspectives on active sensing. Current Opinion in Behavioral Sciences. 11:100–108.
Yeung N, Sanfey AG. 2004. Independent coding of reward magnitude and valence in the human brain. Journal of Neuroscience. 24:6258–6264.
Yoo SBM, Hayden BY. 2018. Economic choice as an untangling of options into actions. Neuron. 99:434–447.
Yttri EA, Dudman JT. 2016. Opponent and bidirectional control of movement velocity in the basal ganglia. Nature. 533:402–406.
Zavala BA, Tan H, Little S, Ashkan K, Hariz M, Foltynie T, Zrinzo L, Zaghloul KA, Brown P. 2014. Midline frontal cortex low-frequency activity drives subthalamic nucleus oscillations during conflict. Journal of Neuroscience. 34:7322–7333.
Zeelenberg M, Pligt J van der, de Vries NK. 2000. Attributions of responsibility and affective reactions to decision outcomes. Acta Psychologica. 104:303–315.
Zeelenberg M, van den Bos K, van Dijk E, Pieters R. 2002. The inaction effect in the psychology of regret. Journal of Personality and Social Psychology. 82:314–327.
Zhou J, Gardner MP, Schoenbaum G. 2021. Is the core function of orbitofrontal cortex to signal values or make predictions? Current Opinion in Behavioral Sciences. 41:1–9.
Zumer JM, Scheeringa R, Schoffelen J-M, Norris DG, Jensen O. 2014. Occipital alpha activity during stimulus processing gates the information flow to object-selective cortex. PLoS Biology. 12:e1001965.
Zuure MB, Hinkley LB, Tiesinga PHE, Nagarajan SS, Cohen MX. 2020. Multiple midfrontal thetas revealed by source separation of simultaneous MEG and EEG. J Neurosci. 40:7702–7713.

NEDERLANDSE SAMENVATTING

Wij nemen als mensen beslissingen op verschillende manieren. Soms nemen we veel tijd, denken lang na en proberen we alle factoren mee te wegen om de beste beslissing te nemen. Op andere momenten moeten we echter snel beslissen en kunnen dus niet lang nadenken of we missen het overzicht. We komen dan vaak tot een keuze die misschien niet de beste is maar goed genoeg voor de gegeven situatie. Het wordt aangenomen dat de mens en andere dieren beschikking hebben over verschillende manieren om te komen tot een beslissing, ofwel besluitvormingsprocessen. Deze besluitvormingsprocessen verschillen in hoe snel ze tot een beslissing komen. De tragere systemen gebruiken meer informatie en komen vaak tot een betere beslissing; de snellere gebruiken minder informatie en maken soms fouten. Afhankelijk van de situatie gebruikt de mens een trager of sneller proces.

Een belangrijk simpel maar snel beslissingsproces is het zogenaamde **Pavloviaanse systeem**, dat zich uit door "**Pavloviaanse neigingen**" in gedrag. Het Pavloviaanse systeem is gevoelig voor prikkels die een mogelijke beloning of straf signaleren. Als er een kans is op een beloning, bijvoorbeeld lekker eten, "activeert" dit systeem gedrag, hetgeen toenadering en verzamelen van beloning meestal faciliteert. Als er daarentegen een dreiging van straf is, bijvoorbeeld de aanwezigheid van een roofdier, onderdrukt dit systeem de neiging tot actie, hetgeen ervoor zorgt dat de mens of dier detectie door of blootstelling aan de dreiging vermijdt. Deze neigingen tot activatie en afremmen van gedrag zijn adaptief, dat wil zeggen dat ze in de meeste situaties tot het gewenste resultaat leiden. Er zijn ook situaties waar deze neigingen in strijd zijn met het geëiste gedrag. Soms moeten we bijvoorbeeld wachten en niets doen om een beloning te ontvangen – bijvoorbeeld het cakebeslag in de oven zetten en wachten tot de cake klaar is in plaats van het beslag op te eten. Omgekeerd moeten we soms actief iets doen om een toekomstige negatieve uitkomst te voorkomen. Bij dreigende werkloosheid of faillissement moeten mensen vaak actief iets ondernemen in plaats van ervoor 'wegduiken' om hun situatie te verbeteren. In dergelijke situaties zijn Pavloviaanse neigingen niet gewenst en moeten ze worden onderdrukt.

Pavloviaanse neigingen zijn waargenomen in vele diersoorten, waaronder apen, knaagdieren en vogels. Deze hoge prevalentie suggereert dat Pavloviaanse neigingen van fylogenetische oorsprong zijn, of dat deze op vroege leeftijd consistent worden aangeleerd, ondanks verschillend in leefomgeving. Deze prevalentie suggereert ook dat "evolutionair oude" **hersengebieden**, die opvallende gelijkenissen vertonen in vele diersoorten, betrokken zijn bij deze neigingen. Een belangrijke kandidaat is het *striatum*, een "*subcorticale*" structuur diep in de hersenen. Het striatum is betrokken bij het stimuleren en remmen van gedrag en ontvangt ook informatie over beloningen en straffen. Er is tot nu toe echter geen duidelijk bewijs dat het striatum de "bron" is van deze Pavloviaanse neigingen. Het is ook niet duidelijk of en hoe andere gebieden die beloningen en straffen verwerken, met name de "*corticale*" gebieden van de hersenen, bijdragen aan Pavloviaanse neigingen.

Dieren kunnen Pavloviaanse neigingen meestal niet **remmen** (of zijn hier in ieder geval erg slecht in). Mensen slagen er veel beter in, maar ook mensen maken vaak "fouten" in situaties waarin de Pavloviaanse neigingen ongepast zijn. Het onvermogen om Pavloviaanse neigingen te remmen (of, omgekeerd, het onvermogen om ze soms "hun gang te laten gaan") kan effect hebben voor het mentale welzijn van mensen. Eerder onderzoek toonde aan dat afwijkende Pavloviaanse neigingen kunnen bijdragen aan psychiatrische stoornissen zoals (alcohol)verslaving,

angststoornissen en depressie. Het is daarom belangrijk te begrijpen hoe mensen hun Pavloviaanse neigingen kunnen bijsturen.

De laatste jaren is het de vraag geworden of verschillende beslissingssystemen – zoals de snelle Pavloviaanse controle en de trage, instrumentele controle – strikt gescheiden zijn en met elkaar concurreren, of dat verschillende systemen juist **in synergie kunnen werken**. Hoewel het Pavloviaanse systeem relatief inflexibel is doordat het acties op een starre manier aan bepaalde prikkels koppelt, is het niet alleen snel maar ook robuust. Deze eigenschap maakt het interessant voor gebruik door andere processen om gedrag te stimuleren of te remmen wanneer dat in het belang is van deze systemen. Het Pavloviaanse systeem wordt hiermee een "instrument" dat door andere beslissingsprocessen kan worden aangeroepen.

Het onderzoek in mijn proefschrift had twee hoofddoelen: Als eerste, een beter begrip verkrijgen van de **neurale oorsprong** van Pavloviaanse neigingen, met name de rol van het striatum en andere hersengebieden. Ten tweede, een beter begrip verkrijgen over de **controle** over Pavloviaanse neigingen door te testen wanneer en hoe mensen deze neigingen onderdrukken, en of ze zich strategisch en doelgericht kunnen blootstellen aan stimuli in de omgeving die deze neigingen automatisch oproepen.

In **hoofdstuk 2** heb ik onderzocht welke neurale processen ten grondslag liggen aan het ontstaan van Pavloviaanse neigingen tijdens actieselectie, en hoe ze worden onderdrukt. Pavloviaanse neigingen tijdens actieselectie ontstaan wanneer mensen een prikkel ontvangen die een kans op beloning of een dreiging van straf signaleert, terwijl ze moeten beslissen of ze actie willen ondernemen (op een knop drukken; een "Go"-actie) of gedrag willen onderdrukken (niet drukken; een "NoGo"-actie). Ik verwachtte dat het striatum de "valentie" (verwachting van positieve vs. negatieve uitkomst) van prikkels zou verwerken en daarmee het actieselectieproces zou beïnvloeden. Ook verwachtte ik dat in situaties waarin de neigingen moeten worden onderdrukt, corticale hersengebieden, met name de *anterieure cingulate cortex* (ACC), actiever zouden worden. Deze toename in ACC-activiteit is in eerdere studies waargenomen in de *theta frequentieband* van de elektrische hersengolven. Deze activiteit wordt gedacht te weerspiegelen hoe sterk mensen de neigingen proberen te onderdrukken en hoeveel het valentiesignaal in het striatum wordt verminderd om neigingen te verzwakken. Om zowel de oorsprong als de onderdrukking van biases te bestuderen, combineerde ik functionele magnetische resonantie beeldvorming (fMRI), waarmee activiteit in gebieden diep in de hersenen met een hoge ruimtelijke resolutie kan worden gemeten, en elektro-encefalografie (EEG), waarmee elektrische hersengolven, gemeten met elektroden op de hoofdhuid, met een hoge temporele resolutie worden gemeten.

Een verassende uitkomst was dat niet het striatum de valentie van prikkels signaleerde, maar juist twee corticale gebieden, de *ventromediale prefrontale cortex* (vmPFC) en de *anterieure cingulate cortex* (ACC), dat deden. Het striatum weerspiegelde de uiteindelijke actie (Go of NoGo) die de proefpersoon zou maken. Ik vond geen duidelijk bewijs dat de ACC actiever was wanneer neigingen werden onderdrukt, waarschijnlijk omdat de deelnemers te veel fouten maakten in situaties waarin neigingen moesten worden onderdrukt. Ook bleek de theta frequentieband in de elektrische hersengolven niet gerelateerd aan de onderdrukking van neigingen maar veel meer aan de uiteindelijke actie (Go of NoGo), net zoals het striatum. Dit suggereert dat de theta-band activiteit in het corticale EEG-signaal gebruikt kan worden als een index van striatale actieselectieprocessen. Samenvattend levert deze studie bewijs voor de oorsprong van Pavloviaanse neigingen in corticale gebieden (vmPFC en ACC), die op hun beurt de actieselectieprocessen in het striatum beïnvloeden.

In **hoofdstuk 3** heb ik de neurale processen onderzocht die betrokken zijn bij het leren van Pavloviaanse neigingen. Niet alleen de acties die mensen ondernemen, maar ook hun leervermogen over hoe die acties tot uitkomsten leiden kan vertekend zijn: Mensen vinden het makkelijker om beloningen toe te schrijven aan iets wat ze net gedaan hebben, dan aan iets wat ze net juist niet gedaan hebben. Omgekeerd vinden ze het moeilijker om een straf toe te schrijven aan terughoudendheid dan aan daadkracht. Deze vooringenomen leermechanismen kunnen ook Pavloviaanse neigingen in gedrag verklaren. Hun neurale oorsprong is echter onduidelijk. Gezien het feit dat het striatum zowel betrokken is bij actieselectie als bij het leren van beloningen en straffen, leek het aannemelijk dat het striatum aanleiding zou geven tot Pavloviaanse neigingen tijdens het leren. Eerdere studies lieten echter zien dat andere corticale gebieden, zoals de ACC, mede bepalen hoeveel het striatum leert. In de gecombineerde EEG-fMRI studie al beschreven in hoofdstuk 2 bestudeerde ik deze thematiek.

Met behulp van wiskundige modellen van keuzegedrag vond ik aanwijzingen dat het gedrag van deelnemers het best werd beschreven door een combinatie van Pavloviaanse neigingen zowel tijdens actieselectie als tijdens het leren. De activiteit in verschillende gebieden, waaronder vmPFC, ACC en striatum, werd beter beschreven door "vooringenomen" leersignalen dan door "standaard" leersignalen zonder vooringenomenheid. De leersignalen in verschillende gebieden werden op verschillende tijdstippen gekoppeld aan het EEG-signaal. Uit de resulaten bleek dat het EEG-signaal in corticale gebieden zoals de ACC voorliep op de activiteit in het striatum. Deze resultaten komen overeen met het idee dat Pavloviaanse neigingen tijdens het leren eerst ontstaan in corticale gebieden, die vervolgens latere leerprocessen in het striatum beïnvloeden, vergelijkbaar met de processen die ten grondslag liggen aan Pavloviaanse neigingen tijdens actieselectie.

In **hoofdstuk 4** heb ik onderzocht of mensen het Pavloviaanse systeem rekruteren om doelen te behalen die door andere beslissystemen zijn gesteld. Soms besluit het doelgerichte, instrumentele systeem om gedrag te stimuleren of te remmen omdat het verwacht dat dit gedrag tot een goed resultaat zal leiden. Het instrumentele systeem kan echter afgeleid raken of er anderszins niet in slagen deze doelen te bereiken. Het zou daarom het Pavloviaanse systeem kunnen inroepen als een "automatische piloot" naar het doel. Omdat het Pavloviaanse systeem automatisch wordt geactiveerd door prikkels die een beloning of straf aangeven, zouden mensen hun aandacht op dergelijke prikkels kunnen richten om het Pavloviaanse systeem te activeren, dat dan automatisch de beoogde actie uitvoert. In twee studies gebruikte ik *eye-tracking*, een techniek die meet waar mensen naar kijken, om te onderzoeken of de actieplannen van mensen – het stimuleren of afremmen van gedrag – beïnvloeden of ze meer naar belonings- of naar strafsignalen kijken. Ook testte ik of, omgekeerd, de hoeveelheid aandacht die ze richten op belonings- of strafsignalen beïnvloedt welke actie ze uiteindelijk zullen uitvoeren.

In beide studies ontdekte ik dat als het nodig was om gedrag te stimuleren (om een beloning te krijgen en de straf te vermijden), ze eerder naar informatie over de mogelijke beloning keken, terwijl als het nodig was om gedrag te remmen, ze eerder naar informatie over mogelijke straf keken. Hoe langer ze naar informatie over de beloning (vergeleken met straf) keken, hoe groter de kans dat ze gedrag activeerden (vergeleken met het afremmen). Mensen van wie het gedrag sterker werd beïnvloed door waar ze naar keken, presteerden ook beter bij de taak. Kortom, ik vond bewijs dat mensen het Pavloviaanse systeem strategisch kunnen rekruteren door aandacht te richten op belonings- en straf informatie op een manier die automatisch de gewenste actie implementeert.

In **hoofdstuk 5** onderzocht ik de neurale processen waarmee het Pavloviaanse systeem ingeschakeld wordt door andere beslissingssystemen. Hiervoor gebruikte ik *magneto-encefalografie*

(MEG), een techniek vergelijkbaar met EEG, maar dat veranderingen in magnetische velden als gevolg van elektrische hersenactiviteit kan meten met een hoge temporele resolutie. Ik hield me vooral bezig met twee specifieke frequentiebanden van deze hersengolven: de "*bèta*-band" activiteit afkomstig van midden-boven in het hoofd geeft aan dat mensen een actie voorbereiden, terwijl de "*alpha*-band" activiteit afkomstig achter uit het hoofd aangeeft of mensen meer naar links of rechts kijken. Ik onderzocht de veranderingen de bèta-band gekoppeld waren aan de verandering in de alfa-band, en vice versa.

Ik ontdekte dat de bèta-band was gekoppeld aan welke actie (Go of NoGo) deelnemers al enkele seconden voordat ze de actie uitvoerden aan het voorbereiden waren. Ook werd de bèta-band activiteit beïnvloed door het aantal beloningen of straffen dat deelnemers ontvingen, wat de invloed van Pavloviaanse neigingen op de voorbereiding van acties weerspiegelt. De alpha-band activiteit had geen relatie met belonings- of strafsignalen. Hoewel de deelnemers werden geïnstrueerd om hun ogen op het midden van het scherm te houden, maakten ze nog steeds "niet toegestane" oogbewegingen in de richting van de belonings- en strafsignalen. Deze oogbewegingen waren afhankelijk van de actieplannen van de deelnemers, vergelijkbaar met de resultaten gerapporteerd in hoofdstuk 4. Samenvattend vond ik Pavloviaanse neigingen in het gedrag van deelnemers en in de hersengolven die actievoorbereiding weerspiegelden, maar ik vond geen verschillen in de hersengolven die aandacht voor beloning en strafsignalen representeerden. Er is verder onderzoek nodig om uit te zoeken waarom sommige beslissingssystemen oogbewegingen gebruiken om Pavloviaanse neigingen op te roepen, zonder dat dit zichtbaar wordt in de meer subtiele aandachtsmechanismen in de alpha-band.

Afsluitend, dit proefschrift werpt een nieuw licht op de neurale oorsprong van Pavloviaanse neigingen en de controle erover. Ik heb bewijs gevonden dat Pavloviaanse neigingen zowel tijdens actieselectie als tijdens het leren eerst ontstaan in corticale gebieden (vmPFC en ACC) en vervolgens worden overgedragen naar het striatum, waar ze actieselectie en leren beïnvloeden. Ik heb ook bewijs gevonden dat mensen niet slechts passief onderworpen zijn aan Pavloviaanse neigingen, maar dat ze in staat zijn dit systeem actief aan te sturen in situaties waarin ze actieplannen willen uitvoeren. In deze situaties gebruiken ze oogbewegingen om naar prikkels te kijken die beloningen en straffen signaleren. Deze prikkels activeren Pavloviaanse neigingen die zorgen voor de uitvoering van de actieplannen. Ik heb signalen gevonden in de elektrische en magnetische hersengolven die deze neigingen al seconden voor de uiteindelijke actie weerspiegelen. Ik vond echter geen bewijs dat de voorbereiding van acties de strategische rekrutering van aandachtprocessen beïnvloedt. Samenvattend beschrijft dit proefschrift de neurale processen die ten grondslag liggen aan Pavloviaanse neigingen en hoe deze neigingen als een "instrument" kunnen dienen voor andere besluitvormingssystemen.

ENGLISH SUMMARY

We humans make decisions in different ways. Sometimes, we put a lot of time and thought into making the most optimal decision, i.e., the one that returns the highest rewards. However, at other times, we have to make a decision quickly, without much thought. We might then arrive at a decision that is suboptimal, but still situationally "appropriate" given that we did not have the chance (time, mental resources) to make a more deliberate decision. It seems that humans and other animals possess different decision-making systems, some of them faster, but less optimal, and others slower, yet more optimal. Humans and animals select which system to use depending on what the situations demands.

One particularly simple, but fast decision-making system is so-called **Pavlovian control**, leading to "**Pavlovian biases**" in behavior. This system is sensitive to any cues signaling the chance for reward or a threat of punishment. If there is a chance for reward, e.g. delicious food, this system invigorates any ongoing behavior, which facilitates approaching and collecting the reward. On the contrary, if there is a threat of punishment, e.g. a dangerous predator being present, this system inhibits any ongoing behavior, which likely protects the agent from being detected by or exposed to the threat. These biases are adaptive and lead to appropriate behavior in a majority of situations, given that rewards usually need to be approached quickly (so they do not get collected by a competitor) while potential threats need to be treated with caution. However, in some situations, these biases actually are in conflict with the behavior that is desired. For example, we sometimes have to wait and do nothing to achieve a positive outcome—e.g. putting the cake batter in the oven and waiting for the cake to be done instead of eating the batter right away. Vice versa, sometimes, we have to actively do something to prevent a future negative outcome. In face of threatening unemployment or bankruptcy, people often need to take active means to improve their situation rather than passively wait for their problems to disappear. In such situations, Pavlovian biases are maladaptive and have to be suppressed.

Pavlovian biases can be observed in various species across the animal realm, including apes, rodents, and birds. This high prevalence suggests that these biases are of phylogenetic origin or alternatively be reliably acquired in various environments at an early age. Furthermore, this prevalence suggests that evolutionary ancient **brain regions** that are preserved across various species might give rise to these biases. One prominent candidate is the *striatum*, a so-called "*subcortical*" structure deep in the brain. The striatum is involved in the invigoration and inhibition of behavior and furthermore has access to information about rewards and punishments. However, so far, there has been no clear evidence showing that the striatum causes Pavlovian biases. Furthermore, it has not been clear whether and how other regions that process rewards and punishments, particularly in the more superficial, "*cortical*" part of the brain that seems to be somewhat unique to humans and other great apes, contribute to Pavlovian biases.

While animals are seemingly unable to (or at least very bad at) **inhibiting** these biases, humans usually manage to do so. Still, they make frequent "mistakes" in situations in which these biases are maladaptive. Failure to inhibit these biases (or, vice versa, an exaggerated tendency to constantly inhibit them) could have consequences for people's mental well-being. Previous research has found hints that aberrant Pavlovian biases might be a factor contributing to psychiatric disorders such as (alcohol) addiction, anxiety disorders, and depression. Hence, it has become important to understand how humans can control the expression of Pavlovian biases.

In recent years, it has become questionable whether different decision systems—such as fast Pavlovian and slower instrumental control (the latter flexibly selecting actions based on their expected outcomes)—are strictly segregated and always in competition with each other, or whether different systems can actually **work in synergy**. For example, while the Pavlovian system is relatively inflexible in that it links actions to certain cues in a rigid fashion, it is also fast and robust. This fact makes it interesting for recruitment by other systems. Possibly, other systems could strategically expose the organism to cues that trigger the Pavlovian system and thus invigorate/inhibit behavior when this is in the interest of these other systems. Hence, the Pavlovian system should probably not be understood in isolation, but as a "tool" that can be recruited by other decision systems.

Taken together, this thesis had two main aims of investigation: Firstly, it aimed to elucidate the **neural origin** of Pavlovian biases, testing whether the striatum and/or other regions process the information necessary to drive biases. Secondly, it aimed to understand the **control** over these biases, testing when and how humans suppress these biases and whether they might even strategically expose themselves to environmental cues that trigger these biases.

In **Chapter 2**, I investigated the neural mechanisms of how Pavlovian response biases arise and how they are suppressed. Response biases arise when humans see a cue that signals the chance for rewards or the threat of punishment and have to decide whether to invigorate behavior (press a button; a "Go" action) or to suppress behavior (no button press; a "NoGo" action). I expected that the striatum would process the "valence" of the cue (signaling positive vs. negative outcomes) and then influence the action selection process. Furthermore, I expected that in situations in which the biases had to be suppressed, more superficial, "cortical" regions, in particular the *anterior cingulate cortex* (ACC), would become more active. This increase in activity has previously been observed in the form of increased "*theta*" band power in the electrical brain waves measured over the scalp. This activity should reflect how strongly humans try to suppress the biases, and how much the valence signal in the striatum is reduced in order to weaken the biases. To study both the origin and the suppression of these biases, I combined functional magnetic resonance imaging (fMRI), which allows to measure activity in regions deep in the brain at a high spatial resolution, and electroencephalography (EEG), which allows measuring electrical brain waves over the scalp at a high temporal resolution.

Surprisingly, I did not find the striatum to signal the valence of cues, but instead found two more superficial, "cortical" regions, the *ventromedial prefrontal cortex* (vmPFC) and the *anterior cingulate cortex* (ACC), to reflect cue valence. In contrast, the striatum reflected whether a person would take ("Go") or withhold ("NoGo") an action. I did not find clear evidence whether the ACC was more active when biases were suppressed, most likely because participants in this study made many mistakes on trials on which biases had to be inhibited. I found that theta power in the electrical brain waves did not reflect the suppression of biases. Instead, theta power reflected the eventual action (Go or NoGo) a person would make. In fact, both theta power and striatal signal were related, suggesting that theta power measured over the scalp could be used to as an index of striatal action selection processes. In sum, I found support for the origin of Pavlovian biases in cortical regions (vmPFC and ACC) which then influence action selection processes in the striatum.

In **Chapter 3**, I investigated the neural mechanisms of how Pavlovian learning biases arise. Not only people's action selection, but also their learning can be biased: People find it easier to attribute rewards to having made an action than to having held back. Vice versa, they find it harder to attribute a punishment to having held back than to having made an action. These biased learning

mechanisms can explain Pavlovian biases in behavior, as well. However, their neural origins are unclear. Again, given that the striatum is both involved in action selection and in learning from rewards and punishments, it seemed plausible that the striatum would give rise to learning biases. However, other cortical regions, such as the ACC, have previously been found to influence how much the striatum learns. Also, cortical regions are involved in learning. Again, I combined fMRI, which allowed me to measure activity in regions at a high spatial resolution, with EEG, which allowed me to measure precisely when certain learning signals occurred.

Using mathematical modeling of choice behavior, I found evidence that participants' behavior was best described by a combination of both Pavlovian response biases and Pavlovian learning biases. The activity in several regions, including vmPFC, ACC, and striatum, was better described by "biased" learning signals than by "standard" learning signals without a bias. The learning signals in different regions were linked to the EEG signal over the scalp at different time points. Importantly, the EEG signal first linked to activity in cortical regions such as the ACC and only later to activity in the striatum. These results are consistent with the idea that Pavlovian learning biases arise first in cortical regions, which then influence later learning processes in the striatum, similar to the mechanisms underlying Pavlovian response biases.

In **Chapter 4**, I investigated whether people "recruit" the Pavlovian system to achieve goals set by other decision systems. Sometimes, the instrumental system decides to invigorate or inhibit behavior because it expects this behavior to lead to a good outcome. However, the instrumental system might become distracted or otherwise fail to implement its goals. It could thus recruit the Pavlovian system as an "auto-pilot" that more reliably achieves its goals. Given that the Pavlovian system gets automatically activated by cues that signal rewards or punishments, people could steer their attention towards such cues in order to trigger the Pavlovian system which then automatically implements the intended action. In two studies, I used *eye-tracking*, a technique that tracks where people look at, to investigate whether people's action plans—to invigorate or inhibit behavior—influence whether they look more at reward or at punishment cues. I also tested whether, vice versa, the amount of attention they direct towards reward or punishment cues influences which action they will eventually execute.

In both studies, I found that, when people needed to invigorate future behavior, they were more likely to look at reward cues, and when they needed to inhibit future behavior, they were more likely to look at punishment cues. The longer they looked at reward (compared to punishment) cues, the more likely were they to invigorate (instead of inhibit) behavior. Participants whose behavior was more strongly influenced by what they looked at were also the ones who performed better at the task. In sum, I found evidence that people can strategically "recruit" the Pavlovian system by using attention to reward and punishment cues in a way that automatically implements the action they plan to do.

In **Chapter 5**, I investigated the neural mechanisms of how the Pavlovian system can be recruited by other decision systems. For this purpose, I used *magnetencephalography* (MEG), a technique similar to EEG which can measure magnetic fields induced by the electrical brain waves at a high temporal resolution. I focused on two specific aspects of these brain waves: "*beta*" power over the center of the head reflects that humans prepare an action, while "*alpha*" power at the back of the head reflects whether humans attend more to the left or the right. I tested whether changes in beta power, reflecting the formation of action plans, predicted changes in alpha power, reflecting attention to reward and punishment cues, and also whether, conversely, changes in alpha power predicted further strengthening of the changes in beta power and the eventual action.

I found that beta power reflected which action (Go or NoGo) participants prepared already several seconds before they executed the action. Also, beta power was influenced by how many rewards or punishments participants could receive, reflecting the influence of Pavlovian biases on action preparation. However, alpha power did not indicate any particular focus on reward or punishment cues. Of note, although participants were instructed to keep their eyes focused on the center of the screen, they still made spontaneous eye movements towards the reward and punishment cues, and these eye movements depended on participants' action plans similar to the results reported in Chapter 4. In sum, I found Pavlovian biases in participants' behavior and in the brain waves that reflected action preparation, but I did not find any differences in the brain waves that reflected attention to reward and punishment cues. Possibly, some decision systems might use eye movements to recruit Pavlovian biases if suitable, but not recruit the more subtle attentional mechanisms that are reflected in alpha power.

In sum, this thesis sheds new light on the neural origin of Pavlovian biases and the control over them. I have found evidence that neural signals underlying Pavlovian response biases and learning biases arise first in cortical regions (vmPFC and ACC) and then get transmitted to the striatum, where they influence action selection and learning. Also, I have found evidence that humans are not merely passive subjects to Pavlovian biases, but can actively recruit these biases in situations in which they can serve the implementation of action plans. In these cases, people use eye movements to cues that signal rewards and punishments, which triggers the biases and "automatically" implements the planned action. I have found signals in the brain waves that reflect these biases already seconds before the eventual action. However, I have not found evidence for how exactly action preparation processes can bias the strategic recruitment of attention. In sum, this thesis shows the neural mechanisms underlying Pavlovian biases and how these biases can become a "tool" for other systems, demonstrating examples of synergy between different decision systems.

DEUTSCHE ZUSAMMENFASSUNG

Menschen verfügen über verschiedene Systeme zur Entscheidungsfindung. In manchen Situationen investieren wir viel Zeit und Mühe darin, die „richtige" Entscheidung zu finden—d.h. in der Regel diejenige, die zu der größten Belohnung führt. In anderen Situationen geht es hingegen darum, möglichst schnell eine Entscheidung zu treffen, ohne lange darüber nachzudenken. Die daraus resultierende Entscheidung mag dann zwar „suboptimal", aber dennoch situativ angemessen sein, da die Gelegenheit, eine wohlüberlegtere Entscheidung zu treffen, gar nicht bestand. Menschen und andere Tiere verfügen also über mehrere verschiedene Systeme zur Entscheidungsfindung, wovon einige schneller sind, aber weniger optimale Ergebnisse erzielen, während andere langsamer sind, dann aber bessere Ergebnisse erzielen. Je nachdem, was eine Situation verlangt, wählen Organismen ein adäquates System zur Entscheidungsfindung aus.

Eine besonders simples, aber dafür sehr schnelles System zur Entscheidungsfindung ist das sogenannte **Pawlowsche Kontrollsystem**, das „**Pawlowsche Tendenzen**" (englisch: Pavlovian biases) im Verhalten hervorruft. Dieses System registriert Umweltreize, die anzeigen, ob gerade die Gelegenheit, eine Belohnung zu erwerben, oder die Gefahr, bestraft zu werden, vorliegt. Wenn Belohnungen verfügbar sind, z.B. in der Form von leckerem Essen, wird dieses System gerade ablaufendes Verhalten weiter verstärken, was es erleichtert, sich der Belohnung zu nähern und sie einzusammeln. Auf der anderen Seite reagiert dieses System auf mögliche Gefahren wie z.B. die Anwesenheit eines gefährlichen Tieres damit, dass es gerade stattfindendes Verhalten unterbricht und damit den Organismus davor schützt, entdeckt oder der Gefahr anderweitig ausgesetzt zu werden. Diese Tendenzen sind adaptiv und führen zu angemessenem Verhalten in der Mehrzahl von Situationen: In der Regel müssen wir bei einer Belohnung schnell „zugreifen" bevor jemand anderes sie uns streitig macht; gleichzeitig müssen wir bei drohender Gefahr eher vorsichtig vorgehen. Nichtsdestotrotz gibt es gewisse Situationen, in denen diese Tendenzen genau das Gegenteil von dem auslösen, was wir eigentlich tun sollten. Manchmal müssen wir nichts tun und einfach warten, damit am Ende ein gutes Ergebnis erzielt wird—beispielsweise den Kuchenteig im Ofen aufgehen lassen anstatt ihn sofort aufzuessen. Umgekehrt müssen wir manchmal aktiv etwas tun, um eine drohe Gefahr oder Strafe zu vermeiden. Unter drohender Arbeitslosigkeit oder Insolvenz müssen Menschen häufig selbst aktiv werden, um ihre Situation zu verbessern, anstatt passiv darauf zu warten, dass sie sich von alleine bessert. In solchen Situationen führen Pawlowsche Tendenzen zu schlechten Ergebnissen und müssen deshalb unterdrückt werden.

Pawlowsche Tendenzen können in einer Vielzahl von Tieren beobachtet werden, darunter auch in Affen, Nagetieren und Vögeln. Diese weite Verbreitung deutet darauf hin, dass diese Tendenzen eine genetische Grundlage haben oder alternativ in den meisten Lebensräumen bereits im frühen Kindesalter erworben werden. Außerdem weist die weite Verbreitung darauf hin, dass evolutionär sehr alte **Gehirnareale**, die in ähnlicher Form in verschiedenen Organismen vorkommen, für diesen Tendenzen verantwortlich sein könnten. Ein wahrscheinlicher Kandidat für den möglichen Ursprung dieser Tendenzen ist das sogenannte *Striatum* (auf Deutsch auch „Streifenkörper"), eine *subkortikale* Struktur tief im Gehirn. Das Striatum ist daran beteiligt, Verhalten zu verstärken oder zu unterdrücken; außerdem hat es Zugriff auf Informationen über mögliche Belohnungen und Bestrafungen. Bisher gab es jedoch keine klare Evidenz dafür, dass das Striatum am Auftreten Pawlowscher Tendenzen beteiligt ist. Darüber hinaus ist die Rolle von anderen Strukturen, die ebenfalls Belohnungen und Bestrafungen verarbeiten, ungeklärt. Dies gilt insbesondere für höhergelegene, *kortikale* Strukturen, wie sie für Menschen und andere Primaten quasi einzigartig sind.

Während andere Tiere große Probleme dabei haben, Pawlowsche Tendenzen zu **unterdrücken** (oder ganz daran scheitern), sind Menschen relativ gut in der Lage, diese zu kontrollieren. Nichtsdestotrotz passieren auch Menschen immer wieder Fehler, wenn sie diese Tendenzen eigentlich unterdrücken sollten. Das Unvermögen, diese Tendenzen zu unterdrücken (aber umgekehrt auch eine zu stark ausgeprägte Veranlagung, sie ständig zu unterdrücken) haben höchstwahrscheinlich Konsequenzen für das mentale Wohlbefinden von Menschen. Bisherige Forschung hat erst Belege dafür geliefert, dass Pawlowsche Tendenzen an der Entstehung von verschiedenen psychischen Krankheiten beteiligt sein könnten, z.B. (Alkohol-) Abhängigkeit, Angststörungen und Depressionen. Aus diesem Grund ist es wichtiger denn je, zu verstehen, wie Menschen den Einfluss von Pawlowschen Tendenzen auf ihr Verhalten gezielt regulieren können.

In den letzten Jahren haben verschiedene Forschungsergebnisse die Annahme, dass verschiedene System der Entscheidungsfindung isoliert voneinander operieren und konstant miteinander im Wettbewerb stehen—darunter auch das schnellere Pawlowsche Kontrollsystem und das langsamere instrumentelle Kontrollsystem (welches Handlungen flexibel auf Basis ihrer zu erwartenden Ergebnisse auswählt) – in Zweifel gezogen. Stattdessen ist es wahrscheinlicher, dass verschiedene Systeme **zusammenarbeiten** und sich gegenseitig zu Nutze machen. Für das Pawlowsche System gilt, dass es recht unflexibel ist, weil es starr mit bestimmten Handlungen auf bestimmte Umweltreize reagiert. Gleichzeitig ist dieses System aber auch schnell und robust gegenüber etwaigen Störungen. Dieser Umstand macht es interessant für andere Systeme, die das Pawlowsche Kontrollsystem „rekrutieren" könnten, um ihre eigenen Ziele zu verfolgen. Beispielsweise könnten andere Systeme den Organismus gezielt solchen Umweltreizen aussetzen, die das Pawlowsche Kontrollsystem aktivieren, und dadurch andere Verhaltensweisen verstärken oder unterdrücken, wenn dies in ihrem Sinne ist. Unter diesem Gesichtspunkt sollte das Pawlowsche Kontrollsystem nicht als isoliert verstanden werden, sondern als ein mögliches „Werkzeug" für andere Systeme.

Zusammengefasst hat diese Dissertation also zwei Ziele: Zum einen sollen die **neuronalen Prozesse**, die zu Pawlowschen Tendenzen im Verhalten führen, näher untersucht werden. Insbesondere soll die Rolle des Striatums sowie anderer Gehirnregionen, die Informationen über Belohnungen und Bestrafungen verarbeiten, näher bestimmt werden. Zum zweiten soll verstanden werden, wie Menschen diese Tendenzen **regulieren** können, was zum einen beinhaltet, zu bestimmen, wie diese Tendenzen unterdrückt werden können, zum anderen aber auch, ob und wie Menschen sich gezielt bestimmten Umweltreizen aussetzen, um diese Tendenzen aktiv hervorzurufen.

In **Kapitel 2** untersuchte ich die neuronalen Prozesse, die Pawlowsche Tendenzen in der Entscheidungsfindung hervorrufen, sowie die Prozesse, die daran beteiligt sind, diese zu unterdrücken. Pawlowsche Tendenzen in der Entscheidungsfindung treten auf, wenn ein Umweltreiz die Chance auf eine Belohnung oder die Gefahr einer Bestrafung signalisiert, während Menschen gleichzeitig entscheiden müssen, ob sie eine aktive Handlung durchführen (eine Taste drücken; „Go") oder lieber zurückhalten sollen (keine Taste drücken; „NoGo"). In dieser Untersuchung erwartete ich, dass das Striatum die Valenz von Umweltreizen (d.h. ob ein Belohnung oder Bestrafung angezeigt wird) verarbeitet, was dann die Handlungsauswahl beeinflusst. Außerdem erwartete ich, dass in solchen Situationen, in den die Tendenzen unterdrückt werden müssen, höhergelegene kortikalen Gehirnareale, insbesondere der *anteriore zinguläre Kortex* (ACC), aktiv werden. Solche Aktivität wurde in der Vergangenheit in Form einer erhöhten Leistung im „*Theta*"-Frequenzbereich in den elektrischen Hirnwellen, die sich an der Kopfoberfläche messen lassen, beobachtet. Diese Aktivität würde dann anzeigen, wie stark

Versuchspersonen versuchen, ihre Pawlowschen Tendenzen zu unterdrücken, und gleichzeitig mit einem reduzierten Valenzsignal im Striatum einhergehen, was die Tendenzen abschwächen würde. Das Ziel dieses Kapitel war es also, den neuronalen Ursprung von Pawlowschen Tendenzen sowie die neuronalen Prozesse, die zu ihrer Unterdrückung beitragen, näher zu bestimmen. Zu diesem Zweck kombinierte ich *funktionale Magnetresonanztomographie* (fMRT), welche es erlaubt, Aktivität mit hoher räumlicher Auflösung sogar in tiefgelegenen Hirnregionen zu messen, mit *Elektroenzephalographie* (EEG), welche es erlaubt, Gehirnwellen über dem Kopf mit hoher zeitlicher Auflösung zu messen.

Zu meiner Überraschung zeigten sich im Striatum keine eindeutigen Signale, die die Valenz von Umweltreizen widerspiegeln würden. Stattdessen fand ich solche Signale in höhergelegenen kortikalen Strukturen, insbesondere im *ventromedialen präfrontalen Kortex* (vmPFC) und im *anterioren zingulären Kortex* (ACC). Das Striatum hingegen zeigte in erster Linie an, welche Handlung (Go oder NoGo) Versuchspersonen ausgewählt hatten. Darüber hinaus fand ich keine eindeutige Evidenz für erhöhte Aktivität im ACC in solchen Situationen, in denen Pawlowsche Tendenzen unterdrückt werden mussten. Dies ist wahrscheinlich darauf zurückzuführen, dass Versuchspersonen in dieser Studie relativ viele Fehler begingen und nur in wenigen Fällen diese Tendenzen erfolgreich unterdrückten. Auch die Theta-Band-Leistung in den elektrischen Gehirnwellen spiegelte nicht wieder, ob Versuchspersonen diese Tendenzen unterdrückten oder nicht, sondern stattdessen, welche Handlung (Go oder NoGo) sie am Ende auswählten. Die Theta-Band-Leistung glich in gewisser Weise der Aktivität des Striatums, was nahelegt, dass man diesen Anteil des EEG-Signals dazu gebrauchen könnte, Aktivität im Striatum während der Handlungsauswahl aus den elektrischen Hirnwellen abzulesen. Zusammengefasst folgt aus den Ergebnissen dieses Kapitels, dass Pawlowsche Tendenzen höchstwahrscheinlich durch Prozesse in kortikalen Hirnarealen (vmPFC und ACC) entstehen, die die Handlungsauswahl im Striatum beeinflussen.

In **Kapitel 3** untersuchte ich die neuronalen Prozesse, durch die Pawlowsche Tendenzen auch im Lernen auftreten können. Nicht nur die Handlungsauswahl, sondern auch Lernprozesse können von diesen Tendenzen beeinflusst werden: Menschen finden es in der Regel einfacher, Belohnungen ihren eigenen Handlungen zuzuschreiben anstatt der Tatsache, nicht gehandelt zu haben. Umgekehrt finden sie es schwieriger, Bestrafungen dafür zu akzeptieren, nicht gehandelt zu haben verglichen mit Fällen, in denen sie eine aktive Handlung gezeigt haben. Solche verzerrten Lernprozesse können gleichfalls erklären, wie Pawlowsche Tendenzen im Verhalten auftreten. Allerdings sind die neuronalen Prozesse, die zu diesen Verzerrungen führen, unklar. Auf der einen Seite scheint – wie bei Tendenzen in der Handlungsauswahl – das Striatum dadurch, dass es an der Handlungsauswahl beteiligt ist, gleichzeitig aber auch beim Lernen von Belohnungen und Bestrafungen eine Rolle spielt, der ideale Kandidat für die Ursprungsregion solcher Verzerrungen zu sein. Auf der anderen Seite hat frühere Forschung aber auch gezeigt, dass kortikale Regionen wie z.B. der ACC beim Lernen von Belohnungen und Bestrafungen eine Rolle spielen und dabei beeinflussen, „wie viel" das Striatum von einer bestimmten Belohnung lernt. Wie schon in Kapitel 2 kombinierte ich fMRT, welches mir erlaubte, Aktivität mit hoher räumlicher Auflösung zu messen, mit EEG, was mir erlaubte, zu bestimmen, wann genau verschiedene Lernsignale auftraten.

Mathematische Modellierungen von Entscheidungsverhalten ergaben, dass das Verhalten von Versuchspersonen am besten dadurch erklärt werden kann, dass Pawlowsche Tendenzen sowohl in der Handlungsauswahl als auch im Lernen von Belohnungen und Bestrafungen auftreten. Darüber hinaus zeigte sich, dass die Aktivität in verschiedenen Gehirnregionen, darunter vmPFC,

ACC, und Striatum, eher „verzerrten" Lernsignalen glich als Lernsignalen ohne solche Verzerrungen. Außerdem bestanden Zusammenhänge zwischen den Lernsignalen in verschiedenen Gehirnregionen und dem EEG-Signal – und zwar zu verschiedenen Zeitpunkten: Das EEG-Signal spiegelte erst Signale aus kortikalen Regionen wie dem ACC wider und erst später Signale aus dem Striatum. Dieser Ergebnisse legen nahe, dass – ähnlich wie Pawlowsche Tendenzen in der Handlungsauswahl – entsprechende Verzerrungen im Lernen von Belohnungen und Bestrafungen zunächst in kortikalen Regionen auftreten und dann später Lernprozesse im Striatum beeinflussen.

In **Kapitel 4** untersuchte ich, ob Menschen das Pawlowsche Kontrollsystem aktiv „rekrutieren" können, um die Ziele anderer Systeme der Entscheidungsfindung zu verfolgen. Manchmal entscheidet sich das instrumentelle Kontrollsystem dafür, Verhalten weiter zu verstärken oder zu unterdrücken, weil es sich davon ein besseres Ergebnis verspricht. Allerdings kann dieses System womöglich abgelenkt werden oder aus anderen Gründen daran scheitern, seinen Handlungsplan in die Tat umzusetzen. Deshalb könnte es das Pawlowsche Kontrollsystem als eine Art „Auto-Pilot" rekrutieren, der das gesetzte Ziel genauso gut erreichen kann, wobei die Wahrscheinlichkeit, zu scheitern, jedoch geringer ist. Aufgrund dessen, dass das Pawlowsche Kontrollsystem in quasi „automatischer" Weise auf Umweltreize reagiert, die mögliche Belohnungen oder Bestrafungen anzeigen, könnten Organismen gezielt ihre Aufmerksamkeit auf solche Reize lenken, die das Pawlowsche System aktivieren, welches dann automatisch die gewünschte Handlung ausführt. Um dies zu testen, führte ich zwei Studien durch. Hierbei benutzte ich *Eye-Tracking*, eine Technik, mit der sich messen lässt, worauf genau Versuchspersonen ihren Blick fokussieren. Ich untersuchte, ob Handlungspläne (Verhalten zu verstärken oder zu unterdrücken) dazu führen, dass Versuchspersonen sich gezielt mehr auf positive oder auf negative Umweltreize konzentrieren. Außerdem testete ich umgekehrt, ob die Aufmerksamkeit, die sie positiven und negative Umweltreizen schenken, einen Einfluss darauf hat, welche Handlung sie am Ende zeigen.

In beiden Studien zeigte sich, dass, wenn Versuchspersonen zukünftigen Handlungen tatsächlich durchführen mussten, sie sich eher auf Belohnungsreize konzentrierten, während sie sich dann, wenn sie in ihrem Verhalten einzuhalten hatten, eher auf Bestrafungsreize konzentrierten. Außerdem erhöhte die Aufmerksamkeit, die sie Belohnungsreizen (verglichen mit Bestrafungsreizen) schenken, die Wahrscheinlichkeit, dass sie am Ende Verhalten verstärken (anstelle von unterdrücken) würden. Zuletzt zeigte sich, dass diejenigen Versuchspersonen, deren Verhalten besonders an ihre Aufmerksamkeit gekoppelt war, bessere Leistungen in der gestellten Aufgabe zeigten. Zusammengefasst legen diese Ergebnisse nahe, dass Menschen ihr Pawlowsches Kontrollsystem strategisch „rekrutieren" können, indem sie ihre Aufmerksamkeit gezielt auf Belohnungs- oder Bestrafungsreize lenken – und zwar dergestalt, dass ihre Handlungspläne automatisch in die Tat umgesetzt werden.

In **Kapitel 5** untersuchte ich, mithilfe welcher neuronalen Prozesse das Pawlowsche Kontrollsystem von anderen Systemen rekrutiert werden kann. Zu diesem Zweck benutzte ich *Magnetenzephalographie* (MEG), eine Technik, die ähnlich wie EEG funktioniert und mit hoher zeitlicher Auflösung magnetische Felder, die von elektrischen Hirnwellen induziert werden, über dem Kopf messen kann. Ich konzentrierte mich auf zwei besondere Aspekte dieser Hirnwellen: zum einen die „*Beta*"-Band-Leistung über der Mitte des Kopfes, die widerspiegelt, inwieweit Menschen gerade eine Handlung planen; zum anderen die „*Alpha*"-Band-Leistung über der Rückseite des Kopfes, die widerspiegelt, ob Menschen sich mehr auf die linke oder die rechte Seite ihres Blickfeldes konzentrierten. Im Einzelnen untersuchte ich, ob Veränderungen in der Beta-

Band-Leistung, welche auf die Formierung von Handlungsplänen hinweisen, zu systematischen Veränderungen in der Alpha-Band-Leistung führten, was auf strategische Aufmerksamkeit auf Belohnungs- oder Bestrafungsreize hindeuten würde. Außerdem testete ich umgekehrt, ob Veränderungen in der Alpha-Band-Leistung bestehende Veränderungen in der Beta-Band-Leistung weiter verstärkten und dadurch die abschließende Handlung beeinflussten.

Meine Ergebnisse zeigten, dass sich anhand des Beta-Band-Leistung der Gehirnwellen bereits mehrere Sekunden, bevor eine abschließende Handlung ausgeführt werden musste, ablesen ließ, welche Handlung (Go oder NoGo) ausgeführt werden würde. Außerdem spiegelte die Beta-Band-Leistung die verfügbaren Belohnungs- und Bestrafungsreize wieder, was zeigte, wann und wie Pawlowsche Tendenzen in neuronalen Prozessen auftreten. In der Alpha-Band-Leistung der Gehirnwellen ergaben sich keine Hinweise darauf, dass Versuchspersonen sich strategisch auf Belohnungs- oder Bestrafungsreize konzentrieren würden. Allerdings ergaben sich Hinweise auf einen solchen „strategischen" Einsatz von Aufmerksamkeit anhand der Augenbewegungen der Versuchspersonen: Obwohl sie angewiesen waren, ihren Blick auf die Mitte des Bildschirm zu fokussieren, zeigten manche Versuchspersonen unwillkürliche Augenbewegungen hin zu den Belohnungs- und Bestrafungsreizen. Diese Augenbewegungen waren von ihren Handlungsplänen beeinflusst – und zwar in gleicher Weise wie in den Ergebnissen in Kapitel 4. Insgesamt zeigte sich in diesen Ergebnissen erneut der Einfluss von Pawlowschen Tendenzen – sowohl im Verhalten der Versuchspersonen als auch in den Gehirnwellen, die die Vorbereitung solcher Handlungen mehrere Sekunden vor ihrer Ausführung widerspiegeln. Allerdings zeigte sich kein strategischer Einsatz von Aufmerksamkeit auf Belohnungs- und Bestrafungsreize in den Gehirnwellen, die räumliche Aufmerksamkeit wiederspiegeln. Möglicherweise können Entscheidungssysteme Einfluss auf Augenbewegungen nehmen, um Pawlowsche Tendenzen strategisch auszulösen, haben aber keinen Zugriff auf subtilere Mechanismen, die sich in der Alpha-Band-Leistung von Gehirnwellen widerspiegeln.

Zusammengefasst tragen die Ergebnisse dieser Dissertation dazu bei, die neuronalen Grundlagen von Pawlowschen Tendenzen sowie die Kontrolle über diese besser zu verstehen. Meine Ergebnisse zeigen, dass Pawlowsche Tendenzen sowohl in der Handlungsauswahl als auch in Lernprozessen zunächst in der Aktivität kortikaler Regionen (vmPFC und ACC) sichtbar sind und diese Signale dann weiter ins Striatum geleitet werden, wo sie die Handlungsauswahl sowie das Lernen von Belohnungen und Bestrafungen beeinflussen. Außerdem zeigen meine Ergebnisse, das Menschen keine passiven „Opfer" dieser Tendenzen sind, sondern sie aktiv rekrutieren können – nämlich dann, wenn sie existierende Handlungspläne dadurch besser in die Tat umsetzen können. In solchen Fällen konzentrieren sie sich auf Umweltreize, die die Chance auf Belohnungen oder die Gefahr von Bestrafungen anzeigen und Pawlowsche Tendenzen auslösen, die automatisch die gewünschte Handlung hervorrufen. Gehirnwellen spiegeln diese Tendenzen bereits Sekunden vor der eigentlichen Handlung wieder. Allerdings ergaben sich keine weiteren Hinweise darauf, wie genau die Aufmerksamkeit in strategischer Weise auf entsprechende Reize gelenkt werden kann. In der Summe enthält diese Dissertation zum einen neue Hinweise auf die neuronalen Prozesse, die Pawlowschen Tendenzen zugrunde liegen. Zum anderen ändert sie unsere Sichtweise auf das Pawlowsche Kontrollsystem als ein „Werkzeug" für andere Systeme der Entscheidungsfindung, was zeigt, dass verschiedene Systeme nicht (immer) im Wettbewerb zueinander stehen, sondern im Gegenteil häufig zusammenarbeiten, um ein bestimmtes Ziel zu verfolgen.

ACKNOWLEDGEMENTS

After some 300 pages so far, here now comes another major section. In the past months, I have seen a few theses submitted in the UK, which usually feature one page of rather concise (and often cryptic) acknowledgements. But like many aspects of the Dutch PhD thesis, also this section seems to have its own "logic" and implicit rules that determine how it should look like. And given this scheme combined with what some people have remarked about my memory of names, they can probably imagine what will come now.

In 2015, I deliberated whether to continue my studies in psychology or in philosophy. One of the reasons why I opted against philosophy was that I anticipated that no one (except for my supervisors and evaluators) would ever read my PhD thesis. Now, looking back, I am also certain that the PhD "experience" would have been very different—probably (a lot more) reading, writing, and teaching all day. This section would have certainly been shorter. And the same would have been true had I written a thesis in psychology several decades ago. However, empirical PhD theses nowadays are so different in many ways. We operate with data that cannot be analyzed with pocket calculators any more. We use expensive machines that break and have to be maintained. There are so many legal restrictions and requirements a university has to meet nowadays. There are so many people that do (or teach us how to do) all these things without whom a PhD would nowadays be absolutely impossible. And this goes beyond official supervisors. Hence the length of this section.

First of all, **Hanneke**: Thanks for allowing me to be your first "own" PhD student (Jenn already claimed the title of "first primary PhD student")! I vividly remember my two interviews without you and how, in the first one, we discussed a few papers (among them the paper featuring what became the "twinkling stars" task) which I thought fitted exactly in what you were probably interested in and what I expected I would be doing during my PhD (but then, of course, things worked out quite differently). I also remember how, in the second one, Pieter asked some questions on optimality and motor control that threw me a bit off, and that I did not have the best feeling afterwards—the more surprised and happier I was when, only two days later, I received your email offering me your open position (it was Wednesday, June 28th 2017, at 17:51)! Our joint journey was for sure rather "special" compared to typical supervisor-student trajectories—two long maternity leaves and COVID meant that, in sum, we had more online than in-person meetings, and I felt that only in my fourth year, things started to feel like I had always imagined a PhD trajectory to be. In some ways, we were a great team and thought alike about many things, for example how to deal with administrative problems or which questions to ask to external speakers about their work (private zoom chats were a COVID innovation I sometimes miss). At other times, we were in strong disagreement (for example about the paradigm that became the basis of chapters 4 and 5) and Pieter had to mediate between us. In that case, I pushed through my will, and probably just got lucky that this version of the paradigm "did work" in the end. In retrospect, I think that not many supervisors would have given me "my will" as often—as well as so many opportunities to try out new things (using almost all the different recording methodologies of human cognitive neuroscience). In this way, I have learned much more than what eventually made it into this thesis. You have definitely shaped my way of doing research, writing, and thinking—probably in many more ways than I can oversee right now. I certainly did not get infected by your enthusiasm for sans serif fonts, "conceptual" figures, and Adobe Illustrator (although I see a dire need to start using it soon). But there are other things which, in case I will ever become a PI myself and have my own group, I will hopefully remember and try to solve in a similar way as you did. Thanks for picking and guiding me through almost five years of

science, eventually convincing me to stay in neuroscience (which, in early 2017, I would have never thought), and thanks for all your efforts to secure me extra time, funding, assistance, and motivation to push this thesis over the finish line! I wish you all the best (including a wonderful lab) for your future career with many more insights into meta-control and "automatisch gedraag".

Secondly, **Pieter**, thanks for being my promotor and keeping an "oversight" on my thesis progress in the second half of my PhD. What a twist of fate that after my second PhD interview (see above), you got more involved into my PhD again and that certain topics returned ("what do you mean with 'optimal'"?). These topics still haunt me today (why exactly does reward invigorate action? see the General Discussion of my thesis), also in my ongoing postdoc work (what do you humans actually want to "maximize" in decision-making?). Your calm and sober approach to dealing with problems and people, especially when things get a bit heated, has in the end ensured the success of my PhD trajectory. When COVID started, you scheduled regular zoom meetings with all PhD students and took our concerns about lab openings, the move to the new Maria Montessori building, and COVID extensions seriously. You made sure that there was a roadmap towards finishing my PhD and in retrospect, it still somewhat surprises me that I managed to submit it on time. I will remember your approach to leadership in case I ever come into a similar position. Also, **Roy**, thanks for having been my promotor and hosting me in your group for the first two years; it was a great time and I learned a few things about neuropsychology that have definitely enriched my PhD horizon!

Next, apart from my "official" supervisors, there are (at least) four more people that deserve special mention for having contributed to the success of my PhD trajectory:

First, **Bernd**: You have been the "longest"-lasting" supervisor I have ever had, and I was extremely happy to have your "non-neuroscience" perspective on my work! When I started the Behavioural Science Research Master, I liked the social psychology in it, but I also wanted something more "cognitive", mathematical to work on. Your two lectures on risky and intertemporal choice plus your course on linear mixed-effects models (probably the most useful "psychology" course I ever took) determined that I wanted to work with you. My research as a faculty assistant with you, Isa, and Hannah gave me the first real taste of doing my own research and strengthened my conviction that this was what I wanted to keep doing. A shame we still have not managed to publish a paper together (but hopefully soon…) Thanks for all the feedback and inspiration I have received from you, all the hints to obscure papers I would otherwise never have come across ("this has already been done 30 years ago by the guy who also did…"), thanks for the high level of methodological rigor and scientific integrity you have taught me, and thanks for providing me a second "home" where to discuss my results and ideas (and rant about other aspects of science) in the last years!

Second, **Robert**: I knew basically nothing about M/EEG analyses when I started my PhD (apart from some scripts I had inherited), and over time, I think I have learned most of the additional skills I now possess—regarding M/EEG, but also data analysis in general—from you. How often have I asked you about how certain things "work" or why "people in EEG research" did things in a certain way, and your answer was often "Well, why don't you simulate it?" This approach has had a huge impact on how I do analyses now. Through you, I feel I have finally understood what happens during M/EEG preprocessing. You were the kind of "supervisory collaborator" who took PhD students and their ideas seriously and encouraged them to try things out of their usual comfort zone—for example directly contributing to Fieldtrip. Moreover, the way Fieldtrip is set up had a lasting impact on how I write code (involving two major "rewrites of

everything" during my PhD). It was great to have you as another "ally" who pushed for more open science (in various flavors) at the Donders and taught me several things on reproducibility and data sharing via the Donders Repository. I think every neuroscience institution should be very happy to have someone like you.

Third, **Roshan**: Thank you that, in absence of my "academic mother", I could spend some more time in the lab of my "grandmother"! Your enthusiasm about all kinds of research vaguely related to motivation and cognitive control (and your patience to answer all my slightly "derailing" questions, even in after-BCS lunches at the Trigon canteen), but also your critical attitude to the things I did myself ("why again do you need that extra analysis?") have certainly made me think harder about what I was doing. I still vividly remember some specific moments (usually during my own BCS presentations) in which you pointed out an idea or a paper I had not yet heard of—for example encoding of time in the striatal signal or residual switch costs—which had a lasting impact on the "big picture" of this thesis (and eventually the General Introduction and Discussion of it). Beyond the content, some traces of Trevor Robbins's writing style might have ended up in mine through you ;-).

Finally, a special mention goes to my first "primary supervisor" during the first year of my PhD: **Jenn**, thanks for taking care of me during Hanneke's first "big absence"! The many questions/ new suggestions I had ("We could do this a bit differently…") and my initial reluctance to use MATLAB have certainly pushed you a bit out of your comfort zone while finishing your own PhD. Thanks for your patience with me. But the biggest thanks goes for the huge data set you (and Emma) inherited me, which kept me busy for five years and hopefully (as the legacy of your own PhD) will eventually be published! Also, thanks for all the Stan and Fieldtrip code I inherited from you and whose spirit lives on in all my data sharing collections. Thanks for teaching me the very first steps of becoming a neuroscientist.

Apart from the people that actively contributed to my PhD trajectory, a few other people deserve special mentions for having had an impact on it already before it started: From my time in Jena, **Klaus Rothermund**, for whom I have worked as a research assistant, was likely the paramount influence that drove my interests towards motivation and learning. Special thanks also to **Thomas Weiß**, who taught me the basics of neuroscience and the very first bits of EEG—although after our joint research project, I was pretty convinced that I would never do neuroscience again! From my Master's period in Nijmegen: **Rob**, thanks for having been an absolutely great head of the Research Master and Master's thesis supervisor! I was rather skeptical about staying in the Dutch academic system during the first year of my Master's, but that changed quickly while working with you on my Master's thesis. You and the department of social psychology (the "9th floor") convinced me that Radboud and Nijmegen were a great work environment and even though I ended up not doing my PhD with you, I was always happy to see you again from time to time when you randomly walked into my office on Friday afternoons and asked me how things were going. Without you, my decision to pursue a research career and to stay in Nijmegen would have likely turned out differently. Also, **Erik**: thanks for teaching me pupillometry and how to publish my first "own" paper! Your pragmatic approach on how to write and how to deal with reviewers (although it involved a lot of "red text") will have a lasting impact on my own writing.

Beyond supervisors, there are a few more people (or groups) from whom I learned a lot about science and data analysis over the years: Most importantly the weekly *M/EEG meeting* on Mondays, organized by Robert, but involving various people that have given me useful many useful pieces

of advice and help on all kinds of data collection and analysis questions of the years—with special mentions to **Jan-Mathijs, René, Phillip, Eelke, Bob, Christoph, Britta, Vitória, Mats, Jakub, and Maria Carla** (your names still appear in my code ;-). Thanks also for allowing me to participate as a tutor in the Donders M/EEG Toolkits, which were educational for me as well. Special thanks for EEG-related help goes also to **Jan-Mathijs Schoffelen**, **Mike Cohen**, **Eric Maris**, and **Tom Marshall** who discussed my ideas and data with me at various stages. Regarding fMRI, I would like to thank **Rogier Mars** for his prompt responses on intricate questions about neuroanatomy (even when I was not satisfied with the answer "just report the Brodmann area") as well as **Maarten Mennes** and **Christian Beckmann** for practical help on FSL, bash, nuisance regressors, and elaborate discussions on how to deal with empty regressors. Lastly, **José**: Thanks for organizing the FAM meetings, which I think any neuroscience institution would be happy to have! (Also, special thanks for running the Zevenheuvelenloop with me twice during COVID!) As a last source of new inspirations, I'd like to thank the *Systems Neuroscience Journal Club (SNJC)*, especially **Mike, Eric, Paul, Nils, Arthur, Teo, Mats, Ashutosh, Nader, Marrit, Mitch, and Jordi**, from whom I have learned how to "read" animal neuroscience papers. Even though this JC never directly related to my daily work, I opened a "new world" to me and definitely shaped this thesis.

Beyond the "content" of my academic work, there were a few people in the background who made it possible in the first place that I could do all the research I did: At the DCC, special thanks go to **Ronny, Miriam**, and **Pauline** for running the labs (even if they move buildings or become suddenly flooded) and trying to make our administrative paperwork easier to manage! Special thanks also to **Hubert**, who taught me everything about eye-tracking and several times made a severe hardware or software problem disappear! **Gerard**, the "good soul" of the technical support group, for helping me master all the obstacles of data collection at the DCC (including custom button boxes and even an unobtrusive camera to monitor from outside when a participant had finished their task)! At the administrative level, thanks to **Jolanda, Vanessa, Saskia, Karin**, and **Maaike** for keeping the DCC running—although administration often appears unnecessarily cumbersome from the researcher perspective, I assume it would be basically unmanageable without people like you. Next, at the DCCN, very special thanks to **Hong** for keeping the DCCN computing cluster running (without which this thesis would not have been possible) and for eventually making Stan available on it; **Edward** for his patience during the long afternoons on which we tried to install certain R packages on the cluster together; and finally **Marek** and **Mike** for running the DCCN TG and always solving every problem I brought when (once again) entering the TG office. Regarding MEG, thanks to **Uriel** for making Python available in the MEG lab, **Kristina** for training Jesse and me on MEG data acquisition, and **Paul** for doing the T1 scans for my MEG participants. Last but not least **Tildie**, thanks for the many small things you do for everyone at the DCCN (which we often don't even see)! From the first year onwards (when I co-organized the Donders Discussions and had to knock on your office door quite a few times), you have always solved any request I had, especially in those delicate things that bridge across centers. Thanks also for hosting the runners at every Zevenheuvelennacht and -loop, it's great to have someone who actively supports the involvement of the Donders in such outside activities!

Lastly, thanks to a few people outside the Donders who helped me through my PhD academically. In all honesty, I think I would not have finished this work without the help of **Google**, **Stackoverflow**, and the **Stan Forum**. It might sound weird to acknowledge these platforms (and the various people who contribute to them), but they are of principle importance for the scientific work we do nowadays and why it looks so different from the work done several decades ago. I could also mention various blogs and podcasts—among those, I probably owe most

to the blogs run by **Andrew Gelman** and **Tal Yarkoni** for fundamentally shaping my view on statistics and neuroscience. Thanks also to three special people for sharing elaborate code with me: **Tobias Hauser**, **Laurence Hunt**, and **Nathaniel Daw**—it is so great to not have to reinvent the wheel, but to be able to profit from the hard work of others! As Laurence has put it in a commentary, I hope that "the life-changing magic of sharing your data" (and code) will soon become the standard in neuroscience.

Next, after this long section of gratefulness to people that "came before me" in science, it's finally time to give thanks to my academic "siblings"—first of all in the *Learning and Decision-Making (LDM) lab*:

Elena: thanks for being my first "academic sister" and second non-Dutch person in our lab to keep the international atmosphere alive ;-) Thanks for all the random Italian moments! I still sometimes have hallucinations of you saying "boh" behind my back. After two and a half years of relative "solitude" as the only PhD in the lab, it was such a relief to receive validation from someone else on all the things that were annoying about doing a PhD. Your trajectory has certainly also not gone as planned, and you might unfortunately not benefit from things like COVID extensions as much as I did, but I'm still very much hoping for a "happy end" for you!

Floortje: thanks for being my "second sister" and relieving me from various "admin" roles in the lab. I still remember our first "real" encounter at the Winter School in which you translated the responses I gave (in English) to questions (posed in Dutch) I got from high school teachers—how great to have someone else that could do all the "Dutch" things I never really wanted to do/learn! You are now finally doing the kind of work (and use the paradigms) I was at the start of my PhD looking forward to do myself, and I am glad that someone is actually finally attacking these "meta-control" questions (a bit of jealousy kicking in). You always were a bit reluctant to accept my "help" (or "long derailing monologues full of new ideas" ;-) and that might potentially have been a good decision to keep your PhD on track. I am very curious about how your thesis will eventually look like!

Ben: we did not spend much time together (probably because your PhD is the complete opposite of mine regarding time spent in the lab vs. office), but thanks for being my first teacher on how to do ultrasound stimulation and for the various random visits to our office. Now that you have my desk space, I hope you know to appreciate your lovely office mates ;-)

And finally **Soha**, my principle paranymph: it was so great to have you around in the lab and finally develop a joint "office routine" in the last year of my PhD! Even though you (smart as you are) soon realized all the annoying things about doing a PhD, you also brought the extra energy that is needed to actually attack some of these problems. When you started in the lab, I felt we finally grow together and started building a real "lab life", and I think that was no coincidence. Unfortunately, our overlap was only for eleven months—I wished we could have worked together for a bit longer (and actually do this "collaboration" you always talked about ;-). Thanks so much for helping me along all thesis- and defense-related things! I hope that, through ultrasound, our paths will cross more often in the future.

Also, thanks to Dr. **Renée**—we did not overlap much in our time in the lab, but your tips on managing Oxford were already of great help, and your own thesis layout was the foundation for how this thesis ended up looking like!

Thanks also to my RAs **Jesse**, **Helena**, and **Nele**, for helping me collect quite some data under stressful circumstances (Covid, lab floodings, MEG warmups). **Jesse**, I am very happy I had you as a my (first ever) master student, it was great to have you around and even though you did not end up in science, I hope you learned a few useful things from me! Thanks also to all former lab members not mentioned yet—**Annelies** (thanks for all the good vibes, the cakes, and all the inside talk I needed to find my way around!), **Emma** (thanks for your fMRI notes and all the data!), **Vanessa** (thanks for letting me not forget my German, for the endless chats about German and Dutch academia, MATLAB, computational psychiatry, all the shared frustrations and words of wisdom on how to survive a PhD… hope to see Dr. Dr. Scholz at conferences in the future!), and **Mojtaba** (I probably talked to much and too fast to you most of the time, but thanks for all the small things you contributed to the lab atmosphere over the last years)!

Beyond my lab, I had a few other flocks of academic "siblings":

Special thanks to my "adopted academic siblings" Sush, Giacomo, and Charlotte who accepted me into "their lab" and for the endless hours at the Cultuurcafé on Friday evenings where we discussed the state of neuroscience and general politics. **Sush**, thanks for teaching me many things about artificial neural networks and mind uploading and making me realize that that is not the direction I want to go into. **Giacomo**, thanks for taming Sush at various moments and then putting things in a coherent order so they actually made sense. **Charlotte**, thanks for the many discussions about the Dutch academic system, Dutch politics and way of life, and all the other things that we should do better in life! Moreover, special thanks to my "big sis" (and long-hours co-worker) **Xiaochen** for being a soul mate in many respects, first on the second floor of Spinoza B and later in the BCS. In retrospect, it seems like quite a coincidence that our paths crossed and stayed parallel for so long! I am curious where your sweet spot for M/EEG will eventually lead you. Finally, **Patricia**: I remember first seeing your name on a multi-people email conversation we were both part of (probably about BCS presentation dates) and me wondering "who is this person that I haven't met her yet?" Similar to Xiaochen, I consider it quite a "twist of fate" that we eventually got to know each other better and became friends! I had even imagined how we might start a collaboration on the drivers of curiosity at some point and eventually would both finish our PhDs together (and even be paranymphs for each other?), which did not work out in the end—still, it was so great to have met you, and I hope our paths will cross again!

Next, thanks to all the (other) people in the *Behavioural Control Seminar (BCS)*—**Esther**, **Andrew** (such a shame you were only in Nijmegen for one year, hope we see each other in the future!), **Eliana** (yet another "twist of fate" we met each other, how often have we said "let's talk more soon" and then something else intervened, hope to see you again at conferences as well!), **Bram**, **Marieke**, **Jessica**, **Monja** (my German confederate and another source of important inside knowledge!), **Ceyda**, **Rebecca** (special thanks again for being my weekend-MEG-buddy!), **René**, **Frank**, **Naomi**, **Claudia**, **Marpessa**, **Lieke 1**, **Lieke 2** (Patricia and me always had our disagreements about who counts as which one), **Sophie**, **Iris**, **Ruben**, **Danae** (again, special thanks for scanning some of my MEG participants on the weekends!), **Jorryt**, **Britt**, **Ping**, and **Natalie**.

Thanks also to the members of the *Decision, Development, and Psychopathology (D2P2) lab* of which I was part for 6 years—thanks for all the dedicated discussions on decision-making, statistics, evolution, Radboud politics, and all the "lab drinks" we had: **Isa**, **Iris**, **Jesse 1**, ~~**Jesse 2**~~ **Achiel**, **Nicole**, **Joppe**, **Edwin**, **Lena**, **Farnaz**, **Floor**, **David E.**, **Felix**, **Leslie**, **Tamara**, **Gaia**, **David**

R., **Fritz**, and **Mingqian**, thanks for all various bits of intellectual stimulation you have been responsible for!

Lastly, thanks to various people who "walked a part of the PhD way" with me at various points and provided other relevant knowledge on how to survive it: First my "floor mates" on the 2nd floor of Spinoza B—**Ileana** (my first office mate! A shame we lost contact through COVID), **Joao**, **Fenny**, **Daniel**, **Nikki**, **Selma**, **Syanah**, **Josi** and **Anna-Sophie**—then my new floor mates in the Maria Montessori building—**Maëlle**, **Marco**, **Lu-Shun**, and **Qiu**—my "class" of fellow PhD candidates—**Peta**, **Randi**, **Lukas**, and **Josh**—and members of the Verhagen lab—**Lennart**, **Andrey**, **Sjoerd**, and **Solenn**—and finally **Marius** for being my academic "mentor" (who actually listened to some of my scientific problems).

Thanks also to "the next generation", i.e., my Bachelor thesis students through whom I learned how to be a supervisor (I hope I was not too bad to begin with; and also that you eventually learned something that proved useful outside of your studies)—**Karlijn**, **David R.**, **Geert**, **Pim**, **Madeleine**, **Theresa**, **Eva**, **Max**, **Sezin**, and **Alex**. I still very much hope to share all the data you collected with the rest of the world at some point!

Finally, there should still be room for acknowledgements of some of those people who had no direct role in my PhD (and actually little clue of what it was doing), but who contributed to my mental sanity and motivation throughout these five years: Most importantly the *Sprittwoch* group:

Ilse, "lievert" with a "t": given our first impression of each other in the Research Master, who would have thought that, one day, I would "follow you" to Oxford! Our time together on the 9th floor was one of the happiest of my life so far, and even though you "left" us straight after the Master, I was looking forward to each of your frequent visits! We are so different in many respects (especially regarding music and food; and I am sometimes very judgmental) so it seems somewhat hilarious that you became one of my paranymphs! Still, our common enthusiasm for the "core business" (open science and other attempts at improving the world) have united us, and your habit of sending me anything that could be vaguely interesting to me shows how much you cared about me. You tried to convince me at the start of my PhD to promise you that, by its end, I should have learned proper Dutch. I think I did not manage that (though I did successfully sell all of my furniture in Dutch!), but I hope you are still somewhat proud of my progress. I hope that, one day, Dr. Pit and Dr. Algermissen will meet again as old (and hopefully not too grumpy) pensioners (if that role will still exist) and fondly look back on good old times in Nijmegen and Oxford.

Julian, Genosse: it was a pleasure to be your ~~colleague~~ comrade for so many years (which was essentially your achievement given that you were more convincing to Rob than me). We definitely should have collaborated on some projects; a pity that COVID came in our way. Your sober way of attacking scientific disputes (and the opposite approach when talking about world politics after 1 am) have certainly made some of the highlights of my past seven years. It was great to have someone at the other end of Mattermost to ask random stats-related questions when Google did not yield a satisfactory answer, and to share one's frustrations about brms/ R/ Python/ Radboud/ academia. I remember how we both hiked through the Ooijpolder (at the beginning of our Vierdaagse preparations) on the Saturday before the COVID lockdown started, and how we talked about picking up online studies for a while, but how hard COVID has eventually hit our PhDs. Though our lives have eventually moved apart quite a bit (and a joint paper will likely never happen), I am very glad to have you as a friend and scientific partner in crime for such a long time, and eventually also as a paranymph!

Peng, Genosse: it was an honor to have been your office mate. You challenged me very hard on various topics (usually related to Christian faith, academia, global politics, and the meaning of life), preferably after 1 am or when I was otherwise very busy. You forced me to explain things clearly and repeatedly and why the glass was overall half-full. I would have been hilarious if you would have become the fourth PhD out of our office in the end (instead of just getting rich). Who knows, let's see what the future brings for both of us and who of us will have to admit at some point in the far future that the other one was right back then on that one night in Nijmegen…

Mandy: your amazing positivity and enthusiasm (the "sunshine" of our group) were just outstanding in the last seven years (while writing this, I realize that this might sounds like one of the gender stereotypes men often use to write about women—but in your case, it reflects a truly special character). When I—usually tired and grumpy about PhD life—eventually arrived at Bloemenburgerhof on some days and everyone else was busy talking/ eating/ playing Mario Kart, for you, it was always priority to welcome me and ask how I was doing. Also, you were a great source on how life outside academia has its not-so-great days, as well… Even though your career has (so far) not led you back to academia or towards a PhD, and even though I can imagine how boring the discussions between Julian and me must have been at some evenings, you always accepted our nerdiness and had an open ear for all our scientific and non-scientific problems. I am hoping to gets hugs from exactly this kind of Mandy still many years from now.

My "acting little brother" **Jonas**: Years ago, I would have never thought that you would be the one who finishes his PhD first! Looking back at the labor division during our Minor Project in the Research Master, I felt reminded of the little brother I had at home, and would have never thought that there once would be a Dr. Wachner to talk to. In retrospect, I must call myself very glad to have been your partner in crime during our Minor (which has also introduced me to the following lovely people) and I hope that your future work will give me some examples I can tell other people about when they ask how "useful" psychological research is for society.

Luise: thanks so much for being the "good soul" of our group (although you should probably stop always cleaning after some of us)—again, this might sound like a gender stereotype, but it is a trait many people should wish they had. You might be less of an aggressive "hugger" than Mandy, but you have always cared when I arrived on a rainy day that had brought enough reasons for being grumpy. You became my primary source for life as a psychologist "outside universities but still doing research", and similar to Jonas, maybe, in a few decades from now, (Dr.?) Luise might have changed other people's lives more than I did…

Roos: I should probably stop making jokes about your age; who knows if, in a few years, I need to ask you for a job…? Thanks for your energy and enthusiasm at various parties and outings, for your honest words when the boys could not stop talking about science ("boooriiing"), but also for your frequent interest in how I was doing ("Hallo Johannes wie geht's?") and telling me I shouldn't always work so much. I am curious whether, at some point, I will start telling other people about my famous friend Roos…

Kris, Dr. Adiasto-san, I still remember when we started calling each other Dr. after you got your position while I was still searching for one, and when I eventually heard I could stay in Nijmegen as well. Thanks for all your endless hospitality at Bloemenburgerhof, for your soberness when Julian and me derailed discussions a bit too much, and for being the self-assigned Methuselah of our group so the kids could always have some fun (and food and good drinks). Also, thanks for normalizing it to just fall asleep on parties, I came to appreciate it more and more.

Maartje, thanks for all the random discussions we had about academia at various moments (looking back, mostly while traveling)! Thanks for introducing me to certain bits of Dutch culture (music) I certainly was not into. Even though you left us quite soon, it was always nice you see you back in Nijmegen and hear your lovely German improve… I hope that, one day, I can send German post cards to Dr. Eijlander.

Tobi, you arrived last became sort-of the "puppy" in our group (though competing with Kris for who complains most about getting old). Thanks for all the "Tobi moments" (among them making me never arrive last) and for asking all the questions I also had as a German coming to the Netherlands for his Master's, but which I sometimes did not dare to ask ;-)

Gillis, man, you did not show up much to our meetings in the end, but I still fondly remember your spirit during our weekend outings (for example when you drove with Luise and me to Haltern) and it has always been a pleasure to have you around.

A few additional people that deserve mentioning: **Julia**, you were the first person I talked to on our first day of the Research Master, and I am so glad I met you! You got me hooked on the Zevenheuvelennacht, you searched for free furniture for my second home in Vossenveld, and you convinced me that we should do the "Intelligent Mobile Apps" course in the Artificial Intelligence Master (through which I only really got to know Ilse and Julian)! Your enthusiasm and "let's just do it" attitude have certainly helped me find my way during my first years in Nijmegen. Being probably the German the most immersed into Dutch culture I have met, it is a funny twist of fate that you will prospectively get a "German" PhD without the whole defense and paranymph business. I wish you all the best on your last steps towards Dr. Norget! Lastly, a special mention to our (inofficial) *StuSti-NL Stammtisch*: **Gloria, Yorick, Fabian**, it was great to meet some Germanz once in a while to gossip about Dutch politics, food, academia, and the general way of life. Thanks for occasionally distracting me from work (though Yorick and me like our work too much probably) and exploring the Netherlands together with me; I hope we stay in touch (maybe even meet at a Sommerakademie?!?)!

I would like to thank the **Studienstiftung des deutschen Volkes** (German National Academic Foundation) for continuous financial support throughout my studies—especially for my stays in Helsinki and New York City. Without this support, I would have probably not gone abroad (and never pursued an Research Master's program) and this thesis might never have happened.

Finally, for my **mother** and **brother**: How often have you asked me during my yearly visits around Christmas what my thesis was actually about and why people should care about this topic. Now, you have gotten a thick book that is unfortunately even in English! I suppose that I could have gotten you more piece of mind if I had just studied something else, started earning actual money after my Master's, and most certainly gotten back to Germany already years ago… Sorry for being a bit stubborn and seemingly "studying" forever, and thanks for letting me pursue the things I like. I hope you can now appreciate the (physical) "outcome" of these years of extra work, plus having a Dr. Johannes at home.

LIST OF PUBLICATIONS

PUBLISHED IN PEER-REVIEWED JOURNALS

Scholz, V., Hook, R.W., Rostami Kandroodi, M., **Algermissen, J.,** Ioannidis, K., Christmas, D., Valle, S. Robbins, T.W., Grant, J.E., Chamberlain, S.R., & den Ouden, H.E.M. (2022). Cortical dopamine reduces the impact of motivational biases governing automated behaviour. *Neuropsychopharmacology*. 10.1038/s41386-022-01291-8.

Algermissen, J., Swart, J.C., Scheeringa, R., Cools., R., & den Ouden, H.E.M. (2022). Striatal BOLD and midfrontal theta power express motivation for action. *Cerebral Cortex*, bhab391. doi: 10.1093/cercor/bhab391.

Wagenmakers, E.-J., Sarafoglou, A., Aarts, S., Albers, C., **Algermissen, J.,** ... & Aczel, B. (2021). Seven steps toward more transparency in statistical practice. *Nature Human Behavior*, 5(11), 1473-1480. doi: 10.1038/s41562-021-01211-8

Algermissen, J., Bijleveld., E. H., Jostmann, N. B., & Holland, R. W. (2019). Explore or reset? Pupil diameter transiently increases in self-chosen switches between cognitive labor and leisure in either direction *Cognitive, Affective, and Behavioral Neuroscience*, 19 (5), 1113-1128. doi: 10.3758/s13415-019-00727-x

Algermissen, J., & Mehler, D. A. M. (2018). May the power be with you: Are there highly powered studies in neuroscience, and how can we get more of them? *Journal of Neurophysiology*, 119 (6), 2114-2117. doi: 10.1152/jn.00765.2017

Koranyi, N., Grigutsch, L. A., **Algermissen, J.**, & Rothermund, K. (2017). Dissociating implicit wanting from implicit liking: Development and validation of the Wanting Implicit Association Test (W-IAT). *Journal of Behavior Therapy and Experimental Psychiatry*, 54, 165-169. doi: 10.1016/j.jbtep.2016.08.008

SUBMITTED

Šoškić, A., Ković, V., **Algermissen, J.**, ..., & Styles, S. (2023). EEGManyPipelines: ARTEMIS for ERP: Agreed Reporting Template for EEG Methodology - International Standard for documenting studies on Event-Related Potentials. *PsyArXiv*.

Trübutschek, D., Yang, Y.-F., Giannelli, C., ..., **Algermissen, J.**, ..., & Nilsonne, G. (2022). EEGManyPipelines: A large-scale grass-root multi-analyst study on EEG analysis practices in the wild. *MetaArXiv*.

Kovacs, M., Jaquiery, M., **Algermissen, J.**, ..., & Aczel, B. (2022). Lab manuals for effcient and high quality science in a happy and safe work environment. *MetaArXiv*.

van Dongen, N., Finnemann, A., de Ron, J., Tiolhin, L., Wang, S., Algermissen, J., ..., & Borsboom, D. (2022). Many Modelers. *PsyArXiv*.

Govaart, G. H., ..., **Algermissen, J.**, ..., & Paul, M. (2022). EEG ERP Preregistration Template. *MetaArXiv*.

Algermissen, J., & den Ouden, H.E.M. (2022). Goal-directed recruitment of Pavlovian biases through selective visual attention. *BioRxiv*.

Algermissen, J., Swart, J.C., Scheeringa, R., Cools., R., & den Ouden, H.E.M. (2021). Biased credit assignment in motivational learning biases arises through prefrontal influences on striatal learning. In revision at Nature Communications. *BioRxiv*.

Algermissen, J., Woyke, I., Niermann, H.C., Tyborowska, A., Roelofs, K., & Figner, B. (2021). Stress decreases adolescent patience in intertemporal choice.

CHAPTERS IN BOOKS ON SCIENTIFIC OUTREACH

den Ouden, H.E.M., & **Algermissen, J.** (2020). Automatisch gedrag. In: van-Baren-Nawrocka, J., Dekker, S., & de Boer, M. (Ed.). Wetenschappelijke doorbraken de klas in! Sport, Slimme computers en Automatisch gedrag. Nijmegen, the Netherlands: Wetenschapsknooppunt Radboud Universiteit.

ABOUT THE AUTHOR

Johannes Algermissen was born on the 30th of October, 1991, in Hildesheim, Germany. He graduated from the Bischöfliches Gymnasium Josephinum in Hildesheim in 2010. Afterwards, he performed nine months of civil service at a center for reintegrating patients with mental disabilities and psychiatric disorders.

In 2011, he started studying psychology (and in 2013 additionally also philosophy with mathematics as a minor) at the Friedrich-Schiller-Universität Jena in Jena, Germany. Throughout his studies, Johannes was supported by a scholarship from the German National Academic Foundation. He finished both Bachelor degrees in 2014 and 2015, respectively, followed by an Erasmus exchange year at the Helsingin Yliopisto in Helsinki, Finland, where he took courses on social psychology, philosophy, musicology, semiotics, Finnish language and culture, and statistics.

In 2015, he enrolled in the Research Master in Behavioural Science at Radboud Universiteit in Nijmegen, the Netherlands, graduating summa cum laude in 2017. In his master thesis, supervised by Rob Holland, Erik Bijleveld, and Nils Jostmann, he investigated whether the decision to transition from cognitive labor to leisure and back can predicted from pupil diameter as an index of the state of the noradrenergic system. Also, during his Master's, he spent two months at Columbia University, New York City, USA, working with Bernd Figner and Eric J. Johnson on the relationship between intertemporal choice and procrastination.

In 2017, after his masters, he started as a PhD candidate working on "Decision making/ computational psychiatry" supervised by Hanneke den Ouden at the Donders Centre for Cognition at the Radboud Universiteit. He used various methods including simultaneous functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), magnetoencephalography (MEG), eye-tracking, and computational modeling of behavior to study the origin and control of Pavlovian biases in learning. During his PhD, he also contributed to various projects aiming at improving psychology and neuroscience, among them the Open Science Community Nijmegen (OSCN), the EEG ManyPipelines Project (EMP), and the ARTEM-IS Project.

Since September 2022, Johannes works as a postdoctoral researcher supervised by Miriam Klein-Flügge at the Department of Psychiatry, University of Oxford in Oxford, United Kingdom. His current research aims to understand the cognitive and neural mechanisms underlying foraging-style decisions in humans, for which he uses large-scale online testing, high-field functional magnetic resonance imaging (fMRI), and transcranial ultrasonic stimulation (TUS).

RESEARCH DATA MANAGEMENT

This research followed the applicable laws and ethical guidelines. Research Data Management was conducted according to the FAIR principles. The paragraphs below specify in detail how this was achieved.

ETHICS

This thesis is based on the results of human studies, which were conducted in accordance with the principles of the Declaration of Helsinki. The studies reported in Chapters 2, 3, and 5 were conducted under the ethical approval of the CMO Arnhem/ Nijmegen ("Imaging Human Cognition", CMO 2014/288). The studies reported in Chapter 4 were conducted after the Ethical Committee of the Faculty of Social Sciences (ECSS) had given a positive advice to conduct these studies to the Dean of the Faculty, who formally approved the conduct of these studies (number ECSW-2018-171). This research was funded by the Dutch Research Council (NWO).

FINDABLE ACCESSIBLE

The table below details where the data and research documentation for each chapter can be found on the Donders Repository (DR). All data archived as a Data Sharing Collection remain available for at least 10 years after termination of the studies.

Chapter	DAC	RDC	DSC	DSC License
2	DAC_3017042.02_747	RDC_3017042.02_137	DSC_3017042.02_604 DSC_3017042.02_620	RU-DI-HD-1.0
3	DAC_3017042.02_747	RDC_3017042.02_137	DSC_3017042.02_604 DSC_3017042.02_796	RU-DI-HD-1.0
4	DAC_2020.00093_358	RDC_2020.00093_303	DSC_2020.00093_989	MIT
5	DAC_3025004.01_772	RDC_3025004.01_697		RU-DI-NH-1.0

DAC = Data Acquisition Collection, RDC = Research Documentation Collection, DSC = Data Sharing Collection

For chapters 2, 3, 4 and 5, research data have also been stored on the Donders project drive. These data were accessible to all members involved in the project. After the finalization of the chapters 2–4, data were removed from the Donders project drives. The publication of chapter 5 is still in preparation. Upon publication, the respective will be made publicly available under the RU-DI-HD-1.0 license as it contains potentially identifiable data.

Informed consent was obtained on paper following the Centre procedure. The forms are archived in the central archive of the Centre for 10 years after termination of the studies.

INTEROPERABLE, REUSABLE

The raw data are stored in the DAC in their original form. For RDC and DSC long-lived file formats (e.g. .csv, .nii, ,.edf, .mat) have been used ensuring that data remains usable in the future. The data of the RDC and DSC are organized with concomitant README files. Results are reproducible by providing a description of the experimental setup, raw data (DAC and DSC), analysis scripts or pipelines (RDC and DSC). Also, the used software including version numbers is specified.

PRIVACY

The privacy of the participants in this thesis has been warranted using random individual subject codes. A pseudonymization key linked this random code with the personal data. This pseudonymization key was stored on a network drive that was only accessible to members of the project who needed access to it because of their role within the project. The pseudonymization key was stored separately from the research data. The pseudonymization keys of chapters 2, 3, 4 were destroyed within one month after finalization of these projects. The key of chapter 5 is still stored on a dedicated restricted network drive and will be destroyed within one month after finalization. Data in chapter 4 are not identifiable and shared without restrictions. MRI-data of chapters 2, 3, and 5 are/ will be defaced, and are/ will be shared under the restricted license RU-DI-HD-1.0, which provides extra statements for the protection of the identity of the participants.

DONDERS GRADUATE SCHOOL FOR COGNITIVE NEUROSCIENCE

For a successful research Institute, it is vital to train the next generation of young scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School for Cognitive Neuroscience (DGCN), which was officially recognised as a national graduate school in 2009. The Graduate School covers training at both Master's and PhD level and provides an excellent educational context fully aligned with the research programme of the Donders Institute.

The school successfully attracts highly talented national and international students in biology, physics, psycholinguistics, psychology, behavioral science, medicine and related disciplines. Selective admission and assessment centers guarantee the enrolment of the best and most motivated students.

The DGCN tracks the career of PhD graduates carefully. More than 50% of PhD alumni show a continuation in academia with postdoc positions at top institutes worldwide, e.g. Stanford University, University of Oxford, University of Cambridge, UCL London, MPI Leipzig, Hanyang University in South Korea, NTNU Norway, University of Illinois, North Western University, Northeastern University in Boston, ETH Zürich, University of Vienna etc.. Positions outside academia spread among the following sectors: specialists in a medical environment, mainly in genetics, geriatrics, psychiatry and neurology. Specialists in a psychological environment, e.g. as specialist in neuropsychology, psychological diagnostics or therapy. Positions in higher education as coordinators or lecturers. A smaller percentage enters business as research consultants, analysts or head of research and development. Fewer graduates stay in a research environment as lab coordinators, technical support or policy advisors. Upcoming possibilities are positions in the IT sector and management position in pharmaceutical industry. In general, the PhDs graduates almost invariably continue with high-quality positions that play an important role in our knowledge economy.

For more information on the DGCN as well as past and upcoming defenses please visit:

https://www.ru.nl/donders/talent-education/graduate-school/

DONDERS INSTITUTE

Radboud University Radboudumc