Aus der Klinik für Psychiatrie und Psychotherapie
der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

DISSERTATION

“The Neural Computation of Value in Decision-Making”
“Die Neuronale Berechnung von Bewertungen in der Entscheidungsfindung”

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Amadeus Magrabi

Datum der Promotion: 23.03.2024

Table of Contents

Figures

Abbreviations

ACC	anterior cingulate cortex
AD	alcohol dependence
BOLD	blood oxygenation level dependent
dlPFC	dorsolateral prefrontal cortex
dmPFC	dorsomedial prefrontal cortex
fMRI	functional magnetic resonance imaging
GLM	general linear model
IT	infero-temporal cortex
ITI	inter-stimulus interval
M	mean
MNI	Montreal Neurological Institute
PCC	posterior cingulate cortex
SD	standard deviation
SEM	standard error of the mean
vmPFC	ventromedial prefrontal cortex
vStr	ventral striatum

Zusammenfassung

Hintergrund: Die Entscheidungsfindung spielt eine zentrale Rolle in unserem Leben und die Qualität unserer Entscheidungen hat einen großen Einfluss auf unsere Zukunft. Um den Entscheidungsprozess besser zu verstehen, hat sich die kognitive Neurowissenschaft bisher auf das Konzept des Entscheidungswerts konzentriert. Einer der wichtigsten Befunde ist, dass der Gesamtwert einer Entscheidungsoption mit der neuronalen Aktivität im ventromedialen präfrontalen Kortex (vmPFC) korreliert. Es ist jedoch immer noch unklar, wie genau dieses Signal im Gehirn berechnet wird.

Methodik: Um diesen Prozess besser zu verstehen, wurden zwei Studien durchgeführt, welche die funktionelle Magnetresonanztomographie nutzen. Die erste Studie basiert auf der Idee, dass der Gesamtwert einer Entscheidungsoption sich in der Regel aus den Werten von verschiedenen Attributen zusammensetzt. Insbesondere sollte untersucht werden, ob die Werte von Attributen im vmPFC berechnet werden, wie Gesamtbewertungen, oder in abgegrenzten, Attribut-spezifischen Regionen. Das Experiment bestand aus einer Entscheidungsaufgabe mit abstrakten Stimuli, die mit monetären Belohnungen verbunden waren und hinsichtlich der Attribute Bewegung (assoziiert mit der Hirnregion V5) und Farbe (Hirnregion V4) variierten. Die zweite Studie untersuchte, wie die Berechnung des Entscheidungswerts bei Patienten mit Alkoholabhängigkeit beeinflusst wird. In der experimentellen Aufgabe mussten Entscheidungen zu verschiedenen Geldangeboten unter Unsicherheit gemacht werden.

Ergebnisse: In der ersten Studie konnte repliziert werden, dass Gesamtwerte mit der Aktivität im vmPFC korrelieren. Allerdings wurde nicht bestätigt, dass Attributwerte systematisch in Attribut-spezifischen Hirnregionen repräsentiert werden. Stattdessen waren Attributwerte mit Aktivitäten im posterioren cingulären Kortex, ventralen Striatum und posterioren inferioren temporalen Gyrus verbunden. In der zweiten Studie gab es keine Unterschiede in der wertbezogenen neuronalen Aktivität oder im Verhalten zwischen gesunden Probanden und Patienten, aber die Patienten zeigten während modellbasierter Entscheidungsprozesse eine geringere Aktivierung des Nucleus caudatus.

Diskussion: Die Ergebnisse deuten darauf hin, dass die Charakterisierung der Funktionen von isolierten Hirnregionen für die Beschreibung von neuronalen Entscheidungsprozessen weniger geeignet ist und Entscheidungsprozesse stattdessen innerhalb dynamischer Netzwerke berechnet werden. Die Ergebnisse des zweiten Experiments zeigen au-

ßerdem, dass Krankheiten wie Alkoholabhängigkeit die neuronalen Prozesse von Entscheidungen auch dann beeinflussen können, wenn keine alkoholbezogenen Stimuli vorliegen. Dies hatte in der Studie allerdings keinen signifikanten Einfluss auf die Qualität der Entscheidungen, was darauf hindeutet, dass die dynamischen Eigenschaften von neuronalen Entscheidungsnetzwerken Kompensationsmechanismen ermöglichen können.

Abstract

Background: Decision-making plays a key role in the human experience and the quality of our choices has a fundamental impact on our future. To better understand the decision-making process, the field of decision neuroscience has focused on the concept of decision value and found that the overall value of a choice option correlates with neural activity in ventromedial prefrontal cortex (vmPFC). However, it is still unclear how exactly this neural signal is computed.

Methods: Two studies using functional magnetic resonance imaging (fMRI) were conducted to investigate this process from different angles. The first study is based on the idea that the overall value of a choice option is typically determined by evaluating and integrating different attributes. In particular, the goal was to investigate whether attribute values are computed in vmPFC, like overall values, or whether they are computed in distinct attribute-specific regions. The experiment consisted of a choice task with abstract stimuli, which were associated with monetary rewards and varied with respect to the attributes motion and color (associated with the brain regions V5 and V4, respectively). The second study investigated how the computation of decision value is affected in patients suffering from alcohol dependence. The task required the evaluation of monetary offers with respect to dynamically changing constraints and different levels of uncertainty.

Results: The first study could replicate the finding that overall values correlate with activity in vmPFC. However, I did not find that attribute values were systematically represented in attribute-specific regions. Instead, attribute values were associated with activity in the posterior cingulate cortex, ventral striatum, and posterior inferior temporal gyrus. In the second study, there were no group differences in value-related neural activity or task performance, but patients showed lower activation associated with model-based decision processes in the caudate nucleus.

Discussion: The results support the idea that the neural mechanisms for choices should be studied from the perspective of neural networks instead of investigating the functional properties of brain regions in isolation. Further, the findings of the second experiment demonstrate that clinical conditions like alcohol dependence can affect regional activations related to choices even in the absence of alcohol-related stimuli. However, this did not significantly affect behavioral task performance in the choice task, which suggests that the dynamic properties of decision-making networks can allow for compensatory mechanisms.

1 Background

Every day, we are faced with a virtually countless number of decisions. For the most part, they happen fast and seemingly automatic, such as the choice between taking the stairs or the elevator, when to drink water, or which chair to sit on. But from time to time, more important decisions must be made, such as choosing a career path or deciding to have children. These choices can require long periods of time, in which we carefully think about our available options, consider short- as well as long-term consequences, and evaluate them with respect to their advantages and disadvantages. We often invest a lot of time and effort into this process, because it can have a long-lasting impact on our future.

Apart from its personal significance, decision-making also plays an important role in scientific research and receives wide interest from a variety of disciplines (Glimcher & Fehr, 2013; Kahneman, 2011). In the early days of decision research, psychologists were primarily interested in conscious and unconscious factors that can influence decisions, philosophers in the properties of free and ethical choices, biologists in the neural basis of decisions and its evolutionary development, economists in the prediction of consumer decisions, and computer scientists in the construction of artificial systems that exhibit intelligent choice behavior. Today, traditional borders between disciplines are collapsing, and researchers engage in interdisciplinary collaboration to develop a more comprehensive account of decision-making. This endeavor has led to the emergence of new research areas, such as the field of neuroeconomics, which has the goal to integrate economic models of decision-making with findings from neuroscience and psychology.

In this chapter, I will outline the conceptual framework that is being used in decision research, introduce typical experimental paradigms and prominent findings in decision neuroscience, and illustrate how my thesis fits into this context.

1.1 Conceptual Framework

In short, decision-making is the process that identifies the most valuable option from a set of alternatives. Conceptually, it has been proven difficult to clearly delineate the precise processes that decision-making consists of. This is in part because decision-making is intricately connected to various other cognitive functions. Choices are not made in isolation, but heavily depend on functions like perception (what information is currently avail-

able?), memory (have similar choices been made in the past?), or reasoning (which consequences could the different choice options result in?). It has therefore been difficult to separate other cognitive functions from computations that can be considered as “pure” decision processes.

One account proposes that decision-making involves five different stages (Fig. 1; Rangel et al., 2008): In the *representation* stage, the structure of the choice problem is defined, including the current state of the individual, the context of the situation, and the set of available options. In the *valuation* stage, each option is evaluated based on the desirability of the respective consequences and the probability of their occurrence. Crucially, this does not necessarily lead to the objective value, but merely the subjective value, which is determined relative to the perspective of the individual and can be prone to errors based on false information. Finally, in the *action selection* stage, the option with the highest expected value is selected as the outcome of the decision, since it is estimated to deliver the largest benefits. After the decision, the experienced consequences are compared with previous expectations in the *outcome evaluation* stage, and potential mismatches are being corrected for in the *learning* stage to improve future decisions. While the separation of these stages is certainly not clear-cut and can be questioned, it nevertheless provides a useful framework to structure research questions on decision-making.

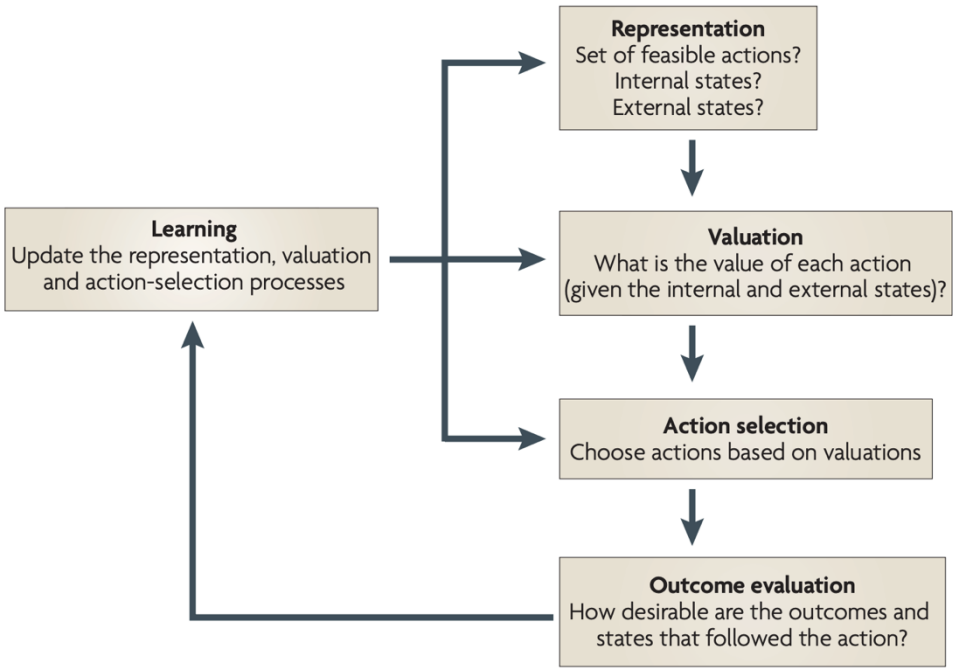

Figure 1. *Stages of Decision-Making. Figure adapted from Rangel et al. (2008).*

1.2 Experimental Paradigms

Studies on decision-making have made use of a range of experimental paradigms, which are generally divided into the categories of *value-based*, *social*, and *perceptual* decision tasks. In this section, I will introduce each of them by describing typical task structures.

Value-based decision tasks are designed to investigate the valuation stage of decision-making. A classical paradigm is the so-called *Becker-DeGroot-Marschak auction* (Becker et al., 1964; Plassmann et al., 2010), in which participants are given a monetary budget and asked to indicate how much they are willing to pay for presented products (such as food or clothing items). Afterwards, their bid competes against a computer-generated bid, which is computed as a random number within the budget of the participant, and participants can only buy the products for their bid if it is higher than the computer-generated bid. Hence, the higher the bid of the participant, the more likely it is that they can buy the product. Through this procedure, the task has shown to provide a reliable estimate of the subjective value that participants associate with specific products, which can be used to investigate how value impacts choice behavior.

In social decision tasks, participants are faced with choices that involve other people. Paradigms are specifically designed such that participants must take different perspectives into account and consider the expectations, intentions, and emotions of others. Economists have investigated these types of decisions under the heading of *game theory* (Camerer, 2003; von Neumann & Morgenstern, 1947). A typical paradigm is the so-called *ultimatum game* (Güth et al., 1982; Sanfey et al., 2003), in which a player A receives a certain amount of money that he can share with another player B. Player A is free to choose how much money he wants to keep for himself and how much he wants to share. But crucially, player B has the option to decline an offer suggested by player A, in which case none of the two players receive any money. Traditional economic theories would predict that player B will accept any offer, since even the smallest offer is a gain for the player compared to not receiving any money. However, it has been shown that players do not attempt to maximize their profit in this way and instead frequently reject offers when they are considered unfair (Nowak et al., 2000).

Finally, perceptual decision tasks are designed to study choices based on sensory information. In the *random-dot motion discrimination task* (Heekeren et al., 2006; Newsome et al., 1989), participants are required to estimate the dominant motion direction within a set of moving dots. A subset of these dots exhibits random motion, whereas

another subset coherently moves into the same direction. If the number of coherently moving dots is low, then the decision is more difficult, because the sensory evidence is low. The paradigm therefore provides an elegant way to manipulate levels of sensory evidence on a continuous scale and investigate how choices are affected.

1.3 Major Findings in Decision Neuroscience

In the following, I will outline some of the milestones in decision neuroscience that are particularly relevant for this thesis. The goal of this section is not to discuss the specific experiments in detail, but merely to sketch the research landscape and provide a general overview of the major findings.

1.3.1 Neural Representation of Value

The valuation stage of decision-making has arguably received the most attention in the literature so far. One of the most robust findings is that neural activity in *ventromedial prefrontal cortex* (vmPFC) correlates with the subjective value of a choice option when it is considered (Bartra et al., 2013; Kable & Glimcher, 2007; Lee et al., 2021; Pelletier et al., 2021; Rangel & Clithero, 2013). This has been demonstrated for a variety of stimuli in neuroimaging research, such as food items (Hare et al., 2009; Plassmann et al., 2007), monetary gambles (Chib et al., 2009; Levy et al., 2010; Tom et al., 2007), clothes (Lim et al., 2013), charity donations (Hare et al., 2010), or wine (Plassmann et al., 2008). Further evidence for this finding was observed in clinical studies on vmPFC lesions (Fellows & Farah, 2007; Gläscher et al., 2012) as well as electrophysiological studies in primates (Leathers & Olson, 2012; Padoa-Schioppa, 2009; Wallis & Miller, 2003).

Based on the consistency of these findings, decision neuroscientists have hypothesized that vmPFC could be an area where multiple sources of evidence converge and get integrated into an overall value signal that drives decision-making (Hare et al., 2009; Levy & Glimcher, 2012; Padoa-Schioppa & Cai, 2011). According to this theory, different parts of a decision problem could be processed in a range of different brain regions, but eventually, all the relevant information is propagated to the vmPFC, where it is combined to a unified value representation. Signals in vmPFC could thus encode a summary of the available evidence and serve as an indicator of the overall value of a choice option. Apart from vmPFC, value signals are also frequently found in the *ventral striatum* (vStr) and the *posterior cingulate cortex* (PCC; Bartra et al., 2013; Clithero & Rangel, 2014). Up to this

point, it is still unclear how exactly these regions interact in the process of value integration and what their differential roles are.

1.3.2 Model-Based vs. Model-Free Decisions

While it is a challenging scientific endeavor to isolate *where* values are encoded in the brain, it is even more challenging to identify *how* they are computed. Which computational strategy does the brain use to solve this task and what are the sub-processes that lead up to an estimate of decision value?

One way this question has been investigated in the literature is based on a distinction between *model-based* and *model-free* decision-making (Daw et al., 2011; Dolan & Dayan, 2013; Gläscher et al., 2010). Model-based choices are based on an explicit cognitive model of the environment, which is used to form predictions about potential consequences of decisions, their likelihood, and their subjective value. This type of decision-making is prospective, goal-directed, and can flexibly adapt to changing circumstances, but typically requires a considerable amount of time and effort. In contrast, model-free choices are based on a relatively simple set of rules to form habitual connections between specific stimuli and behavioral responses. Here, we do not take the full complexity of a choice problem into account and instead apply simplified heuristics that have been shown to produce satisfactory choices in similar situations in the past. Compared to model-based choices, model-free decisions are retrospective, automatic, and fast. But we must accept the risk of oversimplifying the situation and potentially overlooking important details.

For instance, when we decide whether we want to go to a specific restaurant, a model-free mechanism could evaluate the restaurant simply by how well the food tasted on average during our last visits. In contrast, a model-based mechanism could be based on a complex cognitive model that considers recommendations from friends, the specific cook that is preparing the food today, or the restaurant's online ratings. Each of these attributes is first evaluated separately and then finally integrated to a combined overall value to determine optimal choices (Hunt et al., 2014; Kahnt et al., 2011).

A particularly influential mechanism that has been proposed for the computation of value in model-free decisions is known as *reinforcement learning* (Dayan & Niv, 2008; Sutton & Barto, 1998). In reinforcement learning, the *expected value* of a choice option (pre-choice) is compared with the *experienced value* (post-choice). If the two values turned out to be different, the prediction of the expected value is changed for future choices based on the magnitude of the difference between expected and experienced

value (called *prediction error*). The larger the prediction error, the stronger is the change of the expected value towards the experienced value. In the classical formulation, expected values are updated via the following learning rule (known as *temporal difference learning*):

$$V_{t+1} = V_t + \alpha \times \delta_t$$
$$\delta_t = R_t - V_t$$

where V is the expected value, R the experienced value (reward), δ the prediction error, and t a time variable. Further, $0 \leq \alpha \leq 1$ defines the *learning rate* that controls how strongly the expected value should be changed (the expected value would not change at all with $\alpha = 0$, whereas it would be completely adapted to the experienced value with $\alpha = 1$).

One experimental paradigm that is used to study model-based choices is the so-called *two-step* decision task (Deserno et al., 2015), in which participants are required to make two sequential choices between pairs of abstract stimuli to obtain monetary rewards. For example, one option might offer a higher potential reward but also a higher risk of failure, while the other option might offer a lower potential reward but also a lower risk of failure. Since the transition probabilities of the first stage are explicitly instructed, participants can strategically use this knowledge of the task structure to increase their odds of reaching stimuli with a high expected value. By fitting a computational model to the choice behavior, researchers can then quantify to what extent participants were applying a model-based compared to a model-free approach.

It has been observed that people often use a mixture of both strategies, but that there is individual variation in the relative emphasis placed on each strategy (Daw et al., 2011; Gläscher et al., 2010; Smittenaar et al., 2013). On the neural level, experiments have shown that the striatum plays a crucial role for these computations. Model-based decision processes have primarily been associated with activity in the caudate nucleus/ dorsomedial striatum (in humans/rodents), whereas model-free computations have been linked to the putamen/dorsolateral striatum (Dolan & Dayan, 2013; Gahnstrom & Spiers, 2020; Geerts et al., 2020; Sharpe et al., 2019; Wunderlich et al., 2012). Further, in reinforcement learning models of model-free choices, studies have demonstrated that the ventral part of the striatum is associated with the magnitude of prediction errors (Hare et

al., 2008; McClure et al., 2003; O'Doherty et al., 2003), with a particular involvement of dopaminergic neurons (Montague et al., 1996; Schonberg et al., 2007).

1.4 Goal of Thesis

As noted above, the majority of previous studies in decision neuroscience were concerned with the question of where decision values are represented in the brain. This has led to the discovery of robust findings, in particular the correlation between overall decision values and vmPFC activation. However, little is known about the processes that generate this signal. The main goal of this thesis is to shed further light on the neural computations that lead to decision value signals in the brain, and to explore the factors that can influence these signals in the decision-making process. To this end, two studies using functional magnetic resonance imaging (fMRI) were conducted.

1.4.1 Study 1: Neural Integration of Attributes in Decision-Making

The first fMRI experiment aimed to investigate how the brain evaluates and integrates decision attributes. In most situations, the value of a choice option could be determined by evaluating a range of individual attributes separately and then combining these evaluations to arrive at a unified overall value (O'Doherty et al., 2021; Suzuki, 2022). For example, the overall value of a car can depend on how we evaluate its size, speed, or color, and how much subjective weight we assign to each of these attributes. In the experiment, I used a decision task in which monetary values were attached to different perceptual attributes of an artificial stimulus, to examine how attribute values are computed in the brain. The results could then be compared with the neural correlates of 1) attribute identification, which must occur before attribute valuation, and 2) overall value computation, which must occur after attribute valuation.

1.4.2 Study 2: Neural Basis of Decisions in Alcohol-Dependent Patients

The second fMRI experiment examined neural differences in decision processes between alcohol-dependent (AD) patients and healthy control subjects. By studying the functional properties of psychiatric disorders and how they affect cognitive functions, we can not only gain insights for clinical applications, but also improve our understanding of cognitive functions in general. Alcohol addiction is a highly prevalent disease that severely affects the physical and psychological health of patients, but its cause as well as its neural basis

is not well understood. Since AD patients compulsively choose to consume alcoholic stimuli that provide short-term rewards, but fail to make decisions that take the negative long-term consequences of sustained alcohol consumption into account, impairments of model-based or value-related decision processes could be crucial factors. Previous studies showed neural differences between patients and controls when tasks required decisions about alcoholic stimuli (Beck et al., 2012; Schad et al., 2019), but it is unclear to what extent these results generalize to non-alcoholic stimuli as well. To investigate this, a new experimental paradigm was designed, which included decisions about monetary stimuli and specifically relied on the ability to compute model-based decisions and dynamically adapt to changing circumstances.

2 Methods

This section summarizes the most important parts of the experimental methods that were used to conduct the experiments. More details can be found in the attached original publications.

2.1 Study 1: Neural Integration of Attributes in Decision-Making

Twenty-five subjects (14 female; mean age 28.1 ± 4 SD) participated in the fMRI experiment. Participants had to complete a variation of the *random dot task* (e.g., Gold & Shadlen, 2007), in which moving dots varied with respect to two attributes, motion direction and color (with 6 different levels each). Each attribute level was associated with a monetary value from the set {-15, -10, -5, 5, 10, 15} (€ cents). The task of the participants was to identify the attribute levels of a stimulus, remember the corresponding monetary values, and combine them, to decide whether they want to accept or reject a stimulus. For example, if the dots are blue and move to the right, and this would correspond to -15 and +5 cents respectively, then the combined value of the stimulus is -10 cents and should therefore be rejected to avoid a monetary loss. The task consisted of six experimental blocks with 30 trials each. Associations between attribute levels and monetary values were learned separately a few days before scanning (mean days 2 ± 0.4 SD).

The task design allowed to investigate the brain regions responsible for the processing of attribute values and compare them with 1) the regions integrating individual attribute values to combined overall values, 2) the regions associated with the identification and perceptual processing of an attribute (independent of decision value), namely area V4 (McKeefry and Zeki, 1997) for color and area V5 for motion (Watson et al., 1993; Gallivan et al., 2018), and 3) the regions involved in processing attribute salience (the subjective importance of a stimulus that guides attention, corresponding to absolute attribute values; Kahnt and Tobler, 2013; Litt et al., 2011; Maunsell, 2004; Zhang et al., 2017).

To analyze the fMRI data, two *general linear models* (GLM) were computed. GLM1 included a regressor R1 of all trials with correct choices (in which participants accepted positive and rejected negative overall values) and five linear parametric modulators: P1) motion value, P2) color value, P3) motion salience (absolute motion value), P4) color salience (absolute color value), and P5) the absolute difference between motion and color

value. Further, regressors of no interest were included to increase the signal-to-noise ratio of the model, namely regressor R2 including trials with incorrect choices trials and six movement regressors R3-R8 from the MRI image realignment procedure.

Whereas GLM1 was targeted at analyzing individual attribute properties, GLM2 was designed to analyze overall value and overall salience. In contrast to GLM1, it only included two parametric modulators: P1) overall value (sum of motion and color value) and P2) overall salience (absolute overall value). These modulators were not part of GLM1, because the significant correlations between individual and overall attribute values/salience would confound the results of the analyses.

Apart from analyzing whole-brain effects ($p_{unc} < 0.001$, cluster extent threshold $k_E = 15$ voxels, $p_{FWE} < 0.05$), region-of-interest (ROI) analyses were also conducted to analyze whether areas V4 and V5 were involved in computing color and motion value instead of merely attribute identification and perceptual processing. The precise locations of color- and motion-sensitive regions were estimated for each participant with the help of independent localizer tasks. Mean beta weights were extracted from parametric modulators P1 (motion value) and P2 (color value) of GLM1 and analyzed for significance via a repeated-measures analysis of variance (rm-ANOVA) with the factors attribute value (motion/color), region (V5/V4), and hemisphere (left/right).

2.2 Study 2: Neural Basis of Decisions in Alcohol-Dependent Patients

32 detoxified alcohol-dependent (AD) patients (10 female; mean age 46.5 ± 8.9 SD) and 32 healthy control subjects (9 female; mean age 38.9 ± 10.5 SD) participated in the experiment. Patients were diagnosed according to the DSM-IV and ICD-10 (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I); First & Gibbon, 2004). Group comparisons revealed a significant difference with respect to age ($T_{64} = -3.13$, $p < 0.01$), which was controlled for by including age as a covariate in statistical analyses of group differences. In addition to that, patients showed increase smoking behavior (indicated by *pack years of cigarette consumption*; $T_{64} = -2.9$, $p < 0.01$), which is a common finding in studies on AD patients (Batel et al., 1995). But since cigarette consumption was significantly correlated with *lifetime alcohol intake* in AD patients ($r = 0.52$, $p < 0.01$) and can therefore interfere with variance related to alcohol dependence, covariates relating to smoking behavior were not included in my analyses.

In the MRI scanner, participants had to complete a sequential decision-making task in which they had to accept or reject monetary offers between 1 and 99 cents (€). Crucially, for each experimental block of 20 offers, participants were only allowed to accept a maximum of 5 offers. To make optimal choices and maximize the probability of accepting only the highest offers in a block, participants thus had to consider three factors: 1) the value of the current offer, 2) the number of offers that can still be accepted before reaching the limit, and 3) the number of offers that are remaining in the current block. These parameters were included in a computational decision model (Economides et al., 2014) and fitted to the choices of each participant. The model estimates the expected value of accepting an offer V_A by comparing the monetary offer value R with a model threshold M:

$$V_A = R - M$$

Accordingly, a high model threshold indicates that accepting an offer has a low expected value. The model threshold is calculated with this formula:

$$M = c_1 + a \times c_2 - o \times c_3$$

with c_1 being a constant threshold, a the number of offers accepted previously, o the offer index, and c_2 and c_3 as weight parameters for a and o, respectively. In this formulation, the model threshold increases linearly when a increases (since accept choices should be more conservative when many offers have already been accepted), and the model threshold decreases linearly when o increases (since accept choices should be more liberal when the end of a block is near). Finally, the expected value of accepting V_A is used to compute the probability of accepting P_A via a sigmoid function:

$$P_A = \frac{1}{1 + \exp(-\tau \times V_A)}$$

with τ governing the slope of the probability distribution. Parameter estimates of the model were then used as parametric modulators in the analysis of the fMRI data to identify brain regions that compute model-based decision processes and to test for putative differences

between AD patients and controls. This was modelled as a regressor representing the decision phase of valid trials with parametric modulators for the offer value and model threshold. The remaining regressors were of no interest and represented invalid trials (when participants failed to make a response within the time limit), the response phase, the feedback phase, and pauses between experimental blocks.

Based on previous studies, I defined regions of interest (ROIs) and hypothesized that AD patients would show 1) an increased representation of decision value in vmPFC, 2) a decreased representation of model-based decision processes in caudate nucleus.

3 Results

3.1 Study 1: Neural Integration of Attributes in Decision-Making

Participants showed high accuracies in the main decision task (mean 87.5 ± 6.8 % SD; one-sample t-test against chance level, $t_{24} = 27.56$, $p < 0.001$) and responded on average 873 ± 157 ms (SD) after trial onset.

Analyses of fMRI data revealed differential neural correlates for the attribute values of motion and color, respectively (see GLM1; section 2.1). For motion value, activity during correct decision trials showed a significant positive parametric modulation in regions including the posterior cingulate cortex (PCC) and left posterior inferior temporal gyrus (PIT; Fig. 2A), whereas a significant positive parametric modulation by color value was observed in ventral striatum and PCC (anterior to the PCC cluster for motion value; Fig. 2B). However, a direct comparison of motion- and color-related parametric effects using paired t-tests did not reveal significant differences at the whole-brain level. To further investigate this relationship, these regions were used as post-hoc ROIs (the PIT and PCC cluster of the motion value contrast, and the ventral striatum and PCC cluster of the color value contrast, thresholded at $p_{unc} = 0.001$) and mean beta values for motion and color value within these ROIs were compared via paired t-tests (Bonferroni-corrected p-value criterion of 0.05/4 = 0.0125). Consistent with whole-brain results, the analysis did not reveal significant differences between motion- and color-related parametric effects (PIT: $t(24) = 1.5$, $p = 0.15$; PCC_{motion}: $t(24) = 0.4$, $p = 0.71$; PCC_{color}: $t(24) = -2.4$, $p = 0.02$; vStr: $t(24) = -0.7$, $p = 0.5$). Although these findings do not definitively refute the existence of attribute-specific valuation, they imply that the calculation of motion and color value does not appear to involve distinct, attribute-specific valuation modules in this study.

Significant positive modulation of task-related activity by the stimulus' overall value (combined values; GLM2) were found in regions including left dorsolateral prefrontal cortex (dlPFC) and vmPFC (Fig. 2C). Further, the absolute difference between motion and color values was used as a variable in GLM1 to identify comparator regions that estimate differences between attribute values. Here, I observed a significant positive modulation of task-related hemodynamic activity within the dorsomedial prefrontal cortex (dmPFC; Fig. 2D).

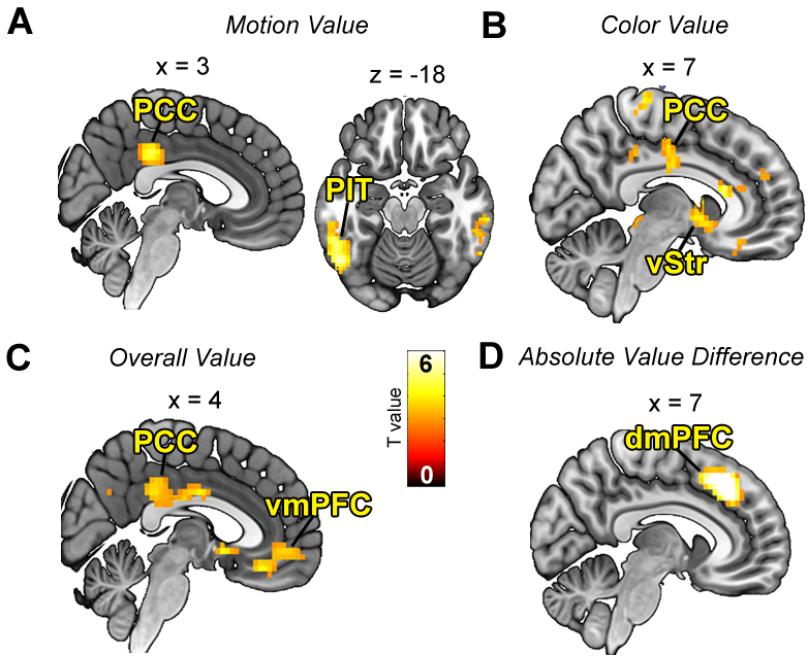

Figure 2. *Brain regions showing significant activations at the group level for A) motion value (GLM1), B) color value (GLM1), C) overall value (GLM2), and D) absolute attribute value differences (GLM1). For illustration purposes, t-maps (from second-level one-sample t-tests on parameter estimates of respective parametric modulators) are thresholded at $p_{unc} < 0.001$ with a cluster extent threshold of $k_E = 15$. Labeled clusters survive cluster-level FWE-correction at $p_{FWE} < 0.05$. Figure adapted from Magrabi et al. (2022a).*

ROI analyses were conducted to test whether regions specialized in attribute identification (i.e., V4 for color and V5 for motion) are also involved in processing the respective attribute values of the stimuli. For this purpose, beta estimates from parametric modulations by motion and color value were extracted from V5 and V4 (Fig. 3) in both hemispheres and included in a rm-ANOVA with factors attribute value (motion/color), region (V5/V4) and hemisphere (left/right). Neither the two-way interaction between attribute value and region ($F(1, 24) = 0.51$, $p = 0.48$) nor the three-way interaction between attribute value, region and hemisphere ($F(1, 24) = 0.53$, $p = 0.47$) were significant, which does not support the hypothesis that attribute values are systematically processed in V5 and V4. Further, there were no significant results for the main effects or the remaining interactions of no interest (attribute value: $F(1, 24) = 0.01$, $p = 0.93$; region: $F(1, 24) = 3.15$, $p = 0.09$; hemisphere: $F(1, 24) = 0.12$, $p = 0.73$; hemisphere x region: $F(1, 24) = 0.51$, $p = 0.48$; hemisphere x attribute value: $F(1, 24) = 2.57$, $p = 0.12$).

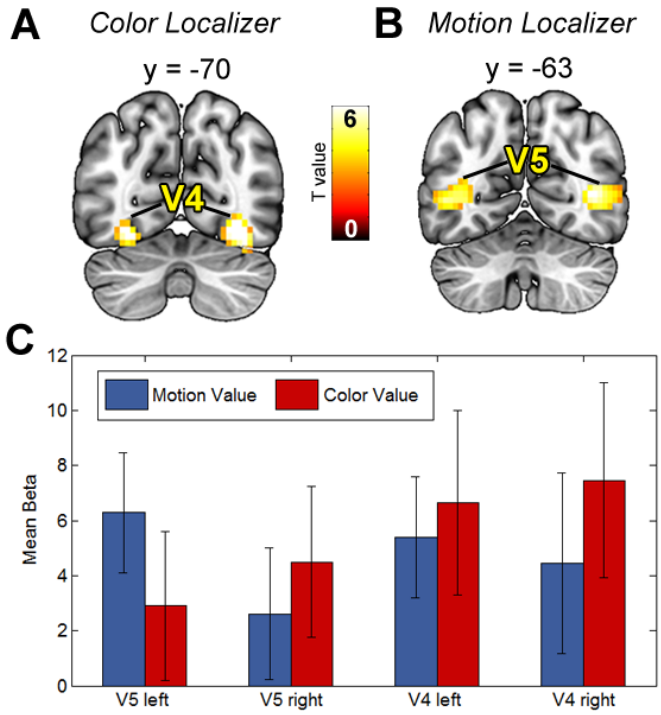

Figure 3. *Activations of localizer tasks in A) bilateral V4 (color localizer) and B) bilateral V5 (motion localizer). For illustration purposes, t-maps are thresholded at $p_{unc} < 0.001$ with a cluster extent threshold of $k_E = 15$. All clusters survive cluster-level FWE-correction at $p_{FWE} < 0.05$. C) Mean beta estimates of parametric modulation by motion and color value in bilateral V5 and V4 (regions adapted to single-subject peaks of localizers; for details see methods section). Results of the rm-ANOVA with factors attribute value (motion/color), region (V5/V4) and hemisphere (left/right) indicate that the data do not reveal a systematic representation of motion and color value in V5 and V4, respectively. Figure adapted from Magrabi et al. (2022a).*

3.2 Study 2: Neural Basis of Decisions in Alcohol-Dependent Patients

On the behavioral level, patients did not show significant differences to control subjects (earned profits: $F(1, 60) = 2.06$, $p = 0.16$; reaction times: $F(1, 60) = 0.47$, $p = 0.49$; model value threshold c_1: $F(1, 60) = 0.35$, $p = 0.56$; model number of accepts c_2: $F(1, 60) = 0.01$, $p = 0.91$; model offer index c_3: $F(1, 60) = 0.4$, $p = 0.55$; model threshold M: $F(1, 60) = 1.25$, $p = 0.27$).

The first goal of the fMRI analysis was to locate regions that process monetary values of offers. In the control group, whole-brain analyses revealed a significant parametric modulation of offer values (P1) in a distributed set of regions including dlPFC, ventral striatum, and dmPFC (Fig. 2A). The patient group showed activation in a largely overlapping set of regions, and a whole-brain comparison of parametric group effects in a two-sample t-test did not reveal significant differences between the two groups.

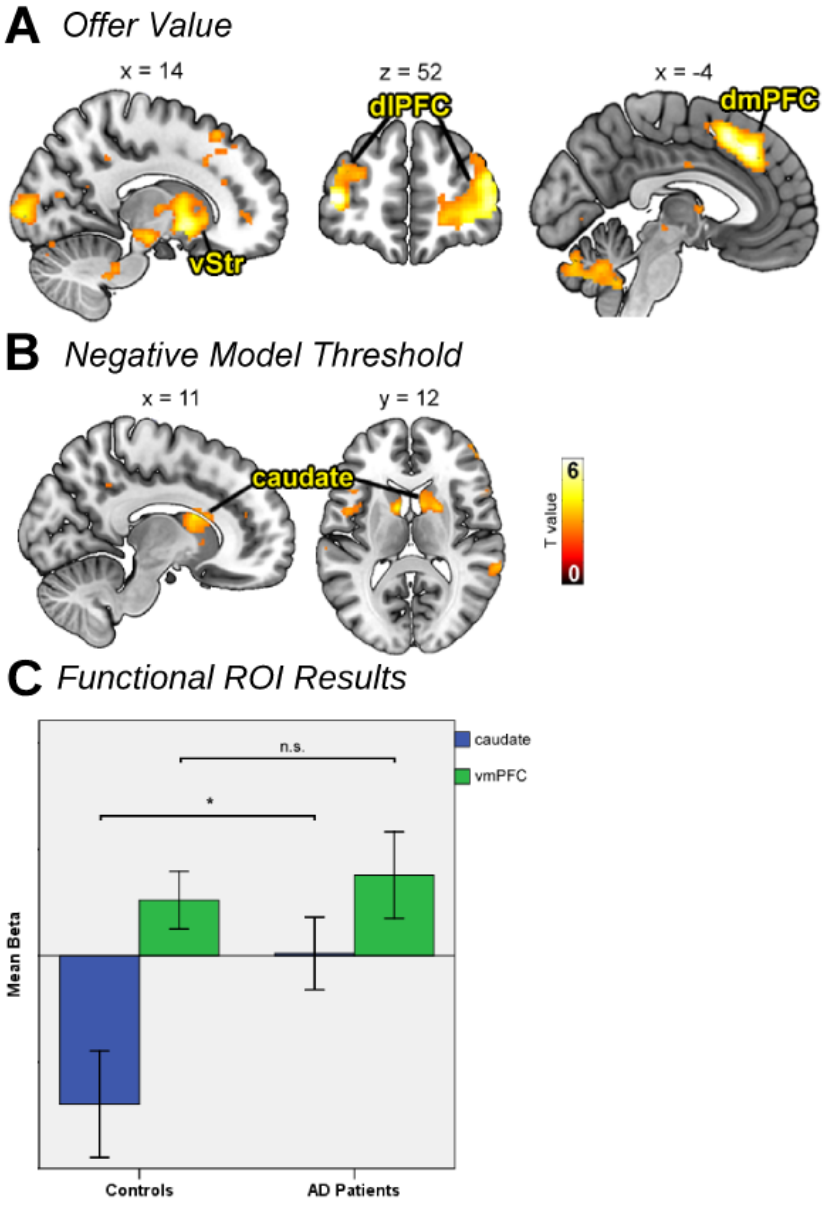

Figure 4. *Brain regions showing parametric effects in the control group for A) offer value and B) negative model threshold. For illustration purposes, t-maps are thresholded at p < 0.001 (uncorrected), k_E = 10. Labelled clusters survive cluster-level FWE-correction at p < 0.05. The patient group showed largely overlapping clusters for offer values, and no significant clusters for negative model thresholds. Abbreviations: dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; vStr, ventral striatum. C) Functional ROI results. Mean beta values in caudate nucleus were extracted from parametric modulators of model thresholds, and beta values in vmPFC from modulators of offer values, respectively. ROIs were defined as 5 mm spheres centered on coordinates from Economides et al (2014). Asterisks denote significant FDR-corrected p-values < 0.05. Figure adapted from Magrabi et al. (2022b).*

Second, I investigated brain areas that demonstrated activation related to model-based decision processes via the model threshold parameter of the GLM (P2). In the control group, I did not observe effects related to positive model thresholds, but activity in caudate nucleus and inferior parietal lobe (IPL) was significantly associated with negative model thresholds (Fig. 2B), indicating stronger neural activity when the threshold was low and participants were more likely to accept offers. This is in line with a previous study that

found stronger effects for negative compared to positive model thresholds (Economides et al., 2014) and can be due to the BOLD signal being highest for go responses. In contrast, the patient group did not exhibit any activity associated with negative model thresholds in the whole-brain brain analysis.

ROI analyses were conducted to investigate group differences in vmPFC (associated with value representation) and caudate activation (associated with model-based decision processes). There were no group differences with respect to parametric effects of offer values in vmPFC ($F(1,60) = 0.1$, $p_{FDR} = 0.834$), but there was a significant difference in parametric effects of model thresholds in the caudate nucleus ($F(1,60) = 4.4$, $p_{FDR} = 0.028$; Fig. 2C) involving stronger negative beta values in the control group.

4 Discussion

4.1 Study 1: Neural Integration of Attributes in Decision-Making

The computation of decision value often relies on multiple attributes. Values of relevant attributes must be computed separately before they can be integrated to an overall value that ultimately determines choices. So far, most studies investigated neural representations of overall values (Kable & Glimcher, 2007; Levy & Glimcher, 2012; Rangel et al., 2008; Sanfey et al., 2006), but little is known about the computation of attribute values (Basten et al., 2010; Kahnt et al., 2011; Lim et al., 2013; Suzuki, 2022).

Regarding the neural basis of attribute valuation, two competing hypotheses are conceivable. On the one hand, attribute values could be computed in distinct, attribute-specific brain regions (Basten et al., 2010; Hanks & Summerfield, 2017; Lim et al., 2013; Persichetti et al., 2015; Philiastides et al., 2010). These could be the same regions that are responsible for the identification and perceptual processing of a particular attribute. For example, this would imply that the fusiform face area, which is known to specifically process facial information (Kanwisher et al., 1997), is also involved in evaluating faces. Consequently, attribute values for a choice option would be calculated in different brain regions that highly depend on the specific attribute. Alternatively, attribute-specific value computations could instead be carried out within the general value network that is known to process overall values. This would mean that different attribute values as well as overall values would be processed in a centralized manner via vmPFC, ventral striatum, and PCC (Bartra et al., 2013; Clithero & Rangel, 2014; Ludwig et al., 2014; Peters & Büchel, 2010).

In the first study, I investigated these hypotheses with an experimental task that required the identification, valuation, and integration of separate decision attributes relating to the visual properties of motion and color. The perceptual processing of these attributes is robustly associated with area V5 for motion (Watson et al., 1993; Gallivan et al., 2018) and area V4 for color (McKeefry and Zeki, 1997). However, I did not find significant correlations with attribute values in these regions, neither in whole-brain nor in dedicated ROI analyses, which does not support the hypothesis that attribute values are computed in distinct, attribute-specific brain regions. Instead, whole brain-analyses showed that activity in PCC and ventral striatum correlated with color value, whereas activity related to motion value correlated with PCC and left PIT. In a direct comparison,

I found no region that had a significantly stronger representation of one attribute value over the other. Even though one cannot draw definitive conclusions from absence of evidence, this lack of specificity is consistent with the idea that the computation of attribute values is not implemented within attribute-specific cortical modules and is instead accomplished in a dynamic manner within a general valuation network including PCC, ventral striatum, and PIT.

In contrast to these results, previous studies have supported the hypothesis that attribute values are computed in attribute-specific regions. In an experiment by Lim et al. (2013), participants had to evaluate t-shirts based on how much they liked both the appearance and meaning of Korean symbols that were printed on them. The authors found that activity in fusiform gyrus correlated with visual values, whereas activity in superior temporal gyrus correlated with semantic values. In a probabilistic choice task, Philiastides et al. (2010) showed that activity in the fusiform face area corresponds to the value of face stimuli, while activity in the parahippocampal place area corresponds to the value of house stimuli. For both studies, the brain regions correlating with attribute values have also been associated with the identification and perceptual processing of the respective attributes, which is not confirmed by the data of my experiment.

There are several possible explanations for this discrepancy. First, the experimental design allowed me to differentiate value- and salience-related effects for each attribute, which was not possible in the experiment by Philiastides et al. (2010). Therefore, the value correlations observed in the fusiform face area and parahippocampal place area may be due to differences in salience rather than value. Second, the motion and color attributes in the experimental paradigm are robustly associated with well-defined regions in area V5 and V4. Arguably, this connection is less evident in the previously mentioned studies, in particular for the semantic attribute of the study by Lim et al. (2013), because the effects are more widespread across neural regions. Therefore, it is more unclear whether these results arise from attribute-specific regions, as it is more difficult to precisely determine regions of interest. Third, I used abstract, novel stimuli in my task, whereas the other studies used more familiar stimuli (faces, t-shirts, houses). One possible explanation for how the brain computes attribute values could be that novel stimuli are initially processed in a domain-general network, where attribute values are computed in a uniform manner. But if stimuli become more familiar and require frequent evaluation,

the computation of attribute values shifts to attribute-specific regions. Thus, the processing of attribute values might change as a function of learning, which could reconcile the effects of my experiment with previous studies.

In addition to analyzing individual attribute values, I also examined brain activity related to the overall value of stimuli and discovered correlations with activity in various regions, including left dlPFC and vmPFC (Fig. 4C). Due to the correlation between attribute values and overall values, my results show an overlap in their effects on the neural level, which cannot be easily disentangled on statistical grounds. However, it is noteworthy that the whole-brain analysis of overall values revealed a significant correlation for a cluster in vmPFC, which was not identified in my analysis of attribute values, neither for color nor motion value. This is in line with the view that vmPFC is one of the main regions that represent integrated value signals, which previous studies have argued for (Hare et al., 2009; Levy & Glimcher, 2012; O'Doherty et al., 2021; Padoa-Schioppa & Cai, 2011; Rangel & Clithero, 2013).

4.2 Study 2: Neural Basis of Decisions in Alcohol-Dependent Patients

Alcohol dependence (AD) is a widespread mental disorder, responsible for approximately 3.3 million deaths annually (World Health Organization, 2018). It is characterized by a loss of control over the consumption of alcohol, compulsive drinking despite repeated harmful consequences, and a negative emotional state (such as anxiety) in the face of withdrawal (Everitt & Robbins, 2005; Koob & Volkow, 2010, Phung et al., 2019).

Neuroscientific studies have suggested that AD develops through a systematic shift in the neural systems that regulate behavior, with increased involvement of the putamen/dorsolateral striatum (in humans/rodents) controlling habitual behavior, and decreased involvement of the caudate nucleus/dorsomedial striatum controlling flexible and model-based behavior (Corbit et al., 2012; DePoy et al., 2013; Everitt & Wolf, 2002; Furlong et al., 2014; Geerts et al., 2020; Gremel & Costa, 2013). Another account suggests that the impulsive behavior of AD patients (Rubio et al., 2008; Virkkunen, 1994) is based on an overactive neural value system (Arcurio et al., 2015; Goldstein & Volkow, 2011; Kamarajan et al., 2020; Seo et al., 2013), which has been associated with ventromedial prefrontal cortex (vmPFC; Bartra et al., 2013; Clithero & Rangel, 2014). However, it is not clear how and under what circumstances these changes in neural information processing occur.

In this study, I used a monetary decision-making task to investigate group differences between AD patients and control subjects with respect to value computation and model-based decision processes. Unlike the majority of previous studies in this area (Goldstein & Volkow, 2011; Schacht et al., 2013), the task did not include alcoholic stimuli, which made it possible to further examine whether AD patients have a general impairment of decision processes, even when alcoholic stimuli are not part of the decision context.

The results showed that patients had decreased functional representation of model-based decision processes in the caudate nucleus, while there were no significant differences between groups in terms of neural value representation or task performance. Previous research has identified that the caudate is a vital region for the computation of goal-directed decisions that involve taking multiple factors into account and long-term planning (Balleine & O'Doherty, 2010; Dolan & Dayan, 2013; Tanaka et al., 2008; Wunderlich et al., 2012). Likewise, in rodents, model-based decision processes have been associated with signals in dorsomedial striatum (Balleine, 2005; Corbit et al., 2012; Yin et al., 2005), which corresponds to the caudate activation that human neuroscience studies have identified. My finding that the neural representation of model-based decision processes in caudate nucleus is decreased for patients therefore suggests that alcohol dependence impairs the neural computations for goal-directed choices and supports the hypothesis that the ability to flexibly adapt choices to long-term consequences is one of the core functions affected by the disorder (Bechara et al., 2001; Goudriaan et al., 2007; Sebold et al., 2014).

The data further revealed that offer value is represented in a distributed set of regions including ventral striatum, dlPFC, and dmPFC for both the patient and the control group. However, I did not observe systematic differences with respect to neural value computations between the two groups, which does not support the hypothesis that alcohol dependence is based on a stimulus-independent overactive valuation system (Arcurio et al., 2015; Goldstein & Volkow, 2011; Kamarajan et al., 2020; Seo et al., 2013).

Surprisingly, patients also did not show deviations in behavioral task performance. One explanation for this finding could be that patients are able to achieve the same level of performance in sequential decision tasks but rely on different neural systems, which are not associated with the model threshold parameter that was employed in the computational model. It remains an open question for future research whether patients generally do not have deficits in performing sequential decision-making tasks, or whether different task designs can reveal systematic deficits in this regard.

4.3 General Discussion

In a nutshell, the goal of the first study was to study the neural mechanisms of computing attribute values: Whereas the *perceptual* processing of attributes can robustly be associated with specific regions (area V4 for perceiving color and V5 for perceiving motion), this does not seem to be the case for the process of attribute valuation during decision-making. Instead of being computed in characteristic regions, attribute values seem to be processed dynamically in a distributed network, comprising regions including PCC, ventral striatum, and PIT.

The core finding of the second study was that AD patients showed decreased functional activation relating to model-based decision processes in the caudate nucleus, whereas there were no neural differences relating to decision value. The effect in the caudate nucleus could be evidence for an impairment of model-based decision-making in AD patients. However, patients did not show behavioral differences in the decision task, which might indicate the involvement of compensatory mechanisms.

A common pattern in both studies and the literature in general is that the neural processes underlying decision-making appear to be very dynamic and distributed across a broad range of brain regions. So far, it does not seem to be possible to establish a clear functional association between the subprocesses involved in decision-making and specific brain regions. This relates to a larger debate in neuroscience about cognitive localization and the question to what extent cognitive functions can be mapped to specific brain regions (Genon et al., 2018). In this context, researchers have distinguished between functional segregation and functional integration (Friston, 2011).

Functional segregation describes the idea that the brain can be divided into regionally distinct modules based on functional or structural properties. In contrast, functional integration refers to the idea that no region is by itself responsible for a cognitive function, and that this requires a dynamic interplay between multiple brain regions in distributed neural networks instead (Siddiqi et al., 2022). In contrast to the traditional idea of phrenology (Gall, 1818), it is now widely accepted that functional segregation alone is not a realistic conceptual framework to explain neural processing. However, it is still common practice to characterize brain regions by their *relative* contribution to cognitive functions because some brain regions seem to be clearly *more* involved in specific cognitive functions than others.

While this appears to be a reasonable approach in some areas of cognitive neuroscience, I think the results of my experiments and the literature in general increasingly show that this approach is less promising and harder to apply in domain of decision neuroscience. The brain regions that are correlated with decision-related processes seem to be more numerous and more widespread, and the findings appear less robust than in other domains (e.g., perception or memory; although it is hard to objectively quantify relations between research domains like these). One explanation for this pattern could be that decision-making is late in the cognitive hierarchy and highly dependent on other cognitive functions. Everything must come together to be able to make good decisions: Perceiving relevant information, remembering past experiences, estimating the likelihood of future events, reasoning about logical implications, evaluating different attributes, etc. This high level of interdependence could explain why decision-related brain activations are more widespread across the brain and have a particularly high variance, which makes it harder to detect strong and robust functional associations in decision neuroscience.

This problem is further exacerbated by 1) the high number of confounding variables that correlate with decision-making processes (O'Doherty, 2014), like the correlation between decision value and salience (Kahnt and Tobler, 2013; Zhang et al., 2017), and 2) the limitations that neuroscientific methods like fMRI and EEG have with respect to spatial and temporal resolution. One could speculate that the neural pathways for decision-making become more simple and easier to detect when the decision scenario is already familiar and very similar to decisions that were made in the past. In that case, decisions become automatic, repetitive, and eventually turn into habits (Guida et al., 2022; Patterson & Knowlton, 2018). But for decisions in novel situations, which have never been encountered before and require the careful evaluation of multiple attributes under uncertainty, the underlying neural processes are likely much more complex.

One example for the issue with establishing reliable findings in decision neuroscience is also apparent in the results of my experiments. It is often considered as one of the most robust findings in the field that neural signals in vmPFC correlate with the overall decision value that determines choices (O'Doherty et al., 2021; Rangel et al., 2008). However, I could only replicate this finding in the first study, but not in the second study. Instead of the vmPFC, offer value in the second study correlated with neural activity in the ventral striatum, dlPFC, and dmPFC. While there could be several potential methodolog-

ical reasons for this (like differences in scanning parameters, trial structure, or task design), inconsistencies like these appear more common in the decision neuroscience literature than in other domains of cognitive neuroscience.

To improve on this, I think the field of decision neuroscience should focus on several aspects in the future. First, we need to work on a more detailed conceptual framework for decision-making. Terms like “value” or “uncertainty” are important for the decision-making process, but more work needs to be done to distinguish between different subtypes of general concepts like these, which would allow us to design more precise experiments with more control over confounding variables and higher signal-to-noise ratios. The distinction between overall values and attribute values in my first study was a step into this direction.

Second, more interdisciplinary work needs to be done. While this has been a popular theme for a while now, too many interdisciplinary efforts are largely superficial and need to go deeper to make real progress. I want to highlight two fields that are especially valuable for decision neuroscience: 1) Philosophy, which could help to make the conceptual framework more precise and inspire new angles for empirical research, and 2) psychiatry, which allows us to investigate how the cognitive and neural processes underlying decision-making are affected by dysfunctions. Like philosophy, psychiatry can also help to reveal blind spots in our conceptual framework for decision-making. My second study was a step into this direction and showed how the clinical disorder of alcoholism can affect the neural processes that drive model-based decision-making.

Third, decision neuroscience needs to put a stronger emphasis on functional integration instead of functional segregation, and focus on understanding the properties of interconnected neural networks rather than trying to characterize specific brain regions in isolation. Needless to say, advances in brain imaging methods that would improve the spatial or temporal resolution of neural measurements would make this paradigm shift a lot easier and open up possibilities for new experimental designs.

Making decisions is one of the most important parts of our lives. Our choices define our actions, our future, our happiness, and our identity. We invest a vast amount of time and energy into deliberating and optimizing our decisions, so a better scientific understanding of this process has a tremendous potential to improve our lives. The field of decision neuroscience faces a lot of challenges, but with the direction I outlined above, I think we are on a good path towards meaningful progress.

5 References

Arcurio, L. R., Finn, P. R., & James, T. W. (2015). Neural mechanisms of high-risk decisions-to-drink in alcohol-dependent women. *Addiction Biology*, *20*(2), 390–406. http://doi.org/10.1111/adb.12121

Balleine, B. W. (2005). Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiology & Behavior*, *86*(5), 717–30. http://doi.org/10.1016/j.physbeh.2005.08.061

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, *35*(1), 48–69. http://doi.org/10.1038/npp.2009.131

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–27. http://doi.org/10.1016/j.neuroimage.2013.02.063

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences*, *107*(50), 21767–72. http://doi.org/10.1073/pnas.0908104107

Batel, P., Pessione, F., Maitre, C., & Rueff, B. (1995). Relationship between alcohol and tobacco dependencies among alcoholics who smoke. *Addiction*, *90*(7), 977-80. http://doi.org/10.1046/j.1360-0443.1995.90797711.x

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S. W., & Nathan, P. E. (2001). Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia*, *39*(4), 376–89. http://doi.org/10.1016/S0028-3932(00)00136-6

Beck, A., Wüstenberg, T., Genauck, A., Wrase, J., Schlagenhauf, F., Smolka, M.N., Mann, K. & Heinz, A. (2012). Effect of Brain Structure, Brain Function, and Brain Connectivity on Relapse in Alcohol-Dependent Patients. *Archives of General Psychiatry*, *69*, 842-52. http://doi:10.1001/archgenpsychiatry.2011.2026

Becker, G. M., Degroot, M. H., & Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, *9*(3), 226–32. http://doi.org/10.1002/bs.3830090304

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton: Princeton UP.

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*(39), 12315–20. http://doi.org/10.1523/JNEUROSCI.2575-09.2009

Clithero, J., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302. http://doi.org/10.1093/scan/nst106

Corbit, L. H., Nie, H., & Janak, P. H. (2012). Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. *Biological Psychiatry*, *72*(5), 389–95. http://doi.org/10.1016/j.biopsych.2012.02.024

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–15. http://doi.org/10.1016/j.neuron.2011.02.027

Dayan, P., & Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. *Current Opinion in Neurobiology*, *18*(2), 185–96. http://doi.org/10.1016/j.conb.2008.08.003

DePoy, L., Daut, R., Brigman, J. L., MacPherson, K., Crowley, N., Gunduz-Cinar, O., Pickens, C. L., Cinar, R., Saksida, L. M., Kunos, G., Lovinger, D. M., Bussey, T. J., Camp, M. C., & Holmes, A. (2013). Chronic alcohol produces neuroadaptations to prime dorsal striatal learning. *Proceedings of the National Academy of Sciences*, *110*(36), 14783–8. http://doi.org/10.1073/pnas.1308198110

Deserno, L., Huys, Q. J. M., Boehme, R., Buchert, R., Heinze, H.-J., Grace, A. A., Dolan, R. J., Heinz, A., & Schlagenhauf, F. (2015). Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. *Proceedings of the National Academy of Sciences*, *112*(5), 1595–600. http://doi.org/10.1073/pnas.1417219112

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–25. http://doi.org/10.1016/j.neuron.2013.09.007

Economides, M., Guitart-Masip, M., Kurth-Nelson, Z., & Dolan, R. J. (2014). Anterior cingulate cortex instigates adaptive switches in choice by integrating immediate and delayed components of value in ventromedial prefrontal cortex. *Journal of Neuroscience*, *34*(9), 3340–9. http://doi.org/10.1523/JNEUROSCI.4313-13.2014

Everitt, B. J., & Robbins, T. W. (2005). Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, *8*(11), 1481–9. http://doi.org/10.1038/nn1579

Everitt, B. J. & Wolf, M. E. (2002). Psychomotor stimulant addiction: a neural systems perspective. *Journal of Neuroscience*, *22*(9), 3312–20. http://doi.org/20026356

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cerebral Cortex*, *17*(11), 2669–74. http://doi.org/10.1093/cercor/bhl176

First, M. B., & Gibbon, M. (2004). The Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I) and the Structured Clinical Interview for DSM-IV Axis II Disorders (SCID-II). In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive handbook of psychological assessment, Vol. 2. Personality assessment* (pp. 134–143). John Wiley & Sons, Inc..

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity, 1(1)*, 13-36. https://doi.org/10.1089/brain.2011.0008

Furlong, T. M., Jayaweera, H. K., Balleine, B. W., & Corbit, L. H. (2014). Binge-like consumption of a palatable food accelerates habitual control of behavior and is dependent on activation of the dorsolateral striatum. *Journal of Neuroscience*, *34*(14), 5012–22. http://doi.org/10.1523/JNEUROSCI.3707-13.2014

Gahnstrom, C.J., & Spiers, H.J. (2020). Striatal and hippocampal contributions to flexible navigation in rats and humans. *Brain and Neuroscience Advances*, *4*, 1–7. https://doi.org/10.1177/2398212820979772

Gall, F. J. (1818). *Anatomie et physiologie du système nerveux en général, et du cerveau en particulier* (Vol. 3). Librairie Grecque-Latine-Allemande.

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, *19(9)*, 519-34. https://doi.org/10.1038/s41583-018-0045-9

Geerts, J.P., Chersi, F., Stachenfeld, K.L. & Burgess, N. (2020). A general model of hippocampal and dorsal striatal learning and decision making. *Proceedings of the National Academy of Sciences*, *117*, 31427–37. https://doi.org/10.1073/pnas.2007981117

Genon, S., Reid, A., Langner, R., Amunts, K., & Eickhoff, S. B. (2018). How to Characterize the Function of a Brain Region. *Trends in Cognitive Sciences, 22(4)*, 350-64. https://doi.org/10.1016/j.tics.2018.01.010

Gläscher, J., Adolphs, R., Damasio, H., Bechara, A., Rudrauf, D., Calamia, M., … Tranel, D. (2012). Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Proceedings of the National Academy of Sciences*, *109*(36), 14681–6. http://doi.org/10.1073/pnas.1206608109

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585–95. http://doi.org/10.1016/j.neuron.2010.04.016

Glimcher, P. W., & Fehr, E. (2013). *Neuroeconomics: Decision making and the brain*. San Diego: Academic Press.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30(1), 535–74. http://10.1146/annurev.neuro.29.051605.113038

Goldstein, R. Z., & Volkow, N. D. (2011). Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews. Neuroscience*, *12*(11), 652–69. http://doi.org/10.1038/nrn3119

Goudriaan, A. E., Grekin, E. R., & Sher, K. J. (2007). Decision making and binge drinking: a longitudinal study. *Alcoholism, Clinical and Experimental Research*, *31*(6), 928–38. http://doi.org/10.1111/j.1530-0277.2007.00378.x

Gremel, C. M., & Costa, R. M. (2013). Orbitofrontal and striatal circuits dynamically encode the shift between goal-directed and habitual actions. *Nature Communications*, *4*, 1–12. http://doi.org/10.1038/ncomms3264

Guida, P., Michiels, M., Redgrave, P., Luque, D., & Obeso, I. (2022). An fMRI meta-analysis of the role of the striatum in everyday-life vs laboratory-developed habits. *Neuroscience & Biobehavioral Reviews*, 104826. https://doi.org/10.1016/j.neubiorev.2022.104826

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–88. http://doi.org/10.1016/0167-2681(82)90011-7

Hanks, T. D., & Summerfield, C. (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron*, *93(1)*, 15–31. https://doi.org/10.1016/j.neuron.2016.12.003

Hare, T. A., Camerer, C. F., Knoepfle, D. T., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience*, *30*(2), 583–90. http://doi.org/10.1523/JNEUROSCI.4089-09.2010

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, *324*(5927), 646–48. http://doi.org/10.1126/science.1168450

Hare, T. A., O'Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *Journal of Neuroscience*, *28*(22), 5623–5630. http://doi.org/10.1523/JNEUROSCI.1309-08.2008

Heekeren, H. R., Marrett, S., Ruff, D. A., Bandettini, P. A., & Ungerleider, L. G. (2006). Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proceedings of the National Academy of Sciences*, *103*(26), 10023–8. http://doi.org/10.1073/pnas.0603949103

Hunt, L. T., Dolan, R. J., & Behrens, T. E. J. (2014). Hierarchical competitions subserving multi-attribute choice. *Nature Neuroscience*, *17*(11), 1613–22. http://doi.org/10.1038/nn.3836

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–33. http://doi.org/10.1038/nn2007

Kable, J. W. & Glimcher, P. W. (2009) The neurobiology of decision: consensus and controversy. *Neuron*, *63(6)*, 733–45. http://doi:10.1016/j.neuron.2009.09.003

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J. D. (2011). Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *NeuroImage*, *56*(2), 709–715. http://doi.org/10.1016/j.neuroimage.2010.05.058

Kahnt, T., & Tobler, P. N. (2013). Salience signals in the right temporoparietal junction facilitate value-based decisions. *Journal of Neuroscience*, *33*(3), 863–9. http://doi.org/10.1523/JNEUROSCI.3531-12.2013

Kamarajan, C., Ardekani, B.A., Pandey, A.K., Kinreich, S., Pandey, G., Chorlian, D.B., Meyers, J. L., Zhang, J., Bermudez, E., Stimus, A. T., & Porjesz, B. (2020). Random forest classification of alcohol use disorder using fMRI functional connectivity, neuropsychological functioning, and impulsivity measures. *Brain Sciences*, *10(2)*, 115. https://doi.org/10.3390/brainsci10020115

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–11.

Koob, G. F., & Volkow, N. D. (2010). Neurocircuitry of addiction. *Neuropsychopharmacology*, *35*(1), 217–38. http://doi.org/10.1038/npp.2009.110

Leathers, M. L., & Olson, C. R. (2012). In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science*, *338*(6103), 132–5. http://doi.org/10.1126/science.1226405

Lee, S., Yu, L.Q., Lerman, C. & Kable, J.W. (2021). Subjective value, not a gridlike code, describes neural activity in ventromedial prefrontal cortex during value-based decision-making. *NeuroImage*, *237*, 118159. https://doi.org/10.1016/j.neuroimage.2021.118159

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–38. http://doi.org/10.1016/j.conb.2012.06.001

Levy, I., Snell, J., Nelson, A. J., Rustichini, A., & Glimcher, P. W. (2010). Neural representation of subjective value under risk and ambiguity. *Journal of Neurophysiology*, *103*(2), 1036–47. http://doi.org/10.1152/jn.00853.2009

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2013). Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *Journal of Neuroscience*, *33*(20), 8729–41. http://doi.org/10.1523/JNEUROSCI.4809-12.2013

Litt, A., Plassmann, H., Shiv, B., & Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cerebral Cortex*, *21*(1), 95–102. http://doi.org/10.1093/cercor/bhq065

Ludwig, V. U., Stelzel, C., Krutiak, H., Magrabi, A., Steimke, R., Paschke, L. M., Kathmann, N., & Walter, H. (2014). The suggestible brain: posthypnotic effects on value-based decision-making. *Social Cognitive and Affective Neuroscience*, *9*(9), 1281–8. http://doi.org/10.1093/scan/nst110

Magrabi, A., Ludwig, V. U., Stoppel, C. M., Paschke, L. M., Wisniewski, D., Heekeren, H., & Walter, H. (2022a). Dynamic Computation of Value Signals via a Common Neural Network in Multi-Attribute Decision-Making. *Social Cognitive and Affective Neuroscience*, *17(7)*, 683–93. https://doi.org/10.1093/scan/nsab125

Magrabi, A., Beck, A., Schad, D. J., Stoppel, C. M, Lett, T. A., Charlet, K., Kiefer, F., Heinz, A. & Walter, H. (2022b). Alcohol Dependence decreases Functional Activation of the Caudate Nucleus during Model-Based Decision Processes. *Alcoholism: Clinical and Experimental Research*, *46(5)*, 749–58. http://doi.org/10.1111/acer.14812

Maunsell, J. H. R. (2004). Neuronal representations of cognitive state: reward or attention? *Trends in Cognitive Sciences*, *8*(6), 261–5. http://doi.org/10.1016/j.tics.2004.04.003

McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal Prediction Errors in a Passive Learning Task Activate Human Striatum. *Neuron*, *38*(2), 339–346. http://doi.org/10.1016/S0896-6273(03)00154-5

McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, *120*, 2229–42. https://doi.org/10.1093/brain/120.12.2229

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–47. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *34*(6237), 52–4.

Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness Versus Reason in the Ultimatum Game. *Science*, *289*(5485), 1773–5. http://doi.org/10.1126/science.289.5485.1773

O'Doherty, J. P. (2014). The problem with value. *Neuroscience & Biobehavioral Reviews*, *43*, 259–68. http://doi.org/10.1016/j.neubiorev.2014.03.027

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron*, *38*(2), 329–37. http://doi.org/10.1016/S0896-6273(03)00169-7

O'Doherty, J. P., Rutishauser, U., & Iigaya, K. (2021). The hierarchical construction of value. *Current Opinion in Behavioral Sciences*, *41*, 71–7. https://doi.org/10.1016/j.cobeha.2021.03.027

Padoa-Schioppa, C. (2009). Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience*, *29*(44), 14004–14. http://doi.org/10.1523/JNEUROSCI.3751-09.2009

Padoa-Schioppa, C., & Cai, X. (2011). The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. *Annals of the New York Academy of Sciences*, *1239*, 130–7. http://doi.org/10.1111/j.1749-6632.2011.06262.x

Patterson, T. K., & Knowlton, B. J. (2018). Subregional specificity in human striatal habit learning: a meta-analytic review of the fMRI literature. *Current Opinion in Behavioral Sciences, 20*, 75–82. https://doi.org/10.1016/j.cobeha.2017.10.005

Pelletier, G., Aridan, N., Fellows, L. K., & Schonberg, T. (2021). A Preferential Role for Ventromedial Prefrontal Cortex in Assessing "the Value of the Whole" in Multiattribute Object Evaluation. *Journal of Neuroscience*, *41(23)*, 5056–68. https://doi.org/10.1523/JNEUROSCI.0241-21.2021

Persichetti, A. S., Aguirre, G. K., & Thompson-Schill, S. L. (2015). Value is in the eye of the beholder: early visual cortex codes monetary value of objects during a diverted attention task. *Journal of Cognitive Neuroscience*, *27(5)*, 893–901. https://doi.org/10.1162/jocn_a_00760

Peters, J., & Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research*, *213*(2), 135–41. http://doi.org/10.1016/j.bbr.2010.04.031

Phelps, E. A., Lempert, K. M., & Sokol-Hessner, P. (2014). Emotion and Decision Making: Multiple Modulatory Neural Circuits. *Annual Review of Neuroscience*, *37(1)*, 263–287. http://doi.org/10.1146/annurev-neuro-071013-014119

Philiastides, M. G., Biele, G., & Heekeren, H. R. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences*, *107(20)*, 9430–5. https://doi.org/10.1073/pnas.1001732107

Phung, Q.H., Snider, S.E., Tegge, A.N. & Bickel, W.K. (2019). Willing to work but not to wait: individuals with greater alcohol use disorder show increased delay discounting across commodities and less effort discounting for alcohol. *Alcoholism: Clinical and Experimental Research*, *43*, 927–36. https://doi.org/10.1111/acer.13996

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *Journal of Neuroscience*, *30*(32), 10799–808. http://doi.org/10.1523/JNEUROSCI.0788-10.2010

Plassmann, H., O'Doherty, J., & Rangel, A. (2007). Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of Neuroscience*, *27*(37), 9984–8. http://doi.org/10.1523/JNEUROSCI.2131-07.2007

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, *105*(3), 1050–4. http://doi.org/10.1073/pnas.0706929105

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–56. http://doi.org/10.1038/nrn2357

Rangel, A., & Clithero, J. (2013). The computation of stimulus values in simple choice. In P. Glimcher & E. Fehr (Eds.), *Neuroeconomics: Decision Making and the Brain: Second Edition* (pp. 125–148). San Diego: Academic Press. http://doi.org/10.1016/B978-0-12-416008-8.00008-5

Rubio, G., Jiménez, M., Rodríguez-Jiménez, R., Martínez, I., Avila, C., Ferre, F., Jiménez-Arriero, M. A., Ponce, G., & Palomo, T. (2008). The role of behavioral impulsivity in the development of alcohol dependence: a 4-year follow-up study. *Alcoholism, Clinical and Experimental Research*, *32*(9), 1681–7. http://doi.org/10.1111/j.1530-0277.2008.00746.x

Sanfey, A. G., Loewenstein, G., McClure, S. M., & Cohen, J. D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends in Cognitive Sciences*, *10*(3), 108–16. http://doi.org/10.1016/j.tics.2006.01.009

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science*, *300*(5626), 1755–8. http://doi.org/10.1126/science.1082976

Schacht, J. P., Anton, R. F., & Myrick, H. (2013). Functional neuroimaging studies of alcohol cue reactivity: a quantitative meta-analysis and systematic review. *Addiction Biology*, *18*(1), 121–33. http://doi.org/10.1111/j.1369-1600.2012.00464.x

Schad, D. J., Garbusow, M., Friedel, E., Sommer, C., Sebold, M., Hägele, C., Bernhardt, N., Nebe, S., Kuitunen-Paul, S., Liu, S., Eichmann, U., Beck, A., Wittchen, H. U., Walter, H., Sterzer, P., Zimmermann, U. S., Smolka, M. N., Schlagenhauf, F., Huys, Q. J. M., Heinz, A., & Rapp, M. A. (2019). Neural correlates of instrumental responding in the context of alcohol-related cues index disorder severity and relapse risk. *European Archives of Psychiatry and Clinical Neuroscience*, *269(3)*, 295-308. http://doi:10.1007/s00406-017-0860-4

Schonberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-Based Decision Making. *Journal of Neuroscience*, *27*(47), 12860–7. http://doi.org/10.1523/JNEUROSCI.2496-07.2007

Sebold, M., Deserno, L., Nebe, S., Nebe, S., Schad, D. J., Garbusow, M., Hägele, C., Keller, J., Jünger, E., Kathmann, N., Smolka, M. N., Rapp, M. A., Schlagenhauf, F., Heinz, A., & Huys, Q. J. M. (2014). Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology*, *70(2)*, 122–31. http://doi.org/10.1159/000362840

Seo, D., Lacadie, C. M., Tuit, K., Hong, K.-I., Constable, R. T., & Sinha, R. (2013). Disrupted ventromedial prefrontal function, alcohol craving, and subsequent relapse risk. *JAMA Psychiatry*, *70*(7), 727–39. http://doi.org/10.1001/jamapsychiatry.2013.762

Sharpe, M.J., Stalnaker, T., Schuck, N.W., 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. https://doi.org/10.1146/annurev-psych-010418-102824

Siddiqi, S. H., Kording, K. P., Parvizi, J., & Fox, M. D. (2022). Causal mapping of human brain function. *Nature Reviews Neuroscience, 23(6)*, 361–75. http://doi.org/10.1038/s41583-022-00583-8

Smittenaar, P., FitzGerald, T. H., Romei, V., Wright, N. D., & Dolan, R. J. (2013). Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. *Neuron*, *80(4)*, 914–19. https://doi.org/10.1016/j.neuron.2013.08.009

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Suzuki, S. (2022). Constructing value signals for food rewards: determinants and the integration. *Current Opinion in Behavioral Sciences*, *46*, 101178. https://doi.org/10.1016/j.cobeha.2022.101178

Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating consequences: brain systems that encode the causal effects of actions. *Journal of Neuroscience*, *28*(26), 6750–5. http://doi.org/10.1523/JNEUROSCI.1808-08.2008

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science*, *315*(5811), 515–8. http://doi.org/10.1126/science.1134239

Virkkunen, M. (1994). Personality Profiles and State Aggressiveness in Finnish Alcoholic, Violent Offenders, Fire Setters, and Healthy Volunteers. *Archives of General Psychiatry*, *51*(1), 28. http://doi.org/10.1001/archpsyc.1994.03950010028004

Von Neumann, J., & Morgenstern, O. (2007). *Theory of games and economic behavior*. In Theory of games and economic behavior. Princeton university press.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience*, *18*(7), 2069–81. http://doi.org/10.1046/j.1460-9568.2003.02922.x

Watson, J. D. G., Myers, R., Frackowiak, R. S. J., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., & Zeki, S. (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex*, *3*(2), 79–94. http://doi.org/10.1093/cercor/3.2.79

World Health Organization (2018). *World Health Statistics*. Geneva: WHO.

Wunderlich, K., Dayan, P., & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience*, *15(5)*, 786–91. http://doi.org/10.1038/nn.3068

Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005). The role of the dorsomedial striatum in instrumental conditioning. *The European Journal of Neuroscience*, *22*(2), 513–23. http://doi.org/10.1111/j.1460-9568.2005.04218.x

Zhang, Z., Fanning, J., Ehrlich, D. B., Chen, W., Lee, D., & Levy, I. (2017). Distributed neural representation of saliency controlled value and category during anticipation of rewards and punishments. *Nature Communications*, *8(1)*, 1–14. https://doi.org/10.1038/s41467-017-02080-4

Eidesstattliche Versicherung

„Ich, Amadeus Magrabi, versichere an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorgelegte Dissertation mit dem Thema „The Neural Computation of Value in Decision-Making“ („Die neuronale Berechnung von Bewertungen in der Entscheidungsfindung“) selbstständig und ohne nicht offengelegte Hilfe Dritter verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel genutzt habe.

Alle Stellen, die wörtlich oder dem Sinne nach auf Publikationen oder Vorträgen anderer Autoren/innen beruhen, sind als solche in korrekter Zitierung kenntlich gemacht. Die Abschnitte zu Methodik (insbesondere praktische Arbeiten, Laborbestimmungen, statistische Aufarbeitung) und Resultaten (insbesondere Abbildungen, Graphiken und Tabellen) werden von mir verantwortet.

Ich versichere ferner, dass ich die in Zusammenarbeit mit anderen Personen generierten Daten, Datenauswertungen und Schlussfolgerungen korrekt gekennzeichnet und meinen eigenen Beitrag sowie die Beiträge anderer Personen korrekt kenntlich gemacht habe (siehe Anteilserklärung). Texte oder Textteile, die gemeinsam mit anderen erstellt oder verwendet wurden, habe ich korrekt kenntlich gemacht.

Meine Anteile an etwaigen Publikationen zu dieser Dissertation entsprechen denen, die in der untenstehenden gemeinsamen Erklärung mit dem/der Erstbetreuer/in, angegeben sind. Für sämtliche im Rahmen der Dissertation entstandenen Publikationen wurden die Richtlinien des ICMJE (International Committee of Medical Journal Editors; www.icmje.og) zur Autorenschaft eingehalten. Ich erkläre ferner, dass ich mich zur Einhaltung der Satzung der Charité – Universitätsmedizin Berlin zur Sicherung Guter Wissenschaftlicher Praxis verpflichte.

Weiterhin versichere ich, dass ich diese Dissertation weder in gleicher noch in ähnlicher Form bereits an einer anderen Fakultät eingereicht habe.

Die Bedeutung dieser eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unwahren eidesstattlichen Versicherung (§§156, 161 des Strafgesetzbuches) sind mir bekannt und bewusst.“

_____ _____

Datum, Ort Unterschrift

Anteilserklärung

Publikation 1

Magrabi, A., Ludwig, V.U., Stoppel, C.M., Paschke, L.M., Wisniewski, D., Heekeren, H., & Walter, H. (2022). Dynamic Computation of Value Signals via a Common Neural Network in Multi-Attribute Decision-Making. *Social Cognitive and Affective Neuroscience*, *17(7)*, 683–93. https://doi.org/10.1093/scan/nsab125

Amadeus Magrabi hatte folgenden Anteil an Publikation 1:

- Entwicklung der Studienidee und Hypothesen
- Literaturrecherche
- Programmierung des Experiments
- Probandenrekrutierung
- Datenerhebung
- Datenvorverarbeitung und statistische Analyse (Verhaltens- und MRT-Daten)
- Verfassung des Manuskripts, inklusive Erstellung aller Abbildungen und Tabellen
- Einreichung bei Zeitschrift, Überarbeitung/Revision

Die Koautoren hatten den folgenden Anteil an Publikation 1:

- Beratung und Diskussion zur Studenidee, Analyse-Strategie und Interpretation der Ergebnisse
- Verbesserungsvorschläge zu einzelnen Formulierungen im Manuskript
- Teilweise Unterstützung bei der Datenerhebung, damit immer mindestens zwei Personen bei den MRT-Messungen anwesend sind

Publikation 2:

Magrabi, A., Beck, A., Schad, D.J., Stoppel, C.M, Lett, T.A., Charlet, K., Kiefer, F., Heinz, A. & Walter, H. (2022). Alcohol Dependence decreases Functional Activation of the Caudate Nucleus during Model-Based Decision Processes. *Alcoholism: Clinical and Experimental Research*, *46(5)*, 749–58. http://doi.org/10.1111/acer.14812

Amadeus Magrabi hatte folgenden Anteil an Publikation 2:

- Entwicklung der Hypothesen
- Literaturrecherche

- Datenvorverarbeitung und statistische Analyse (Verhaltens- und MRT-Daten), mit der Ausnahme der Berechnung des Entscheidungsmodells zu den Verhaltensdaten
- Verfassung des Manuskripts, inklusive Erstellung aller Abbildungen und Tabellen
- Einreichung bei Zeitschrift, Überarbeitung/Revision

Die Koautoren hatten den folgenden Anteil an Publikation 2:

- Datenerhebung (im Rahmen des „NGFN“-Forschungsprogramms)
- Beratung und Diskussion zur Analyse-Strategie und Interpretation der Ergebnisse
- Die Berechnung des Entscheidungsmodells zu den Verhaltensdaten (angelehnt an Economides et al., 2014) wurde von einem Koautor übernommen, Amadeus Magrabi hat die berechneten Modellparameter mit der MRT-Analyse integriert
- Verbesserungsvorschläge zu einzelnen Formulierungen im Manuskript

_____

Unterschrift des Doktoranden/der Doktorandin

Amadeus Magrabi

Original Publications

Publication 1

Magrabi, A., Ludwig, V.U., Stoppel, C.M., Paschke, L.M., Wisniewski, D., Heekeren, H., & Walter, H. (2022). Dynamic Computation of Value Signals via a Common Neural Network in Multi-Attribute Decision-Making. *Social Cognitive and Affective Neuroscience*, *17(7)*, 683–93. https://doi.org/10.1093/scan/nsab125

Impact Factor (2019): 3,571

Platz 13 von 77 im Fachgebiet *Psychology* (Top 17%)

Publication Rank

Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI
Selected Categories: **"PSYCHOLOGY"** Selected Category Scheme: WoS
Gesamtanzahl: 77 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1	PSYCHOLOGICAL BULLETIN	52,600	20.850	0.027120
2	Annual Review of Psychology	21,277	18.156	0.019500
3	PSYCHOTHERAPY AND PSYCHOSOMATICS	4,275	14.864	0.006480
4	Annual Review of Clinical Psychology	6,126	13.692	0.009570
5	JOURNAL OF CHILD PSYCHOLOGY AND PSYCHIATRY	19,837	7.035	0.021080
6	PSYCHOLOGICAL REVIEW	29,567	6.857	0.009080
7	PSYCHOLOGICAL MEDICINE	26,702	5.813	0.039350
8	DEPRESSION AND ANXIETY	9,355	4.702	0.013860
9	JOURNAL OF MEMORY AND LANGUAGE	9,767	3.893	0.007610
10	PSYCHOSOMATIC MEDICINE	12,560	3.702	0.009890
11	PSYCHOPHYSIOLOGY	14,586	3.692	0.012670
12	INTERNATIONAL JOURNAL OF EATING DISORDERS	9,613	3.668	0.010750
13	Social Cognitive and Affective Neuroscience	7,347	3.571	0.019570
14	JOURNALS OF GERONTOLOGY SERIES B-PSYCHOLOGICAL SCIENCES AND SOCIAL SCIENCES	9,435	3.502	0.009930
15	HUMAN FACTORS	6,763	3.165	0.005330
16	HEALTH PSYCHOLOGY	11,888	3.056	0.015480
17	COGNITIVE PSYCHOLOGY	7,784	3.029	0.004590
18	PSYCHO-ONCOLOGY	11,286	3.006	0.016340
19	INTERNATIONAL PSYCHOGERIATRICS	7,341	2.940	0.009920

OXFORD

Social Cognitive and Affective Neuroscience, 2022, **17**, 683–693
DOI: https://doi.org/10.1093/scan/nsab125
Advance Access Publication Date: 27 November 2021
Original Manuscript

Dynamic computation of value signals via a common neural network in multi-attribute decision-making

Amadeus Magrabi,[1,2] Vera U. Ludwig[1,2,3,4] Christian M. Stoppel,[2] Lena M. Paschke,[1,2,5] David Wisniewski,[6] Hauke R. Heekeren,[1,7] and Henrik Walter[1,2,8]

[1]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin 10117, Germany
[2]Department of Psychiatry and Psychotherapy, Charité—Universitätsmedizin Berlin, Berlin 10117, Germany
[3]Wharton Neuroscience Initiative, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[5]Department of Psychology, Humboldt-Universität zu Berlin, Berlin 12489, Germany
[6]Department of Experimental Psychology, Ghent University, Gent 9000, Belgium
[7]Department of Education and Psychology, Freie Universität Berlin, Berlin 14195, Germany
[8]Berlin Center for Advanced Neuroimaging, Charité—Universitätsmedizin Berlin, Berlin 10119, Germany
Correspondence should be addressed to Amadeus Magrabi, Department of Psychiatry and Psychotherapy, Charité—Universitätsmedizin Berlin, Berlin 10117, Germany. E-mail: amadeus.magrabi@gmail.com.

Abstract

Studies in decision neuroscience have identified robust neural representations for the value of choice options. However, overall values often depend on multiple attributes, and it is not well understood how the brain evaluates different attributes and integrates them to combined values. In particular, it is not clear whether attribute values are computed in distinct attribute-specific regions or within the general valuation network known to process overall values. Here, we used a functional magnetic resonance imaging choice task in which abstract stimuli had to be evaluated based on variations of the attributes color and motion. The behavioral data showed that participants responded faster when overall values were high and attribute value differences were low. On the neural level, we did not find that attribute values were systematically represented in areas V4 and V5, even though these regions are associated with attribute-specific processing of color and motion, respectively. Instead, attribute values were associated with activity in the posterior cingulate cortex, ventral striatum and posterior inferior temporal gyrus. Furthermore, overall values were represented in dorsolateral and ventromedial prefrontal cortex, and attribute value differences in dorsomedial prefrontal cortex, which suggests that these regions play a key role for the neural integration of attribute values.

Key words: decision-making; value; attribute; salience; fMRI

Introduction

Valuation is a crucial part of decision-making. To make beneficial choices, available options need to be accurately evaluated, and the ones with the highest value need to be selected. Studies in decision neuroscience have investigated valuation processes extensively and found neural representations of value in the ventromedial prefrontal cortex (vmPFC), posterior cingulate cortex (PCC) and ventral striatum (Bartra *et al.*, 2013; Clithero and Rangel, 2014). Most studies addressed the question as to where overall values of choice options are processed in the brain (Rangel *et al.*, 2008; Kable and Glimcher, 2009). However, overall values are often based on values of different attributes. For example, the overall value of a car can depend on the evaluation of its size, speed or color. In these cases, values of relevant attributes have to be computed separately, before they can be integrated to a combined value that ultimately determines choices. So far, the majority of studies investigated neural representations of overall values (Sanfey *et al.*, 2006; Kable and Glimcher, 2007; Rangel *et al.*, 2008; Levy and Glimcher, 2012), but little is known about the computation and integration of attribute values (Basten *et al.*, 2010; Kahnt *et al.*, 2011; Lim *et al.*, 2013; Suzuki *et al.*, 2017; Vaidya *et al.*, 2018; Pelletier *et al.*, 2021).

With regard to known functional specializations of different brain regions, two hypotheses concerning neuronal attribute valuation are conceivable. On the one hand, attribute values could be computed in distinct, attribute-specific brain regions (Basten *et al.*, 2010; Philiastides *et al.*, 2010; Lim *et al.*, 2013). From this perspective, attribute values are processed within regions that are also specialized in processing objective properties of the particular attributes. For instance, it would be predicted that the fusiform face area, which is known to selectively process faces (Kanwisher *et al.*, 1997), is also responsible for the evaluation of faces. As a

Received: 2 March 2021; **Revised:** 12 October 2021; **Accepted:** 25 November 2021

result, attribute values for a choice option would be computed in distinct neural regions that highly depend on the particular attribute, which is consistent with studies that found evidence of value correlations in sensory regions (Gold and Shadlen, 2007; Serences, 2008; Persichetti *et al.*, 2015; Hanks and Summerfield, 2017). On the other hand, all attribute-specific value computations could instead be performed within the general valuation network that is known to process overall values (Bartra *et al.*, 2013; Clithero and Rangel, 2014). As such, different attribute values as well as overall values would be processed in a homogeneous and centralized manner via vmPFC, PCC and ventral striatum (Peters and Büchel, 2010; Ludwig *et al.*, 2014).

Here, we used model-based functional magnetic resonance imaging (fMRI) to distinguish between these hypotheses and investigate how attribute values are computed in the brain. Participants were presented with a dot stimulus varying in two constituent perceptual attributes: motion direction and dot color. Each attribute level (i.e. each particular motion direction and each color) was associated with a specific monetary gain or loss. Based on these individual attribute values, participants had to determine the overall value of the stimulus and decide to accept or reject the offer. In addition, we also conducted separate localizer tasks to identify regions specialized in the processing of the physical properties of the attributes, independent of attribute valuation.

Compared to previous studies on attribute valuation (Hare *et al.*, 2009; Basten *et al.*, 2010; Philiastides *et al.*, 2010; Kahnt *et al.*, 2011; Park *et al.*, 2011; Lim *et al.*, 2013; Hutcherson *et al.*, 2015; Suzuki *et al.*, 2017; de Berker *et al.*, 2019; Pelletier *et al.*, 2021), this approach combines two methodological advantages:

(i) Our decision task is based on two well-investigated stimulus attributes, which are processed in separate, well-defined cortical modules, namely Area V5 for motion (Watson *et al.*, 1993) and Area V4 for color (McKeefry and Zeki, 1997). By using independent localizer tasks, we were thus able to specifically address whether valuation of individual attributes proceeds separately within these well-defined perceptual areas, or instead within the network known to compute overall values (comprising vmPFC, PCC and ventral striatum).

(ii) The values assigned to each attribute in our task span a range of both positive and negative values, which allows us to disentangle the effects of value and salience. Salience, in contrast to value, refers to the subjective importance of a stimulus, which ultimately guides the amount of attentional resources deployed to a certain stimulus or event (Maunsell, 2004; Zink *et al.*, 2006; Kahnt and Tobler, 2013). When valuation processes are studied only by means of positive values, salience and value are indistinguishable, as more positive stimuli are also more salient (Litt *et al.*, 2011; Leathers and Olson, 2012; Kahnt *et al.*, 2014; Zhang *et al.*, 2017). For that reason, salience is a common experimental confound in the majority of decision neuroscience studies (O'Doherty, 2014). However, if both positive and negative values are involved, salience- and valuation-related mechanisms of decision processes can easily be distinguished, because stimuli with a high negative value are of low value, but of high salience (as there is a high incentive to avoid them). Hence, our design allows us to dissociate regions that compute the value and salience of each attribute, to provide a more elaborate account of attribute valuation in the human brain.

Methods

Participants

Twenty-five right-handed subjects (14 female; mean age 28.1 ± 4 s.d.) participated in the study. All subjects had normal or corrected-to-normal vision, had no history of psychiatric or neurological illnesses, were free from medication interfering with fMRI performance, were native German speakers, and provided informed consent before participation. All experimental procedures were approved by the local ethics committee. Participants were compensated with 25€ for study participation and could receive an additional performance-dependent bonus in the range of 0–17€.

Stimulus material and experimental design

All stimuli were presented using MATLAB (version 8; MathWorks, Natick, MA) and the Psychophysics Toolbox extension (version 3; Brainard, 1997).

Main decision task

We decided to use a version of the random dot task, because we found it to be a well-established paradigm in the perceptual decision-making literature to elicit motion-related activation in Area V5 (e.g. Gold and Shadlen, 2007; Gallivan *et al.*, 2018). The task was designed to require the computation and integration of two distinct attribute values. For this purpose, stimuli varying in the attribute dimensions color and motion direction were employed. Each stimulus consisted of a set of 200 dots presented within a circular aperture in front of a black background (dot radius 0.07° of visual angle, aperture radius 2°, Figure 1A). Across trials, color and motion of the stimulus varied with respect to six different levels (color: blue, red, turquoise, green, brown and pink; motion direction: upward, downward, up-left, up-right, down-left and down-right with a uniform angle of 60° between directions). During each trial, all dots were constantly displayed in the same color and moved coherently into one direction (dot velocity 4°/s). Dots reaching the aperture limit were randomly replotted at the opposite semicircle (orthogonal to the current motion direction) according to a beta function ($\alpha = 1.9$, $\beta = 1.9$) to maintain an even density distribution of dots within the circular display. For both color and motion, each of the six attribute levels was associated with a monetary value taken from the set [–0.15, –0.10, –0.05, 0.05, 0.10, 0.15€]. These associations between particular monetary values and attribute levels were counterbalanced across participants and had to be acquired in a separate learning session (described in the section "learning task").

In each trial, participants had to identify the attribute values of the current color and motion direction, while being required to maintain ocular fixation throughout the trial at a cross displayed in the center of the circular aperture. These values then had to be integrated, as the sum of attribute values indicated the overall value of the stimulus. Stimuli covered all possible combinations of attribute values except those summing to 0€ (i.e. $6 \times 6 = 36$ unique combinations, out of which the 30 combinations without zero-sums were included in the experiment). Note that motion and color value are orthogonal to each other in this design, which allows for an independent assessment of their respective effects. Based on the integrated overall value, participants had to decide whether to accept or reject the current offer (i.e. optimal choices result in accepting all positive overall values and rejecting all negative ones). Participants indicated their choices by pressing one of two designated buttons on a response box using their right index/middle finger (accept/reject), which terminated

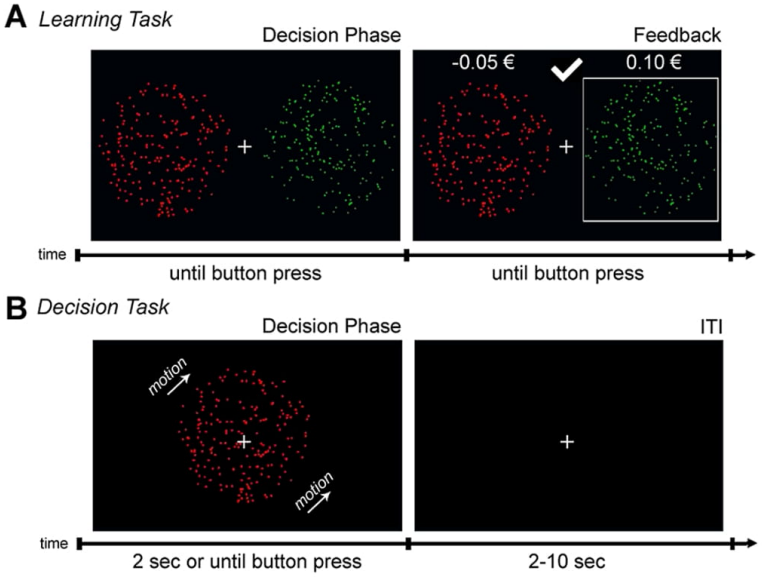

Fig. 1. Experimental design. A) Learning task. Participants had to indicate whether the left or right dot field represented a higher value and received feedback on the values of both attributes and their choice accuracy. The figure shows an example of a color value trial in which dots within both circular apertures were static but varied in their constituent color attribute. Motion value trials (which are not displayed) were designed analogously, except that dots were uniformly displayed in gray and varied with respect to their motion direction. B) Main decision task. In contrast to the learning task, participants were presented with a single circular aperture, within which the dots varied on a trial-by-trial basis with regard to both attributes. The subjects' task was to indicate whether they want to accept or reject a stimulus based on the sum of both attribute values. The decision screen terminated either by button press responses or after reaching a time limit of 2 s.

the current trial. If no response was given within 2 s after stimulus onset, the trial was automatically terminated and classified as if the inferior option had been chosen (i.e. stimuli with negative overall values were counted as accepted, and those with positive ones as rejected). Duration of inter-trial intervals (during which only the fixation cross remained on the screen) was randomized between 2 and 10 s according to a truncated exponential function ($\lambda = 6$, mean sec 5.1 ± 2.2 s.d.; Dale, 1999). After completion of the experiment, the overall values chosen for each trial were summed up and paid out as a monetary bonus (possible range: 0–17€) in addition to the monetary compensation for participation.

Our design makes use of a categorical manipulation of attribute levels. A dimensional manipulation would have also been possible (i.e. a spectrum from weak to strong motion opposed to six different motion directions), but we hypothesized that it might add more uncertainty to our task, because it can be more difficult for participants to assess the precise value of a stimulus in a dimensional design. By reducing this level of uncertainty, we reasoned that we can be more confident that participants made their choices with the intended attribute values and reduce noise in our data.

Learning task

To establish the associations between each of the individual attribute levels (i.e. the particular colors and motion directions) to one of the six monetary values, subjects completed an offline learning task (Figure 1B) in the days before scanning (mean days 2 ± 0.4 s.d.). Monetary values for color and motion levels were learned in separate blocks. During motion blocks, participants were presented with two apertures (4.5° to the left and right of the central fixation cross) each composed of 200 moving dots drawn in gray (dot and aperture size identical to the main decision task). For each trial, motion was coherent within each aperture (one of the six directions described above), but simultaneously presented apertures never showed the same direction. The participants' task was to indicate whether the left (index finger) or right (middle finger) dot field embodied the higher monetary value. After button presses, participants received feedback whether their choice was correct and corresponding monetary values were displayed above both apertures. Trials of color blocks were designed in an analogous manner with the difference that dots within each aperture remained stationary, but varied in their constituent color. Participants did not practice on stimuli combining both attributes in the learning task, to make sure that the combined experimental stimuli are not over-learned and attributes need to be actively integrated in the main decision task.

Each block consisted of 30 trials, which included two occurrences of all possible attribute value combinations. Participants completed a minimum of six blocks, including three motion and color blocks arranged in an alternating order (with the starting block type being counterbalanced across participants). After the sixth block, the task ended if participants achieved an accuracy of at least 95% during the last two blocks. If the accuracy criterion was not achieved, participants had to complete another motion and color block until it was satisfied.

Scanning session

Repetition of learning task

On the day of scanning, participants first repeated one motion and color block of the learning task outside of the scanner with 15 trials per block.

Decision task

After the learning task repetition, participants were placed in the MRI scanner. Before the recording of the first run, participants completed 15 practice trials (randomly taken from the available stimulus set) after each of which trial-wise feedback on their earnings and the values of presented attributes were displayed. The subsequent main decision task was separated into five runs with 60 trials each, with all of the 30 unique attribute value combinations (see experimental design of the main decision task) occurring twice per run. As a consequence, each motion and color value type (–0.15, –0.10, –0.05, 0.05, 0.10 and 0.15€ for each attribute) was presented 50 times throughout the entire experiment. Trial ordering was fully randomized with the exception that unique attribute value combinations were not allowed to occur twice within the first 30 trials of a run. There was no trial-wise feedback on participants' performance during the main decision task, but the total amount of earnings was displayed during brief pauses between runs.

Localizers

After completion of the decision task, motion and color localizers were conducted (order counterbalanced between participants). Both localizer tasks consisted of ten 24-s trials separated by a 12-s inter-trial interval. During the motion localizer task, subjects were presented with the same circular aperture as during the main experiment, which in contrast contained 200 dots drawn in gray. During the 24-s motion phase, dots moved coherently into a randomly chosen direction which was changed every second. During the 12-s static phase (i.e. the inter-trial interval), dots were repositioned to a new random location within the aperture after every second.

The color localizer task was designed in an analogous manner with a 24-s color phase and a 12-s inter-trial interval (achromatic phase). Stimuli consisted of a 6 × 6 checkerboard with each of the particular squares (0.4° × 0.4°) drawn in colors of random RGB values, and each square changing its color every second. The achromatic stimulus was geometrically identical, but squares changed their appearance every second only in achromatic space.

fMRI data acquisition and preprocessing

Functional data

Imaging was conducted on a 3-T Siemens Tim Trio MRI scanner (Siemens, Erlangen, Germany) with a 12-channel head coil. Functional volumes consisted of 33 continuous slices that were acquired in descending order by using a T2*-weighted gradient-echo sequence [repetition time (TR): 2 s; echo time (TE): 30 ms; matrix size: 64 × 64; field of view (FOV): 192 mm; flip angle: 78°; inter-slice gap: 0.75 mm; final voxel size: 3 × 3 × 3.75 mm]. For each participant, 133 volumes were recorded for each localizer task and an average number of 977 volumes for the main decision task (dependent on reaction times). To allow for steady-state magnetization, two dummy scans were acquired at the beginning of each run and discarded.

Structural data

For registration purposes, a high-resolution, T1-weighted structural volume was acquired from every subject after completion of the decision and localizer tasks using a magnetization-prepared rapid gradient-echo (MPRAGE) sequence (192 slices; TR: 1900 ms; TE 2.52 ms, matrix size: 256 × 256; FOV: 256 mm; flip angle: 9°; final voxel size: 1 × 1 × 1 mm).

fMRI data analysis

Preprocessing

Preprocessing was performed using SPM12 (Wellcome Trust Centre for Neuroimaging, UCL, London) and MATLAB (version 8; MathWorks, Natick, MA). Functional images were realigned, slice-time corrected, spatially normalized to the template of the Montreal Neurological Institute (MNI) and smoothed using a Gaussian kernel of 8-mm full-width at half-maximum.

GLM analysis of decision task

The decision task data were analyzed by means of two different general linear models (GLMs) for each participant. For both GLMs, event regressors were constructed as boxcar functions beginning at the time of stimulus onsets and the duration of the respective choice period.

For GLM1, regressor R1 comprised all trials during which participants made correct choices (accepting positive and rejecting negative overall values). Five linear parametric modulators of regressor R1 were included in the model to analyze neural correlates of the following decision variables: P1) motion value, P2) color value, P3) motion salience (absolute motion value), P4) color salience (absolute color value), and P5) absolute difference between motion and color value. The latter parametric modulator was included to investigate comparator regions, which could be responsible for the integration of the two attribute values. Note that P5 is not significantly correlated with P1 ($r = 0$), P2 ($r = 0$), P3 ($r = 0.09$) or P4 ($r = 0.09$). All variables were z-transformed before they were added to the model. In addition, to minimize the error term of GLM1, an additional regressor R2 comprising all incorrect choice trials was included, as well as six movement regressors R3–R8 from the realignment procedure.

GLM2 was created to analyze overall value and overall salience. These variables were not included in GLM1 due to multicollinearity which would result from significant correlations between overall values and attribute values (for each attribute $r = 0.77$, $P < 0.001$), and between overall salience and attribute salience ($r = 0.3$, $P < 0.001$). GLM2 was designed in an analogous manner to GLM1, but only included two instead of four parametric modulators: P1) overall value (sum of motion and color value) and P2) overall salience (absolute overall value).

For both models, all regressors were convolved with the canonical hemodynamic response function (HRF) and regressed against the blood-oxygen-level-dependent (BOLD) signal in each voxel. Parametric modulators were not orthogonalized to each other, allowing regressors to fully compete for explained variance. First-level contrasts were constructed by weighting all parametric modulators over baseline and submitted to second-level random-effects group analyses for statistical analysis. All statistical parametric maps from group analyses were thresholded at $P < 0.001$ (uncorrected) for voxel-level inference with a minimum cluster-size criterion of 15 contiguous voxels, and subsequent cluster-level family-wise error (FWE) correction for multiple testing at $P < 0.05$.

GLM analysis of localizer tasks

The GLMs for analyses of motion and color localizers included the following regressors: R1) boxcar function for the motion/color phase, R2) boxcar function for the static/achromatic phase and R3–R8) movement regressors as covariates of no interest. Again, regressors were convolved with the canonical HRF and regressed against the BOLD signal in each voxel. First-level contrasts were constructed by separately weighting R1 and R2 over baseline, as well as R1 > R2. These contrasts were subsequently submitted to second-level random-effects group analyses (paired *t*-tests) for statistical evaluation.

ROI analyses

To investigate whether attribute values are systematically represented in Areas V4 and V5, a region of interest (ROI) analysis was performed in two steps. First, regions maximally responsive to functional localizers were identified. To this end, 9-mm spheres were centered at peak activations in left and right V5 derived from the group analysis of the motion localizer, and at peak activations in left and right V4 derived from the group analysis of the color localizer. Within each of these four spheres, peak activations of the respective localizer were identified for each individual participant, and 6-mm spheres were centered at these coordinates. Second, these individually adapted spherical ROIs were used to extract mean beta weights from parametric modulators for motion and color value (P1 and P2 of GLM1) for each participant. The extracted beta values were then analyzed for significance via a repeated-measures analysis of variance (rm-ANOVA) with the factors attribute value (motion/color), region (V5/V4) and hemisphere (left/right). Accordingly, the two-way interaction between attribute value and region indicates whether the respective attribute values are systematically represented in V5 and V4, and the three-way interaction between all factors further allows testing for a hemisphere-specific effect. The remaining interactions (hemisphere × region, hemisphere × attribute value) were included in the analysis, but were of no interest to our research questions.

Results

Behavioral results

In the learning task, the mean accuracy for learning color values (92.8 ± 3.8% s.d.) was higher than the mean accuracy for motion values (85.9 ± 6.6% s.d.; paired *t*-test, $t_{24} = -5.57$, $P < 0.001$) across all blocks. However, all participants successfully completed the

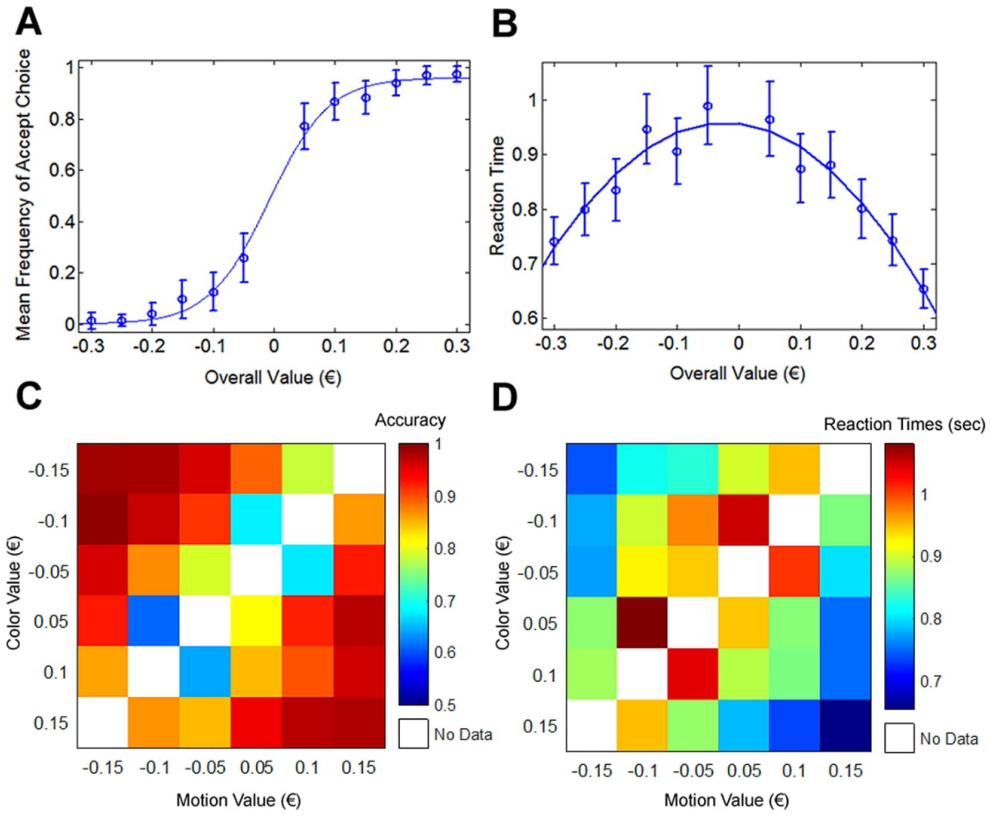

Fig. 2. Behavioral results. A) Mean frequency of accept choices plotted against overall values (fitted with a sigmoid function). Error bars represent the standard error of the mean. B) Mean reaction time plotted against overall values (fitted with a quadratic function). C) Mean accuracy for each combination of attribute value types. D) Mean reaction time for each combination of attribute value types.

learning task and achieved an accuracy of 98.1 (s.d. 2.5%) for both attributes during their last two blocks (mean of additional block number 4.4 ± 4 s.d.).

Participants showed a high level of accuracy in the main decision task (mean 87.5 ± 6.8% s.d.; one-sample *t*-test against chance level, $t_{24} = 27.56$, $P < 0.001$; Figure 2A and C) and responded on average 873 ± 157 ms (s.d.) after trial onset. Single-subject multiple linear regression models were used to estimate the effects of the overall value and the absolute attribute value difference of a stimulus (which are orthogonal to each other) on reaction times (Figure 2B and D). This analysis was performed to test whether overall value or attribute value similarity would facilitate choices. Regression coefficients showed a negative effect for overall value (one-sample *t*-test, $t_{24} = -4.27$, $P < 0.001$) and a positive effect for absolute attribute values difference ($t_{24} = 3.87$, $P < 0.001$), suggesting that participants were able to respond faster for stimuli with high overall values and high attribute value similarity. Furthermore, the effect of overall value and absolute attribute value difference on decision accuracy was tested using logistic regression models. In this analysis, absolute attribute value difference had a significant negative influence on accuracy (one-sample *t*-test, $t_{24} = -4.34$, $P < 0.001$), suggesting that participants were more accurate when attribute values were similar, but the effect of overall value was not significant ($t_{24} = -0.92$, $P = 0.367$).

Additional regression models were used to analyze differences in the processing of motion and color values. These variables were not included in the regression models above, since they are highly correlated with overall value ($r = 0.77$) and the absolute attribute value difference ($r = 1$), respectively. In a logistic regression model predicting task accuracy, neither motion (one-sample *t*-test, $t_{24} = -0.38$, $P = 0.704$) nor color value ($t_{24} = -0.42$, $P = 0.677$) were significant. In a linear regression model predicting reaction times, motion value was not significant ($t_{24} = -1.86$, $P = 0.076$), but color value had a significant negative impact ($t_{24} = -2.80$, $P = 0.009$). However, a paired *t*-test between the regression coefficients of motion and color value did not reveal a significant difference ($t_{24} = 0.65$, $P = 0.522$). This suggests that, on average, participants paid approximately equal attention to both attributes.

fMRI results

Value

The first goal of the fMRI analysis was to identify regions that are involved in valuation processes. Using GLM1, we were able to identify those regions that are specifically involved in the valuation of individual stimulus attributes (i.e. the particular values assigned to the stimulus' motion and color). We only report positive parametric modulations, because significant negative modulations were not observed. For motion value, activity during correct decision trials showed a significant positive parametric modulation in regions including PCC and left posterior inferior temporal gyrus (PIT; Figure 3A, Table 1), whereas a significant positive parametric modulation by color value was observed in ventral striatum and PCC (anterior to the PCC cluster for motion value; Figure 3B, Table 1). However, a direct comparison of motion- and color-related parametric effects using paired *t*-tests did not reveal significant differences at the whole-brain level. To further explore this relationship, the aforementioned regions were used as *post hoc* ROIs (PIT and PCC cluster of the motion value contrast, and ventral striatum and PCC cluster of the color value contrast, thresholded at $P_{unc} = 0.001$) and mean beta values for motion and color value within these ROIs were

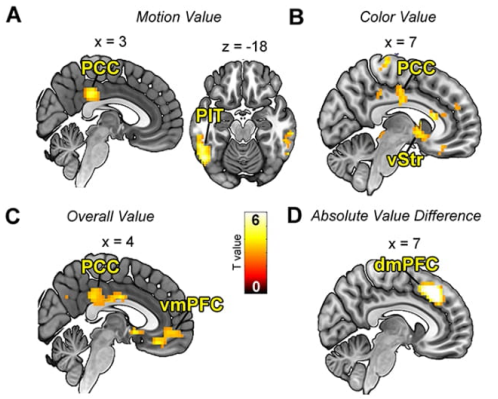

Fig. 3. Brain regions showing significant activations at the group level for A) motion value (GLM1), B) color value (GLM1), C) overall value (GLM2) and D) absolute attribute value differences (GLM1). For illustration purposes, *t*-maps (from second-level one-sample *t*-tests on parameter estimates of respective parametric modulators) are thresholded at $P_{unc} < 0.001$ with a cluster extent threshold of $k_E = 15$. Labeled clusters survive cluster-level FWE correction at $P_{FWE} < 0.05$.

compared via paired *t*-tests (Bonferroni-corrected *P*-value criterion of 0.05/4 = 0.0125). In line with the whole-brain results, the analysis did not reveal significant differences between motion- and color-related parametric effects [PIT: $t(24) = 1.5$, $P = 0.15$; PCC_{motion}: $t(24) = 0.4$, $P = 0.71$; PCC_{color}: $t(24) = -2.4$, $P = 0.02$; ventral striatum: $t(24) = -0.7$, $P = 0.5$]. While these results do not ultimately disprove the existence of attribute-specific valuation, they nevertheless indicate that motion and color value computations do not seem to recruit clearly separable, attribute-specific valuation modules within the current study.

Beyond the analyses of attribute valuation, GLM2 allowed to investigate which neural regions take part in the computation of overall value (i.e. the integrated value in terms of summed attribute values). We observed clusters showing a significant positive modulation of task-related activity by the stimulus' overall value in regions including left dorsolateral prefrontal cortex (dlPFC) and vmPFC (Figure 3C, Table 2). Compared to the analyses of individual attribute values, the results reveal partially overlapping neural regions (such as PCC and left PIT), which is to be expected based on the intrinsic correlation between overall and attribute-specific values. Due to this correlation, it cannot be directly assessed in a statistically valid way whether processes of overall and attribute-specific valuation show systematic neural differences. However, on the descriptive level, a significant cluster in vmPFC was only observed for parametric modulation of the stimulus' overall value, whereas this cluster was not significant in analyses of attribute-specific valuation. This pattern fits well to previous studies which suggested that vmPFC integrates information from multiple sources of evidence to an overall value (Hare *et al.*, 2009; Padoa-Schioppa and Cai, 2011; Levy and Glimcher, 2012; Rangel and Clithero, 2013).

Absolute attribute value difference

The absolute difference between motion and color values was used as a variable in GLM1 to identify comparator regions that estimate differences between attribute values (Basten *et al.*, 2010;

Table 1. Brain regions showing task-related activation in GLM1. Height threshold: $P_{unc} < 0.001$, $T_{24} = 3.47$. Extent threshold: $k_E = 15$ voxels. All activations survive whole-brain correction for multiple comparisons at the cluster level ($P_{FWE} < 0.05$). Abbreviation: MOG, middle occipital gyrus

Region	Side	MNI coordinates			k_E	T_{max}	P_{FWE} (cluster level)
		x	y	z			
Motion value							
Posterior inferior temporal gyrus	L	−51	−61	−18	171	5.74	0.003
Posterior cingulate cortex		0	−31	35	105	5.03	0.025
Superior parietal lobe	L	−39	−76	46	157	4.68	0.005
Color value							
Ventral striatum		−9	14	−3	127	6.17	0.007
Superior parietal lobe	R	18	−58	65	90	5.74	0.028
Posterior/middle cingulate cortex		−15	−22	35	171	5.49	0.002
Superior frontal sulcus	R	27	11	43	103	5.48	0.017
Anterior cingulate cortex		9	23	16	85	5.37	0.034
Dorsolateral prefrontal cortex	L	−48	35	16	167	3.91	0.002
Motion salience							
Posterior cingulate cortex		−9	−46	31	240	9.02	<0.001
LG/TPJ/MOG	R	21	−58	−14	1670	8.56	<0.001
Posterior/middle cingulate cortex		−3	−16	39	373	6.92	<0.001
Superior temporal gyrus/IPL/TPJ/mid-insular cortex/vmPFC	L	−54	−46	20	1504	5.87	<0.001
Superior frontal gyrus	L	−18	50	35	495	5.80	<0.001
Inferior frontal gyrus	R	48	38	1	76	5.75	0.022
Mid-insular cortex	R	33	−1	1	133	4.93	0.002
Superior frontal gyrus	R	15	41	46	90	4.50	0.011
Color salience							
IPL/middle temporal gyrus	L	−54	−13	−29	1610	9.08	<0.001
IPL/middle temporal gyrus	R	63	−52	20	1466	7.40	<0.001
Posterior cingulate cortex		−12	−49	31	322	6.59	<0.001
Inferior frontal gyrus	R	45	38	−10	84	6.50	0.009
Superior frontal gyrus/vmPFC	L	−18	29	58	345	6.06	<0.001
Posterior cingulate cortex		0	−16	39	215	5.91	<0.001
Hippocampus/putamen	L	−21	−4	5	262	5.67	<0.001
Fusiform gyrus	R	42	−55	−18	189	5.60	<0.001
Cerebellum	R	21	−82	−33	84	5.35	0.009
Inferior frontal gyrus	L	−51	32	−14	70	5.27	0.02
Middle frontal gyrus	L	−33	29	46	111	5.05	0.002
Absolute attribute value difference							
Dorsomedial prefrontal cortex		0	32	43	287	8.85	<0.001
Inferior frontal gyrus	R	33	26	−6	97	7.78	0.012
Middle frontal gyrus	R	45	14	43	114	5.89	0.006
Inferior frontal gyrus	L	−33	20	1	72	5.75	0.038
Superior frontal gyrus	R	27	17	58	92	5.66	0.015
IPL	R	48	−52	58	113	5.56	0.006

Philiastides *et al.*, 2010). By this means, we observed a significant positive modulation of task-related hemodynamic activity within the dorsomedial prefrontal cortex (dmPFC; Figure 3D, Table 1).

Salience

A significant positive modulation of task-related activity by motion salience (GLM1) was observed in regions including bilateral temporoparietal junction (TPJ), right lingual gyrus (LG) and PCC (Figure 4A, Table 1), whereas a positive modulation by color salience (GLM1) was found in bilateral inferior parietal lobe (IPL), bilateral anterior temporal cortex (AT) and PCC (see Figure 4B and Table 1). Direct comparisons of motion and color salience effects by paired *t*-tests revealed no activations surviving our significance criterion, suggesting that neither attribute salience had a significantly stronger effect nor relies on specialized processing modules in our current decision task. For modulation by overall stimulus salience (GLM2), partially overlapping regions including bilateral TPJ and PIT (Table 2) were observed.

Attribute processing in V5/V4

Motion and color localizers were conducted to identify regions specifically involved in the processing of physical stimulus attributes (i.e. motion and color), independent of valuation processes (Figure 5A and B). As expected, the motion localizer revealed significant activation in the group analysis within bilateral V5 (left V5: $T_{24} = 5.89$, $k = 107$, cluster-level $P_{FWE} = 0.036$, $x = -48$, $y = -67$, $z = 5$; right V5: $T_{24} = 6.28$, $k = 109$, cluster-level $P_{FWE} = 0.034$, $x = 42$, $y = -64$, $z = 9$), and the color localizer within bilateral V4 (left V4: $T_{24} = 8.66$, $k = 753$, cluster-level $P_{FWE} < 0.001$, $x = -30$, $y = -73$, $z = -14$; right V4: $T_{24} = 8.63$, $k = 753$, cluster-level $P_{FWE} < 0.001$, $x = 30$, $y = -73$, $z = -14$).

ROI analyses were performed to test whether regions specialized in processing the physical attributes of the stimuli (i.e. V5 for motion and V4 for color) also compute the respective attribute values. For this purpose, beta estimates from parametric modulations by motion and color value were extracted from V5 and V4 in both hemispheres, and entered into an rm-ANOVA with factors attribute value (motion/color), region (V5/V4) and hemisphere (left/right). Neither the two-way interaction between

Table 2. Brain regions showing task-related activation in GLM2. Height threshold: $P_{unc} < 0.001$, $T_{24} = 3.47$. Extent threshold: $k_E = 15$ voxels. All activations survive whole-brain correction for multiple comparisons at the cluster level ($P_{FWE} < 0.05$)

		MNI coordinates					
Region	**Side**	**x**	**y**	**z**	k_E	T_{max}	P_{FWE} **(cluster level)**
Overall value							
Dorsolateral prefrontal cortex	L	−51	29	24	136	6.12	0.023
Superior parietal lobe	L	−27	−76	46	187	5.30	0.007
Posterior inferior temporal gyrus	R	60	−25	−25	216	5.27	0.004
Posterior cingulate cortex		−21	−22	35	499	5.14	<0.001
Inferior temporal gyrus	L	−54	−58	−18	215	4.97	0.004
Ventromedial prefrontal cortex		0	35	−18	212	4.74	0.004
Overall salience							
Inferior temporal gyrus/TPJ	R	48	−34	1	1607	6.73	<0.001
TPJ/postcentral gyrus	L	−51	−22	35	415	6.11	<0.001
Inferior temporal gyrus	L	−33	−46	−18	181	5.87	<0.001
Occipital lobe	L	−33	−88	1	302	5.46	<0.001
Rostral anterior cingulate cortex		18	32	−10	108	5.24	0.007
Hippocampus	R	30	−10	−18	96	4.92	0.012
Hippocampus	L	−27	−16	−14	81	4.53	0.024
Middle temporal gyrus	L	−39	−64	5	69	4.11	0.042

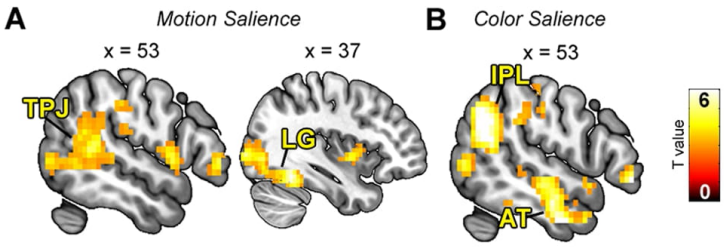

Fig. 4. Brain regions showing significant group-level activations for A) motion salience and B) color salience (GLM1). For illustration purposes, *t*-maps (from second-level one-sample *t*-tests on parameter estimates of the respective parametric modulator) are thresholded at $P_{unc} < 0.001$ with a cluster extent threshold of $k_E = 15$. Labeled clusters survive cluster-level FWE correction at $P_{FWE} < 0.05$.

attribute value and region [$F(1, 24) = 0.51$, $P = 0.48$] nor the three-way interaction between attribute value, region and hemisphere [$F(1, 24) = 0.53$, $P = 0.47$] were significant, which does not support the hypothesis that attribute values are systematically processed in V5 and V4. Furthermore, there were no significant effects for the main effects or the remaining interactions of no interest [attribute value: $F(1, 24) = 0.01$, $P = 0.93$; region: $F(1, 24) = 3.15$, $P = 0.09$; hemisphere: $F(1, 24) = 0.12$, $P = 0.73$; hemisphere × region: $F(1, 24) = 0.51$, $P = 0.48$; hemisphere × attribute value: $F(1, 24) = 2.57$, $P = 0.12$].

Discussion

The current experiment investigated how attribute values are processed and integrated in the brain during decision-making. In particular, we tested the competing hypotheses whether (i) distinct regions that specialize in the processing of a particular physical attribute also compute the respective attribute value (Philiastides *et al.*, 2010; Lim *et al.*, 2013), or whether (ii) attribute values are collectively processed in a general valuation network consisting of vmPFC, PCC and ventral striatum (Levy and Glimcher, 2012; Bartra *et al.*, 2013; Clithero and Rangel, 2014). To differentiate between these hypotheses, we used a choice task in which monetary values were associated with the attributes motion and color, whose physical properties are known to be processed in specialized brain regions, namely Area V5 for motion (Watson *et al.*, 1993) and Area V4 for color (McKeefry and Zeki, 1997).

Whole-brain analyses showed that activity in PCC and ventral striatum correlated with color value, whereas activity related to motion value occurred in PCC and left PIT. In a direct comparison, we did not detect any region that had a specifically stronger representation of one compared to the other attribute value. This lack of specificity suggests that the computation of particular attribute values is not realized within specialized cortical modules, but is instead accomplished in a dynamic manner within a network comprising PCC, PIT and ventral striatum. Consistent with this idea, ROI analyses did not reveal a systematic representation of motion value in V5 and color value in V4, which does not support the hypothesis that attribute values and physical properties of attributes are computed in the same regions. Taken together, our data thus provide concordant evidence for the hypothesis that values are homogeneously processed within a general valuation network.

In contrast to our results, previous studies have supported the hypothesis that attribute values are computed in attribute-specific regions. In an experiment by Lim *et al.* (2013), participants had to evaluate t-shirts based on how much they liked both the appearance and meaning of Korean symbols that were printed on them. The authors found that activity in fusiform gyrus correlated with visual values, whereas activity in superior temporal gyrus correlated with semantic values. Furthermore, a study by

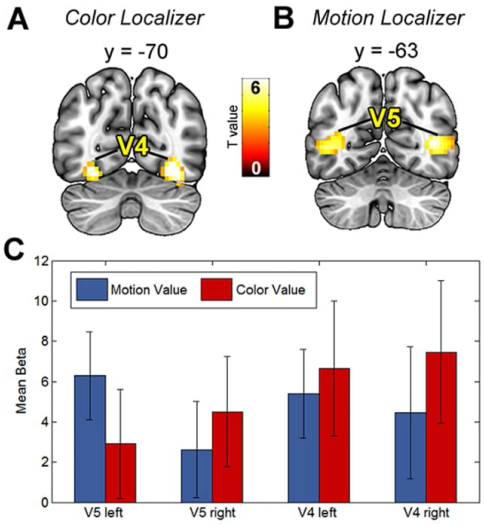

Fig. 5. Activations of localizer tasks in A) bilateral V4 (color localizer) and B) bilateral V5 (motion localizer). For illustration purposes, t-maps are thresholded at $P_{unc} < 0.001$ with a cluster extent threshold of $k_E = 15$. All clusters survive cluster-level FWE correction at $P_{FWE} < 0.05$. C) Mean beta estimates of parametric modulation by motion and color value in bilateral V5 and V4 (regions adapted to single-subject peaks of localizers; for details see Methods section). As the corresponding results of the rm-ANOVA with factors attribute value (motion/color), region (V5/V4) and hemisphere (left/right) indicate, the data do not reveal a systematic representation of motion and color value in V5 and V4, respectively.

Philiastides *et al.* (2010) demonstrated in a probabilistic choice task that activity in fusiform face area correlates with the value of face stimuli, and parahippocampal place area with the value of house stimuli. For both studies, the brain regions correlating with attribute values have also been associated with the processing of physical attribute properties, which is not confirmed by data of the present experiment.

This discrepancy could originate from several factors. First, our design allowed us to differentiate value- and salience-related effects for each attribute, which was not possible in the experiment by Philiastides *et al.* (2010). As such, value correlations in fusiform face area and parahippocampal place area might originate from differences in salience instead of value. Second, the motion and color attributes in our paradigm are robustly associated with well-defined regions in Areas V5 and V4. Arguably, this connection is less straightforward in the aforementioned studies, in particular for the semantic attribute of the study by Lim *et al.* (2013), because the effects are more distributed across brain regions. Therefore, it is less clear whether these results actually arise from attribute-specific regions, as it is more difficult to precisely determine ROIs. Third, we used abstract, novel stimuli in our paradigm, whereas the other studies used more common stimuli (faces, houses and t-shirts). A possible mechanism for neural attribute valuation could be that novel stimuli are first computed in a domain-general network, in which attribute values are processed homogeneously, but if stimuli are more familiar and need to be evaluated frequently, the computation of attribute values shifts toward the respective attribute-specific regions. Thus, the processing of attribute values might change as a function of learning, which could reconcile the effects of our experiment with previous studies.

Apart from analyses of single attribute values, we also investigated brain activity related to the overall value of stimuli and found correlations with activity in, among other regions, left dlPFC and vmPFC. Due to the correlation between attribute values and overall values, our results show an overlap in their effects on the neural level, which cannot be easily disentangled on statistical grounds. Nonetheless, it is of note that whole-brain analyses of overall values revealed a significant correlation for a cluster in vmPFC, which was not found in our analyses of attribute values (neither for motion nor for color value). This is in line with the view that vmPFC is one of the main regions that represent integrated value signals, which previous studies have argued for (Hare *et al.*, 2009; Padoa-Schioppa and Cai, 2011; Levy and Glimcher, 2012; Rangel and Clithero, 2013; Pelletier and Fellows, 2019). From this perspective, decision problems are deconstructed into subproblems (such as the computation of single attribute values) that are resolved in distributed brain regions, and the available evidence is then ultimately combined to a unified value signal in vmPFC, which is consistent with our findings.

Furthermore, we analyzed brain activity related to the absolute difference between simultaneously presented attribute values. This variable could be an indicator for regions that compare attribute values and thus estimate difference signals that drive the process of value integration. Consistent with findings from other studies (Basten *et al.*, 2010; Hare *et al.*, 2011), we found that absolute differences between attribute values were mainly associated with activity in dmPFC. Moreover, the behavioral data revealed that larger differences between attribute values were associated with longer reaction times and a lower choice accuracy, which suggests that these trials were more difficult for participants. One explanation for this relationship could be that larger attribute value differences induce a higher need for value integration and thus demand more cognitive resources. When attribute value differences are small, each attribute value in isolation already provides a good estimate for a stimulus' overall value, and value integration is less important for effective decision-making. But when attribute value differences are large, it is crucial to integrate the underlying attribute values into a representative overall value to make optimal decisions. Thus, dmPFC could be responsible for the estimation of attribute value difference signals that indicate the need for value integration and form the basis for the computation of overall values.

In addition to analyses of value-related effects, we also explored salience-related effects. Salience refers to the subjective importance of a stimulus (Maunsell, 2004; Zink *et al.*, 2006; Litt *et al.*, 2011; Leathers and Olson, 2012; Kahnt and Tobler, 2013; Kahnt *et al.*, 2014) that guides attention and prioritizes the processing of particular stimuli over less important ones. In the context of this experiment, both high positive as well as high negative attribute values have a high salience, because both have a large impact on decisions. The more salient an attribute value is (i.e. the higher the absolute attribute value is), the less likely it is that the other attribute value will outweigh its influence. Hence, when time is limited, it is efficient to selectively process attributes with higher salience and pay less attention to attributes with lower salience (Kahnt and Tobler, 2013). We investigated brain regions that could realize such a selection mechanism by analyzing neural correlates of absolute attribute values for motion and color. As a result, motion salience was associated with bilateral TPJ, right LG and PCC, whereas bilateral IPL, AT and PCC were correlated with color salience. This network could therefore be

responsible for allocating attentional resources and assigning priority to the attribute that is most relevant in a given situation. Notably, activity in PCC was related to attribute value, attribute salience as well as overall value, which suggests that it plays a central part for neural information processing in value-based decision-making.

There are some limitations that have to be taken into account in the interpretation of our findings. First, our color and motion attributes have different perceptual properties, but they were both encoded via monetary values. However, in real-world decisions, the attributes that have to be combined often have different types of value encoding. For example, in the evaluation of a car, the price has a monetary value encoding, the design an aesthetic value encoding, and the safety rating is encoded as risk to our health. Arguably, the attributes in our experiment still qualify as different attributes, in the same way that the interior and exterior design of a car can be evaluated independently, even though both have an aesthetic value encoding. But it is an open question in what way our results would differ if the attributes had different types of value encodings. Second, while we did not observe representations of attribute values in V4 or V5, we cannot decisively rule out that possibility, since strong conclusions from null findings are not statistically justified. Third, we modeled events in our GLM for the duration of the whole choice period, but it is possible that explicitly modeling different stages of the decision-making process (such as attribute value identification and attribute integration) could reveal stronger and more precise effects. There is also evidence that decision attributes are processed at different rates (e.g. Sullivan *et al.*, 2014), which could have played a role in our task as well. For example, while color can, in principle, be perceived immediately, motion perception requires the observation of visual frames during a longer time span. The low temporal resolution of fMRI makes it challenging to better incorporate factors like these, but a dedicated experimental design and methods like the combination of electroencephalography and fMRI could help to disentangle subprocesses of decision-making in more detail.

So far, most studies on attribute integration in value-based decision-making (Hare *et al.*, 2009; Basten *et al.*, 2010; Philiastides *et al.*, 2010; Kahnt *et al.*, 2011; Park *et al.*, 2011; Lim *et al.*, 2013; Hutcherson *et al.*, 2015) have argued for a feed-forward model, in which attribute values are separately computed in dedicated regions, and only afterward integrated in a unifying region like vmPFC. However, some studies have argued for a more flexible model (Hunt *et al.*, 2014; Siegel *et al.*, 2015). In this view, attribute values are determined in a dynamic process that includes continuous feed-forward and feedback projections as well as competitive inhibition between attributes. Hence, the model proposes that attribute values are not computed sequentially and in isolation. Instead, there is a constant exchange of information in which value predictions are continuously updated and re-evaluated, dependent on concurrent neural computations that process factors like salience, memory or affective states. Since we observed a uniform neural network for different types of attribute values, the results of our study are in line with the latter model and support the idea that attribute values are computed in an interdependent and contextualized manner.

Acknowledgements

We thank Christine Stelzel and Ilya Veer for valuable feedback during discussions, and Gabriele Bellucci for his help in data acquisition.

Funding

This work was supported by the Berlin School of Mind and Brain, Charité Berlin, Bernstein Computational Neuroscience Program of the German Federal Ministry of Education and Research (grant 01GQ1001C), the German Research Foundation within the Collaborative Research Center 'Volition and Cognitive Control: Mechanisms, Modulations, Dysfunctions' (DFG grant SFB 940/1 2014 and 2015), the Flemish Science Foundation and the European Union's Horizon 2020 Research and Innovation Program (Marie Skłodowska-Curie grant 665501).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

A.M., C.M.S., V.U.L., H.R.H. and H.W. designed the experiment. A.M. programmed the experiment, acquired the data and analyzed the results. C.M.S., V.U.L., L.M.P., D.W., H.R.H. and H.W. provided methodological and conceptual advice for the analysis and discussion of the data. The manuscript was written by A.M. and edited by all authors.

References

Bartra, O., McGuire, J.T., Kable, J.W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, **76**, 412–27.

Basten, U., Biele, G., Heekeren, H.R., Fiebach, C.J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences USA*, **107**, 21767–72.

Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, **10**, 433–6.

Clithero, J., Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, **9**, 1289–302.

Dale, A.M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping*, **8**, 109–14.

de Berker, A.O., Kurth-Nelson, Z., Rutledge, R.B., Bestmann, S., Dolan, R.J. (2019). Computing value from quality and quantity in human decision-making. *Journal of Neuroscience*, **39**, 163–76.

Gallivan, J.P., Chapman, C.S., Wolpert, D.M., Flanagan, J.R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, **19**, 519–34.

Gold, J.I., Shadlen, M.N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, **30**, 535–74.

Hanks, T.D., Summerfield, C. (2017). Perceptual decision making in rodents, monkeys, and humans. *Neuron*, **93**, 15–31.

Hare, T.A., Camerer, C.F., Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, **324**, 646–8.

Hare, T.A., Schultz, W., Camerer, C.F., O'Doherty, J.P., Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences USA*, **108**, 18120–5.

Hunt, L.T., Dolan, R.J., Behrens, T.E.J. (2014). Hierarchical competitions subserving multi-attribute choice. *Nature Neuroscience*, **17**, 1613–22.

Hutcherson, C.A., Montaser-Kouhsari, L., Woodward, J., Rangel, A. (2015). Emotional and utilitarian appraisals of moral dilemmas

are encoded in separate areas and integrated in ventromedial prefrontal cortex. *Journal of Neuroscience*, **35**, 12593–605.

Kable, J.W., Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, **10**, 1625–33.

Kable, J.W., Glimcher, P.W. (2009). The neurobiology of decision: consensus and controversy. *Neuron*, **63**, 733–45.

Kahnt, T., Heinzle, J., Park, S.Q., Haynes, J.-D. (2011). Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *NeuroImage*, **56**, 709–15.

Kahnt, T., Park, S.Q., Haynes, J.-D., Tobler, P.N. (2014). Disentangling neural representations of value and salience in the human brain. *Proceedings of the National Academy of Sciences USA*, **111**, 5000–5.

Kahnt, T., Tobler, P.N. (2013). Salience signals in the right temporoparietal junction facilitate value-based decisions. *Journal of Neuroscience*, **33**, 863–9.

Kanwisher, N., McDermott, J., Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, **17**, 4302–11.

Leathers, M.L., Olson, C.R. (2012). In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science*, **338**, 132–5.

Levy, D.J., Glimcher, P.W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, **22**, 1027–38.

Lim, S.-L., O'Doherty, J.P., Rangel, A. (2013). Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *Journal of Neuroscience*, **33**, 8729–41.

Litt, A., Plassmann, H., Shiv, B., Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cerebral Cortex*, **21**, 95–102.

Ludwig, V.U., Stelzel, C., Krutiak, H., *et al.* (2014). The suggestible brain: posthypnotic effects on value-based decision-making. *Social Cognitive and Affective Neuroscience*, **9**, 1281–8.

Maunsell, J.H.R. (2004). Neuronal representations of cognitive state: reward or attention? *Trends in Cognitive Sciences*, **8**, 261–5.

McKeefry, D.J., Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, **120**, 2229–42.

O'Doherty, J.P. (2014). The problem with value. *Neuroscience and Biobehavioral Reviews*, **43**, 259–68.

Padoa-Schioppa, C., Cai, X. (2011). The orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. *Annals of the New York Academy of Sciences*, **1239**, 130–7.

Park, S.Q., Kahnt, T., Rieskamp, J., Heekeren, H.R. (2011). Neurobiology of value integration: when value impacts valuation. *Journal of Neuroscience*, **31**, 9307–14.

Pelletier, G., Aridan, N., Fellows, L.K., Schonberg, T. (2021). A preferential role for ventromedial prefrontal cortex in assessing the value of the whole in multiattribute object evaluation. *Journal of Neuroscience*, **41**, 5056–68.

Pelletier, G., Fellows, L.K. (2019). A critical role for human ventromedial frontal lobe in value comparison of complex objects based on attribute configuration. *Journal of Neuroscience*, **39**, 4124–32.

Persichetti, A.S., Aguirre, G.K., Thompson-Schill, S.L. (2015). Value is in the eye of the beholder: early visual cortex codes monetary value of objects during a diverted attention task. *Journal of Cognitive Neuroscience*, **27**, 893–901.

Peters, J., Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research*, **213**, 135–41.

Philiastides, M.G., Biele, G., Heekeren, H.R. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences USA*, **107**, 9430–5.

Rangel, A., Camerer, C., Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, **9**, 545–56.

Rangel, A., Clithero, J. (2013). The computation of stimulus values in simple choice. In: Glimcher, P., Fehr, E., editors. *Neuroeconomics: Decision Making and the Brain*, 2nd edn, San Diego: Academic Press, 125–48.

Sanfey, A.G., Loewenstein, G., McClure, S.M., Cohen, J.D. (2006). Neuroeconomics: cross-currents in research on decision-making. *Trends in Cognitive Sciences*, **10**, 108–16.

Serences, J.T. (2008). Value-based modulations in human visual cortex. *Neuron*, **60**, 1169–81.

Siegel, M., Buschman, T.J., Miller, E.K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science*, **348**, 1352–5.

Sullivan, N., Hutcherson, C., Harris, A., Rangel, A. (2014). Dietary self-control is related to the speed with which attributes of healthfulness and tastiness are processed. *Psychological Science*, **26**, 122–34.

Suzuki, S., Cross, L., O'Doherty, J.P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature Neuroscience*, **20**, 1780–6.

Vaidya, A.R., Sefranek, M., Fellows, L.K. (2018). Ventromedial frontal lobe damage alters how specific attributes are weighed in subjective valuation. *Cerebral Cortex*, **28**, 3857–67.

Watson, J.D.G., Myers, R., Frackowiak, R.S.J., *et al.* (1993). Area V5 of the human brain: evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cerebral Cortex*, **3**, 79–94.

Zhang, Z., Fanning, J., Ehrlich, D.B., Chen, W., Lee, D., Levy, I. (2017). Distributed neural representation of saliency controlled value and category during anticipation of rewards and punishments. *Nature Communications*, **8**, 1–14.

Zink, C.F., Pagnoni, G., Chappelow, J., Martin-Skurski, M., Berns, G.S. (2006). Human striatal activation reflects degree of stimulus saliency. *NeuroImage*, **29**, 977–83.

Publication 2

Magrabi, A., Beck, A., Schad, D.J., Stoppel, C.M, Lett, T.A., Charlet, K., Kiefer, F., Heinz, A. & Walter, H. (2022). Alcohol Dependence decreases Functional Activation of the Caudate Nucleus during Model-Based Decision Processes. *Alcoholism: Clinical and Experimental Research*, *46(5)*, 749–58. http://doi.org/10.1111/acer.14812

Impact Factor (2019): 3,035

Platz 14 von 42 im Fachgebiet *Substance Abuse* (Top 33%)

Publication Rank

Journal Data Filtered By: **Selected JCR Year: 2019** Selected Editions: SCIE,SSCI Selected Categories: **"SUBSTANCE ABUSE"** Selected Category Scheme: WoS
Gesamtanzahl: 42 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfactor Score
1.	TOBACCO CONTROL	9,207	6.726	0.018580
2.	ADDICTION	19,861	6.343	0.030820
3.	ADDICTION	19,861	6.343	0.030820
4.	INTERNATIONAL JOURNAL OF DRUG POLICY	5,658	4.444	0.014970
5.	Alcohol Research-Current Reviews	899	4.214	0.002220
6.	ADDICTION BIOLOGY	4,329	4.121	0.008280
7.	NICOTINE & TOBACCO RESEARCH	10,026	4.079	0.020870
8.	DRUG AND ALCOHOL DEPENDENCE	20,269	3.951	0.040630
9.	Harm Reduction Journal	1,512	3.818	0.003540
10.	ADDICTIVE BEHAVIORS	13,899	3.645	0.022950
11.	ADDICTIVE BEHAVIORS	13,899	3.645	0.022950
12.	Addiction Science & Clinical Practice	646	3.088	0.001690
13.	JOURNAL OF SUBSTANCE ABUSE TREATMENT	5,696	3.083	0.010000
14.	ALCOHOLISM-CLINICAL AND EXPERIMENTAL RESEARCH	14,315	3.035	0.015690
15.	Journal of Addiction Medicine	1,672	3.014	0.005140
16.	AMERICAN JOURNAL OF DRUG AND ALCOHOL ABUSE	2,780	2.925	0.004250
17.	JOURNAL OF GAMBLING STUDIES	2,877	2.836	0.003710
18.	PSYCHOLOGY OF ADDICTIVE BEHAVIORS	5,301	2.780	0.007640
19.	Substance Abuse	1,595	2.652	0.004540

Received: 12 October 2021 | Accepted: 15 March 2022
DOI: 10.1111/acer.14812

ALCOHOLISM CLINICAL & EXPERIMENTAL RESEARCH

ORIGINAL ARTICLE

Alcohol dependence decreases functional activation of the caudate nucleus during model-based decision processes

Amadeus Magrabi[1,2] | Anne Beck[3] | Daniel J. Schad[3] | Tristram A. Lett[1] | Christian M. Stoppel[1] | Katrin Charlet[1,4] | Falk Kiefer[5] | Andreas Heinz[1] | Henrik Walter[1,2]

[1]Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Berlin, Germany
[2]Berlin School of Mind and Brain, Humboldt-Universität zu Berlin, Berlin, Germany
[3]HMU Health and Medical University Potsdam, Potsdam, Germany
[4]Section on Clinical Genomics and Experimental Therapeutics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland, USA
[5]Department of Addictive Behavior and Addiction Medicine, Medical Faculty Mannheim, Central Institute of Mental Health, Heidelberg University, Mannheim, Germany

Correspondence
Amadeus Magrabi, Division of Mind and Brain Research, Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany.
Email: amadeus.magrabi@gmail.com

Funding information
Berlin School of Mind and Brain; Charité Berlin; Ministry of Education and Research, Grant/Award Number: 01ZX1311E/e and 01EE1406A; Canadian Institute of Health Research; German Research Foundation, Grant/Award Number: Exc 257, HE2597/14-2, 402170461, TRR 265 and 1936/1-1

Abstract

Background: Impaired decision making, a key characteristic of alcohol dependence (AD), manifests in continuous alcohol consumption despite severe negative consequences. The neural basis of this impairment in individuals with AD and differences with known neural decision mechanisms among healthy subjects are not fully understood. In particular, it is unclear whether the choice behavior among individuals with AD is based on a general impairment of decision mechanisms or is mainly explained by altered value attribution, with an overly high subjective value attributed to alcohol-related stimuli.

Methods: Here, we use a functional magnetic resonance imaging (fMRI) monetary reward task to compare the neural processes of model-based decision making and value computation between AD individuals (n = 32) and healthy controls (n = 32). During fMRI, participants evaluated monetary offers with respect to dynamically changing constraints and different levels of uncertainty.

Results: Individuals with AD showed lower activation associated with model-based decision processes in the caudate nucleus than controls, but there were no group differences in value-related neural activity or task performance.

Conclusions: Our findings highlight the role of the caudate nucleus in impaired model-based decisions of alcohol-dependent individuals.

KEYWORDS
alcohol, computational modeling, decision making, fMRI, value

INTRODUCTION

Alcohol dependence (AD) is a highly prevalent psychiatric disorder, accounting for about 3 million deaths per year worldwide (World Health Organization, 2018). It is characterized by a loss of control over the consumption of alcohol, a negative emotional state (such as anxiety) during withdrawal, and continued drinking despite repeated harmful consequences (Everitt & Robbins, 2005; Koob & Volkow, 2010).

Regarding the neurobiological basis of addiction, multiple studies have investigated the neural response of AD patients to alcoholic stimuli and related conditioning processes (Beck et al., 2012; Chase et al., 2011; Kühn & Gallinat, 2011; Schad et al., 2019) indicating an increased incentive salience and value attribution to those cues (Wrase et al., 2007) as well as aberrant processing of nonalcoholic stimuli, in terms of diminished responsiveness toward nonalcoholic reinforcements (Goldstein & Volkow, 2011; Luijten et al., 2017; Schacht et al., 2013; Sebold et al., 2017). However, it was also observed that substance use disorders were associated with increased limbic system sensitivity to reward and loss delivery (Bjork et al., 2008).

A further important and putatively related aspect of addictive disorders are maladaptive choices that oppose the explicitly stated desires of the patients, such as continuing consumption despite the desire to abstain. Here, two extensively explored components of decision making are of importance that have been characterized using computational modeling methods: (1) a flexible planning system integrating all available information to find the most appropriate decision and considering the consequences of actions: the goal-directed or model-based system and (2) a rigid habitual system that simply repeats actions that were rewarded in the past without taking a model of the environment into account: the habitual or model-free system (e.g., see Sebold et al., 2014). In addictive behaviors, it was observed that there is a shift from goal-directed (i.e., model-based) toward habitual (model-free) decision making (e.g., Voon et al., 2017). On the neuronal level, it has been suggested that AD develops through a systematic shift in the neural systems that regulate behavior, with increased involvement of the dorsolateral striatum/putamen (in rodents/humans) controlling habitual behavior, and decreased involvement of the dorsomedial striatum/caudate controlling flexible and goal-directed behavior (Corbit et al., 2012; DePoy et al., 2013; Everitt & Wolf, 2002; Furlong et al., 2014; Gahnstrom & Spiers, 2020; Geerts et al., 2020; Sharpe et al., 2019; Vollstädt-Klein et al., 2010). In rodent studies, it has been shown that lesions of the dorsomedial striatum (comparable to human's caudate) block goal-directed behavior (Yin et al., 2005), while in contrast, lesions of the dorsolateral striatum (comparable to human's putamen) disrupt habit formation (Yin et al., 2004). Thus, the capacity for decision making in terms of goal-directed behavior seems to be a core function affected in AD (Mollick & Kober, 2020; Sebold et al., 2014). In particular, AD patients continuously choose to consume alcohol and neglect the long-term consequences of sustained consumption on their physical and psychological health (Amlung et al., 2017; Phung et al., 2019).

Other studies suggest that the choice behavior of AD patients (Kamarajan et al., 2020; Rubio et al., 2008; Virkkunen, 1994) is based on an overactive neural value system (Arcurio et al., 2015; Goldstein & Volkow, 2011; Seo et al., 2013), which has been associated with ventromedial prefrontal cortex (vmPFC; Bartra et al., 2013; Clithero & Rangel, 2014; Lee et al., 2021). However, it is not well understood how exactly and under what conditions these shifts in neural information processing can occur.

Here, we developed a sequential decision-making task to investigate this process in AD patients and healthy control subjects via functional magnetic resonance imaging (fMRI). To detect behavioral and neural differences between AD patients and controls, we designed a task that specifically relies on the ability to flexibly adapt choices to multiple factors and their associated consequences. To further contribute to the core question of whether AD affects decision networks in general, beyond choices that are specifically related to alcohol, we used a task that relies on monetary incentives instead of alcoholic stimuli.

In this decision task, participants had to decide whether to accept or reject various monetary offers that were presented to them. Crucially, for each experimental block of 20 offers, participants were only allowed to accept a maximum of 5 offers. To make optimal choices and maximize the probability of accepting only the highest offers in a block, participants thus had to consider three factors: (1) the value of the current offer, (2) the number of offers that can still be accepted before reaching the limit, and (3) the number of offers that are remaining in the current block. These parameters were included in a decision model (Economides et al., 2014) and computed for the choice data. Parameter estimates of the model were then used as parametric modulators in the analysis of the functional MRI data to identify brain regions that compute model-based decision processes, and to test for putative differences between AD patients and controls.

Based on previous studies, we defined regions of interest (ROIs) and hypothesized that AD patients would show (1) a decreased representation of model-based decision processes in caudate nucleus and (2) an increased representation of decision value in vmPFC.

MATERIALS AND METHODS

Participants

The experimental sample consisted of 32 detoxified AD patients and 32 healthy control subjects (Table 1). The sample was acquired as part of the *National Genome Research Network* (Spanagel et al., 2010) at *Charité—Universitätsmedizin Berlin*. All subjects were right-handed, had a normal or corrected-to-normal vision, and provided informed consent before participation. AD patients were diagnosed with AD according to DSM-IV and ICD-10

TABLE 1 Descriptive statistics of alcohol-dependent patients and healthy control subjects

	AD patients (22 male, 10 female)			Control subjects (23 male, nine female)			Group difference	
	Mean	SD	Missing data	Mean	SD	Missing data	*p*	*T*
Lifetime drinking history (consumption in kg)	909.9	885.6	2	83.8	99	0	<0.01[a]	−5.2
Education level[a]	*Median:* 2	*IQR:* 1	0	*Median:* 3	*IQR:* 1	0	*M-W U-test:* 0.137	*Z-value:* −1.49
Age	46.5	8.9	0	38.9	10.5	0	<0.01[a]	−3.13
Pack years of cigarette consumption[b]	21.6	18.1	0	9.6	14.7	0	<0.01[a]	−2.9
Number of smokers	25	–	0	10	–	0	–	–
Duration of dependence (years)	6.6	5.6	4	–	–	–	–	–
Age of dependence onset	40.8	7.8	4	–	–	–	–	–
Percentage of invalid trials[c]	24.27	9.40	0	25.12	6.91	0	0.68	0.41

Abbreviations: AD, alcohol dependence; IQR, interquartile range; M-W *U*-test, Mann–Whitney *U*-test; SD, standard deviation.

[a]Ordinal variable corresponding to education levels in the German school system (from lowest to highest): 0 = no graduation, 1 = *Hauptschule*, 2 = *Realschule*, 3 = *Abitur*.

[b]Definition of pack years: ((number of consumed cigarettes per day/18) × c number of years smoked), with 18 as the standard amount of cigarettes in one pack.

[c]Definition of invalid trials: maximum number of offers were already accepted or no response was given in time.

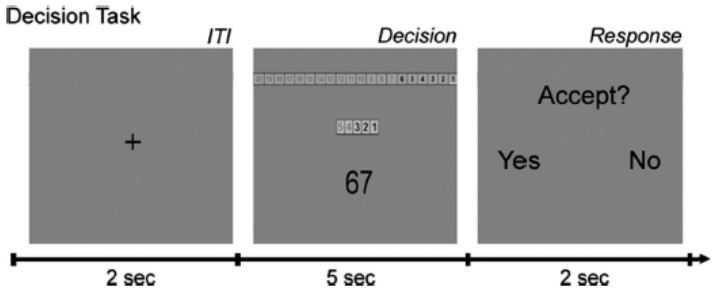

FIGURE 1 Trial structure of the decision task. In the decision phase, participants were presented with offer values (in cents) as well as indicators for the remaining number of trials (maximum of 20 for each of the nine blocks) and for the number of offers that can still be accepted (maximum of five in each block). In the response phase, participants specified their choice to accept or reject an offer with their left or right index finger. ITI, inter-trial interval

(Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I); First et al., 2001) and completed medically supervised detoxification (mean detoxification days when data was acquired: 12.71 ± 4.93 SD). Exclusion criteria for all participants were DSM-IV Axis-I disorders (excluding alcohol and nicotine dependence in AD patients, and excluding only nicotine dependence in control subjects), use of cannabinoids, benzodiazepines, barbiturates, cocaine, amphetamines, opiates (tested by urine screening) or psychotropic medication, claustrophobia, epilepsy, other neurological or psychiatric illnesses, and pregnancy. Data sets of subjects that showed excessive head motion in the scanner or failed to follow task instructions (i.e., subjects that continuously tried to accept offers even though the acceptance limit was already reached) were excluded from further analyses. The amount of lifetime alcohol consumption was assessed using the Lifetime Drinking History (Skinner & Sheu, 1982) and the experiment was approved by the local ethics committee.

Group comparisons revealed a difference with respect to age (Table 1), which was controlled for by including age as a covariate in statistical analyses of group differences. Further, as is commonly found in studies on AD patients (Batel et al., 2006), patients showed increased smoking behavior (Table 1) as indicated by pack years of cigarette consumption (definition of pack years: (number of consumed cigarettes per day/18) × number of years smoked; with 18 as the standard amount of cigarettes in one pack). Since cigarette consumption was significantly correlated with lifetime alcohol intake in AD patients ($r = 0.52$, $p = 0.003$), and can therefore interfere with

variance related to AD in statistical analyses, pack years of cigarette consumption were not included as a covariate in our analyses.

Task

The decision task was implemented in Presentation software (Neurobehavioral Systems) and consisted of nine blocks including 20 trials each. During each trial, participants had to decide whether to accept or reject a monetary offer between 1 and 99 cents (€). However, they were only allowed to accept a maximum of five out of 20 offers in each block. The main challenge of the task was thus to evaluate whether one should accept offers in early trials or one should wait instead for potentially higher offers in later trials. To allow for strategic decision making and rough estimates of upcoming monetary offers, consecutive offers never exceeded a value difference of more than 11 cents, with possible value changes being taken from the set [−11, −7, −3, 3, 7, 11]. Unknown to participants, the sequences of monetary offers followed a predefined pattern for each block (Figure S1), but the order in which participants were presented with block-specific patterns was randomized between participants.

Participants could base their decision whether to accept or reject a given offer on primarily three factors: (1) *offer value*, indicating the monetary amount of the current offer, (2) *offer index*, indicating how many offers were already presented in a block, and (3) the *number of accepts*, indicating how many offers were already accepted in a block. Information about the current state of these three factors was presented to participants in each trial for 5 s (Figure 1). After that, in a separate response phase (2 s), participants had to indicate their choice to accept or reject the offer via button presses with their left or right index finger. When the choice was not indicated within the time limit of 2 s, the offer was counted as rejected. The inter-trial interval consisted of a simple fixation cross and lasted for 2 s. Every block ended with a feedback screen (5 s), indicating the monetary earnings of the respective block, and an empty pause screen (10 s), in which participants could prepare for the next block.

Computational model

The decisions in the task allowed for strategic use of the variables offer value, offer index, and the number of accepts. To estimate to what extent participants took these variables into account, instead of basing their choices solely on the offer value, a decision model that has been validated for a similar sequential decision task (Economides et al., 2014) was computed for the choice data. The model estimates the expected value of accepting an offer V_A by comparing the monetary offer value R with a model threshold M:

$$V_A = R - M.$$

Accordingly, a high model threshold indicates that accepting an offer has a low expected value.

The model threshold is computed in the following way:

$$M = c_1 + a \times c_2 - o \times c_3,$$

with c_1 being a constant threshold, a the number of offers accepted previously, o the offer index, and c_2 and c_3 as weight parameters for a and o, respectively. In this formulation, the model threshold increases linearly when a increases (since accept choices should be more conservative when many offers have already been accepted), and the model threshold decreases linearly when o increases (since accept choices should be more liberal when the end of a block is near).

Finally, the expected value of accepting V_A is used to compute the probability of accepting P_A via a sigmoid function:

$$P_A = \frac{1}{1 + \exp\left(-\tau \times V_A\right)}$$

with τ governing the slope of the probability distribution. Thus, the computational model has four free parameters: a constant value threshold (c_1), weight parameters for the number of accepts (c_2) and the number of offers (c_3), and a parameter for the slope of the sigmoid function (τ).

Invalid trials (i.e., in which the maximum number of offers were already accepted or no response was given in time) were not included in the model. To test for differences in task performance between the patient and control group, the following behavioral variables were analyzed: profit, mean reaction time, mean index of accepted offers (indicating how long subjects were willing to wait), parameter estimates of the computational decision model (c_1, c_2, and c_3), and mean model threshold (M). Group differences were tested via a general linear model (GLM) including a fixed factor for group membership (1 = controls, 2 = patients) and age as a covariate of no interest.

In addition to the analysis of behavioral data, the formula for the model threshold M and the decision value R were also used in fMRI analyses (as parametric modulators P2 and P1, see below). To identify neural correlates of these processes, the formula was applied for each trial based on the participant's extant behavior to that point in the block, to create an idealized value that was entered into the hemodynamic model as an idealized BOLD signal waveform.

MRI data acquisition and preprocessing

Functional data

Functional imaging was conducted in a 3 Tesla Siemens Tim Trio MRI scanner (Siemens, Erlangen, Germany) with a 12-channel head coil. 32 contiguous slices were acquired in ascending order using a T2*-weighted gradient-echo sequence. For each participant, 940 volumes were recorded with the following imaging parameters: repetition time (TR): 1.9 s; echo time (TE): 30 ms; matrix size: 64 × 64; field of view (FOV): 192 mm; flip angle: 80°; voxel size: 3.1 × 3.1 × 2.8 mm^3; inter-slice gap: 0.7 mm.

TABLE 2 Behavioral data of alcohol-dependent patients and healthy control subjects

	Control subjects (23 male, nine female)		AD patients (22 male, 10 female)		Group difference[a]	
	Mean	SD	Mean	SD	F	p
Profit (cents; mean)	2725	134.6	2660.6	122.5	2.06	0.157
Index of accepted offers (mean)	10.5	1.2	10.2	1.5	0.47	0.498
Reaction time (ms; mean)	596.9	125.7	605.5	130.3	0.07	0.787
Model parameter c_1: constant value threshold	66.8	6	67.7	7	0.35	0.556
Model parameters c_2: number of accepts	3.1	1.4	3.1	1.2	0.01	0.908
Model parameter c_3: offer index	1.3	0.1	1.3	0. 1	0.4	0.551
Model threshold M: (cents; mean)	59.3	3.8	60.2	4.9	1.25	0.268

[a]Group differences of behavioral data were analyzed via a GLM including a fixed factor for group membership and age as a covariate of no interest.

Structural data

For registration purposes, a high-resolution, T1-weighted structural scan was acquired from every subject with a three-dimensional magnetization prepared rapid gradient-echo sequence (192 slices; TR: 2.3 s; TE 3.03 ms, matrix size: 256 × 256; FOV: 256 mm; flip angle: 9°; voxel size: 1 × 1 × 1 mm^3).

Preprocessing

The data were analyzed in Matlab (MathWorks) using SPM12 (Wellcome Department of Imaging Neuroscience, Institute of Neurology). Functional images were realigned to the first volume, slice-time corrected, coregistered to the structural data, spatially normalized to the template of the Montreal Neurological Institute (MNI), resampled to a voxel size of 3 × 3 × 3 mm^3, and smoothed using a Gaussian kernel of 8 mm full-width at half-maximum.

fMRI data analysis

GLM analysis

The fMRI data were analyzed via a GLM for each participant. Different event regressors were constructed as box-car functions with onsets and durations of the respective choice periods:

Regressor R1 corresponded to the decision phase (Figure 1) of all trials for which participants made valid choices, excluding trials in which no response was given in the response phase, or trials for which offers could not be accepted anymore because the limit was already reached. To identify neural correlates of specific decision variables, linear parametric modulators of regressor R1 were included in the GLM for the offer value (P1) and the model threshold (P2). Thus, P1 comprises the raw monetary value of the offer, whereas P2 represents model-based decision processes that take the offer index and the number of accepts into account. Both parametric modulators were *z*-transformed before they were added to the model. In addition, to minimize the error term of the GLM, regressors of no interest were included for the decision phase of invalid trials (R2), the response phase (R3), the feedback (R4), and pause (R5) phase between blocks, as well as six movement regressors R6 to R11 from the realignment procedure. R2 was included as a regressor of no interest because it is uncertain whether meaningful decision-making processes were present in invalid trials (since participants either did not indicate their choice in time or did not need to make a decision at all when they already accepted the maximum of five offers per block). Likewise, regressors R3 to R5 were of no interest since they were not part of the decision phase.

All regressors were convolved with the canonical hemodynamic response function and regressed against the BOLD signal in each voxel. Parametric modulators were not orthogonalized to each other, allowing regressors to fully compete for explained variance. First-level contrasts were constructed for offer values (P1) and model thresholds (P2) by weighting parametric modulators over baseline and submitted to second-level *t*-tests at the group level. One-sample *t*-tests were conducted separately for the patient and control group, and differential group effects were tested via two-sample *t*-tests that included age as a covariate of no interest. All statistical parametric maps from group analyses were thresholded at $p < 0.001$ (uncorrected) for voxel-level inference with a minimum cluster-size criterion of 10 contiguous voxels, and subsequent cluster-level family-wise error rate -correction for multiple testing at $p < 0.05$.

ROI analysis

ROIs were defined via spheres centered on coordinates from Economides et al. (2014). In particular, caudate nucleus ($x = -12$, $y = -6$, $z = 18$, radius = 5 mm) was used as an ROI for model-based decision processes, and vmPFC ($x = 4$, $y = 52$, $z = 14$, radius = 10 mm) as an ROI for value representation. Mean beta values of modulator P1

(offer value) and P2 (model threshold) were extracted from vmPFC and caudate nucleus, respectively. To analyze differences between groups, the beta values of each subject were entered into a GLM with a fixed factor for group membership and a covariate of no interest for age.

RESULTS

Behavioral results

Behavioral measures did not show significant group differences (Table 2), indicating that patients did not show impairments in task performance. In line with this result, summed accept responses for each trial in each block showed a similar distribution for both groups (Figure S1).

fMRI results

Whole-brain results

The first goal of the fMRI analysis was to identify regions that process the monetary value of offers. In the control group, whole-brain analyses revealed significant parametric modulation of offer values (P1) in a distributed set of regions including the dorsolateral prefrontal cortex (dlPFC), ventral striatum (vStr), and dorsomedial prefrontal cortex (dmPFC; Table 3; Figure 2A). The patient group showed activation in a largely overlapping set of regions (Table 3), and a whole-brain comparison of parametric group effects in a two-sample *t*-test did not reveal significant differences between the two groups.

Second, we investigated brain areas that demonstrated activation related to model-based decision processes via the model threshold parameter of the GLM (P2). In the control group, we did not observe effects related to positive model thresholds, but activity in the caudate nucleus and inferior parietal lobe was significantly associated with negative model thresholds (Table 3; Figure 2B), indicating stronger neural activity when the threshold was low and participants were more likely to accept offers. This is consistent with a previous study that found stronger effects for negative compared to positive model thresholds (Economides et al., 2014) and can be due to the BOLD signal being highest for go responses. The patient group, in contrast, did not show any activity related to negative model thresholds in the whole-brain analysis, but there was a significant cluster in the middle occipital gyrus associated with positive model thresholds (Table 3).

ROI results

ROI analyses were conducted to investigate group differences in vmPFC (associated with value representation) and caudate activation (associated with model-based decision processes). There were no group differences with respect to parametric effects of offer values in vmPFC, F (1, 60) = 0.1, p_{FDR} = 0.834), but we observed a significant difference in parametric effects of model thresholds in caudate nucleus, F (1, 60) = 4.4, p_{FDR} = 0.028 (Figure 2C), with stronger negative beta values for the control group.

DISCUSSION

This study was designed to compare neural processes of value computation and model-based decision making between alcohol-dependent patients and healthy control subjects. Participants performed an fMRI decision task, in which monetary offers had to be evaluated with respect to dynamically changing constraints. The results showed that patients had decreased functional representation of model-based decision processes in the caudate nucleus, whereas there were no group differences in terms of neural value representation or task performance.

Previous studies have found that the caudate is a crucial area for the computation of goal-directed choices that require the consideration of multiple factors and long-term planning (Balleine & O'Doherty, 2010; Dolan & Dayan, 2013; Geerts et al., 2020; Sharpe et al., 2019; Wunderlich et al., 2012). Likewise, in rodents, model-based decision processes have been associated with signals in the dorsomedial striatum (Balleine, 2005; Corbit et al., 2012; Gahnstrom & Spiers, 2020), which corresponds to the caudate activation that human neuroscience studies have identified. Our finding that the neural representation of model-based decision processes in the caudate nucleus is decreased for patients only therefore suggests that AD impairs the neural computations in the medial dorsal striatum for goal-directed choices, and supports the hypothesis that the neural mechanism underlying the ability to flexibly adapt choices to long-term consequences is one of the core functions affected by the disorder (Bechara et al., 2001; Goudriaan et al., 2007; Reiter et al., 2016; Sebold et al., 2014, 2017). Surprisingly, the occipital gyrus was associated with positive model thresholds in patients. Although this region clearly is affected by AD (e.g., Hermann et al., 2007), so far little is known about its role in decision-making processes, which makes it an interesting research question for future studies.

Our data further revealed that offer value is represented in a distributed set of regions including vStr, dlPFC, and dmPFC for both the patient and the control group. However, we did not observe systematic differences with respect to neural value computations between the two groups. Previous studies have suggested that AD could be based on an overactive valuation system (Arcurio et al., 2015; Goldstein & Volkow, 2011; Seo et al., 2013). Since we did not observe systematic group differences with respect to neural value representations, the results of the current experiment do not support these hypotheses, and speak for a uniform processing of nonalcoholic stimulus values in patients and healthy controls. In line with that, Bjork et al. (2008) showed that reward and loss anticipation during a monetary incentive delay task elicited similar activation of vStr in patients and controls as well as similar mood responses. This finding underlines the notion that value representation might not be a characteristic marker of addiction.

TABLE 3 Brain regions showing task-related activation

Region	Side	MNI coordinates x	Y	Z	T_{max}	p_{FWE} (cluster-level)
Controls						
Offer value						
Middle frontal gyrus	L	−24	−6	54	7.24	<0.001
Middle frontal gyrus	R	48	14	40	7.12	<0.001
Dorsolateral prefrontal cortex	R	44	40	24		
Dorsomedial prefrontal cortex	M	−2	30	42		
Ventral striatum	R	14	6	−6		
Inferior parietal lobe	R	40	−44	48	6.85	<0.001
Dorsolateral prefrontal cortex	L	−40	52	6	6.54	<0.001
Inferior parietal lobe	L	−44	−46	46	6.32	<0.001
Middle occipital gyrus	R	20	−94	10	5.26	<0.001
Cerebellum	L	−38	−66	−38	4.98	<0.001
Cerebellum	L	−12	−78	−34	4.96	<0.001
Middle occipital gyrus	L	−38	−92	0	4.95	<0.001
Substantia nigra	M	8	−18	−16	4.89	0.031
Negative model threshold						
Caudate	R	10	8	18	4.36	0.032
Inferior parietal lobe	L	42	−54	48	4.1	0.039
AD patients						
Offer value						
Dorsolateral prefrontal cortex	R	26	48	−8	6.51	<0.001
Dorsomedial prefrontal cortex	M	−2	32	38		
Ventral striatum	R	−10	10	−4		
Inferior parietal lobe	L	58	−44	50	5.65	<0.001
Cerebellum	L	−40	−72	−32	4.69	<0.001
Cerebellum	R	34	−56	−38	4.63	<0.001
Posterior cingulate cortex	M	4	−26	32	3.99	0.02
Model threshold						
Middle occipital gyrus		−30	−80	10	4.06	0.022

Note: Height threshold, $T_{24} = 3.23$; extent threshold, $k_E = 10$ voxels. All clusters survive whole-brain correction for multiple comparisons based on cluster-level FWE-control.

Interestingly, patients also did not show deviations in behavioral task performance. One explanation for this finding is that patients rely on different neural systems to achieve the same level of performance. Compensatory effects like these could also be found in other studies investigating decision making in AD patients (Charlet et al., 2014; Claus et al., 2018; Sebold et al., 2017), but the literature still shows mixed results (Galandra et al., 2018). The inconsistencies in the studies could result from the relatively large variance of decision-making tasks that were used since the tasks might recruit slightly different neural systems.

The absence of group differences in behavior and caudate responding could also be interpreted in the light of a finding by Gilman et al. (2015). They also observed that alcohol-dependent patients and controls showed few differences in behavior or in mesolimbic activation by choice for and receipt of (risky) gains. Interestingly, a history of rewarded instrumental responses boosted the activation of motivational neurocircuitry for additional reward in terms that patients exhibited heightened striatal activation that correlated with total earnings during the task.

The lack of behavioral group differences in our data could be related to the novel decision-making paradigm that we used, which might not have been optimal to detect differences on the behavioral level. One could speculate that even though the neural differences in the caudate were not associated with behavioral group differences in our task, they could have been revealed with a decision task that includes more uncertainty and requires more complex long-term

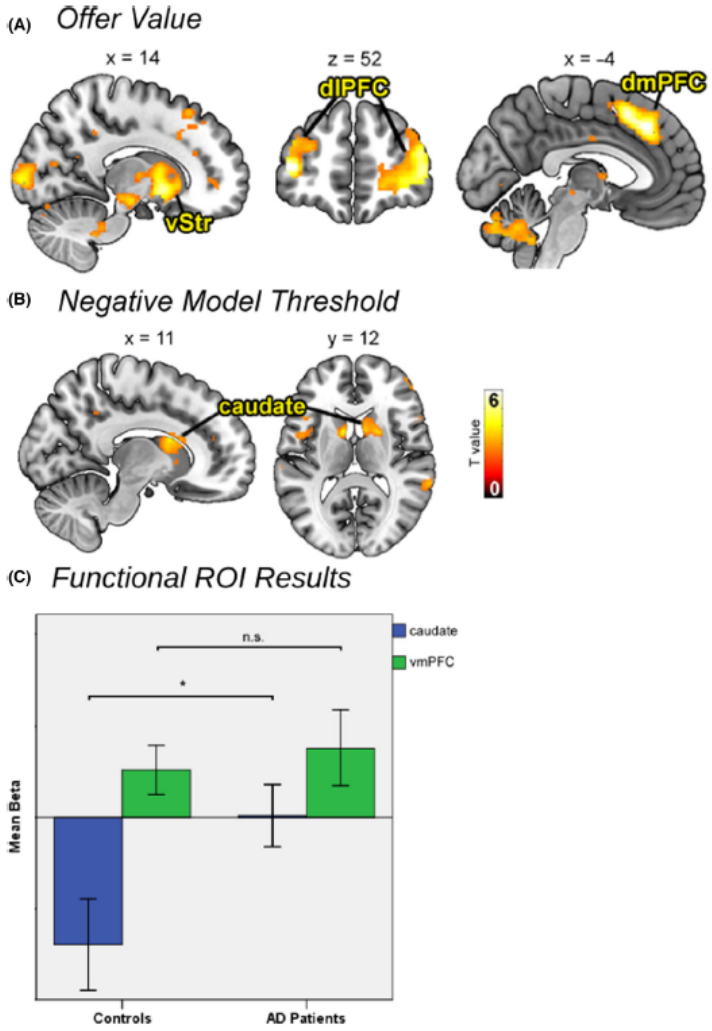

FIGURE 2 Brain regions showing parametric effects in the control group for (A) offer value and (B) negative model threshold. For illustration purposes, *t*-maps are thresholded at $p < 0.001$ (uncorrected), $k_E = 10$. Labelled clusters survive cluster-level FWE-correction at $p < 0.05$. The patient group showed largely overlapping clusters for offer values (Table 3), and no significant clusters for negative model thresholds. Abbreviations: dlPFC, dorsolateral prefrontal cortex; dmPFC, dorsomedial prefrontal cortex; vStr, ventral striatum. (C) Functional ROI results. Mean beta values in caudate nucleus were extracted from parametric modulators of model thresholds, and beta values in vmPFC from modulators of offer values, respectively. ROIs were defined as 5 mm spheres centered on coordinates from Economides et al. (2014). Asterisks denote significant FDR-corrected *p*-values < 0.05

planning. In the framework of our task, this could for example be tested in a future study by making trial-to-trial changes in monetary values more erratic, by increasing the number of trials per block, or by adding uncertainty to the number of offers that can be accepted (e.g., an unknown randomized number between 5 and 10 instead of the fixed number of 5).

Another limitation that has to be acknowledged in the interpretation of our study is that we did not control for nicotine dependence, because it is significantly correlated with AD and would have interfered with the analysis of alcohol-related effects. Even though this is a common issue in the study of AD (Batel et al., 2006), it is a source of uncertainty that limits the strength of the conclusions that can be drawn.

To conclude, this study highlights the role of the caudate nucleus in computing goal-directed choices and integrating multiple factors into adaptive choices. AD patients showed a decreased functional

representation of model-based decision processes in this region, which could be a key factor that characterizes decision-making dysfunctions related to AD.

ACKNOWLEDGMENTS

This work was supported by the Berlin School of Mind and Brain, Charité Berlin, the Ministry of Education and Research (BMBF; 01ZX1311E/e:Med-program Alcohol Addiction; Spanagel et al., 2013; and 01EE1406A), the Canadian Institute of Health Research, and the German Research Foundation (DFG Exc 257, DFG HE2597/14-2 as part of DFG FOR 1617). This work was further in part supported by the German Research Foundation (DFG: Project-ID 402170461 – TRR 265; CH 1936/1-1).

CONFLICT OF INTEREST

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

AM, CMS, and HW designed the study. AM was responsible for the data analysis, DJS conducted the computational modeling procedure, and all authors contributed to the discussion of the results. The manuscript was written by AM and edited by all authors.

ORCID

Amadeus Magrabi https://orcid.org/0000-0002-1229-7708

REFERENCES

Amlung, M., Vedelago, L., Acker, J., Balodis, I. & MacKillop, J. (2017) Steep delay discounting and addictive behavior: a meta-analysis of continuous associations. *Addiction*, 112, 51–62. https://doi.org/10.1111/add.13535

Arcurio, L.R., Finn, P.R. & James, T.W. (2015) Neural mechanisms of high-risk decisions-to-drink in alcohol-dependent women. *Addiction Biology*, 20, 390–406. https://doi.org/10.1111/adb.12121

Balleine, B.W. (2005) Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. *Physiology & Behavior*, 86, 717–730. https://doi.org/10.1016/j.physbeh.2005.08.061

Balleine, B.W. & O'Doherty, J.P. (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology*, 35, 48–69. https://doi.org/10.1038/npp.2009.131

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

Batel, P., Pessione, F., Maître, C. & Rueff, B. (2006) Relationship between alcohol and tobacco dependencies among alcoholics who smoke. *Addiction*, 90, 977–980. https://doi.org/10.1046/j.1360-0443.1995.90797711.x

Bechara, A., Dolan, S., Denburg, N., Hindes, A., Anderson, S.W. & Nathan, P.E. (2001) Decision-making deficits, linked to a dysfunctional ventromedial prefrontal cortex, revealed in alcohol and stimulant abusers. *Neuropsychologia*, 39, 376–389. https://doi.org/10.1016/S0028-3932(00)00136-6

Beck, A., Wüstenberg, T., Genauck, A., Wrase, J., Schlagenhauf, F., Smolka, M.N. et al. (2012) Effect of brain structure, brain function, and brain connectivity on relapse in alcohol-dependent patients. *Archives of General Psychiatry*, 69, 842–852. https://doi.org/10.1001/archgenpsychiatry.2011.2026

Bjork, J.M., Smith, A.R. & Hommer, D.W. (2008) Striatal sensitivity to reward deliveries and omissions in substance dependent patients. *NeuroImage*, 42, 1609–1621. https://doi.org/10.1016/j.neuroimage.2008.06.035

Charlet, K., Beck, A., Jorde, A., Wimmer, L., Vollstädt-Klein, S., Gallinat, J. et al. (2014) Increased neural activity during high working memory load predicts low relapse risk in alcohol dependence. *Addiction Biology*, 19, 402–414. https://doi.org/10.1111/adb.12103

Chase, H.W., Eickhoff, S.B., Laird, A.R. & Hogarth, L. (2011) The neural basis of drug stimulus processing and craving: an activation likelihood estimation meta-analysis. *Biological Psychiatry*, 70, 785–793. https://doi.org/10.1016/j.biopsych.2011.05.025

Claus, E.D., Feldstein Ewing, S.W., Magnan, R.E., Montanaro, E., Hutchison, K.E. & Bryan, A.D. (2018) Neural mechanisms of risky decision making in adolescents reporting frequent alcohol and/or marijuana use. *Brain Imaging and Behavior*, 12, 564–576. https://doi.org/10.1007/s11682-017-9723-x

Clithero, J. & Rangel, A. (2014) Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, 9, 1289–1302. https://doi.org/10.1093/scan/nst106

Corbit, L.H., Nie, H. & Janak, P.H. (2012) Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. *Biological Psychiatry*, 72, 389–395. https://doi.org/10.1016/j.biopsych.2012.02.024

DePoy, L., Daut, R., Brigman, J.L., MacPherson, K., Crowley, N., Gunduz-Cinar, O. et al. (2013) Chronic alcohol produces neuroadaptations to prime dorsal striatal learning. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 14783–14788. https://doi.org/10.1073/pnas.1308198110

Dolan, R.J. & Dayan, P. (2013) Goals and habits in the brain. *Neuron*, 80, 312–325. https://doi.org/10.1016/j.neuron.2013.09.007

Economides, M., Guitart-Masip, M., Kurth-Nelson, Z. & Dolan, R.J. (2014) Anterior cingulate cortex instigates adaptive switches in choice by integrating immediate and delayed components of value in ventromedial prefrontal cortex. *Journal of Neuroscience*, 34, 3340–3349. https://doi.org/10.1523/JNEUROSCI.4313-13.2014

Everitt, B.J. & Robbins, T.W. (2005) Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neuroscience*, 8, 1481–1489. https://doi.org/10.1038/nn1579

Everitt, B.J. & Wolf, M.E. (2002) Psychomotor stimulant addiction: a neural systems perspective. *Journal of Neuroscience*, 22, 3312–3320.

First, M., Spitzer, R., Gibbon, M. & Williams, J. (2001) *Structured clinical interview for DSM-IV-TR Axis I disorders, research version, patient edition with psychotic screen (SCID-I/PW/PSY SCREEN)*. New York: New York State Psychiatric Institute, Biometrics Research.

Furlong, T.M., Jayaweera, H.K., Balleine, B.W. & Corbit, L.H. (2014) Binge-like consumption of a palatable food accelerates habitual control of behavior and is dependent on activation of the dorsolateral striatum. *Journal of Neuroscience*, 34, 5012–5022. https://doi.org/10.1523/JNEUROSCI.3707-13.2014

Gahnstrom, C.J. & Spiers, H.J. (2020) Striatal and hippocampal contributions to flexible navigation in rats and humans. *Brain and Neuroscience Advances*, 4, 1–7. https://doi.org/10.1177/2398212820979772

Galandra, C., Basso, G., Cappa, S. & Canessa, N. (2018) The alcoholic brain: neural bases of impaired reward-based decision-making in alcohol use disorders. *Neurological Sciences*, 39, 423–435. https://doi.org/10.1007/s10072-017-3205-1

Geerts, J.P., Chersi, F., Stachenfeld, K.L. & Burgess, N. (2020) A general model of hippocampal and dorsal striatal learning and decision making. *Proceedings of the National Academy of Sciences of the United States of America*, 117, 31427–31437. https://doi.org/10.1073/pnas.2007981117

Gilman, J.M., Smith, A.R., Bjork, J.M., Ramchandani, V.A., Momenan, R. & Hommer, D.W. (2015) Cumulative gains enhance striatal response

to reward opportunities in alcohol-dependent patients. *Addiction Biology*, 20, 580–593. https://doi.org/10.1111/adb.12147

Goldstein, R.Z. & Volkow, N.D. (2011) Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nature Reviews Neuroscience*, 12, 652–669. https://doi.org/10.1038/nrn3119

Goudriaan, A.E., Grekin, E.R. & Sher, K.J. (2007) Decision making and binge drinking: a longitudinal study. *Alcoholism, Clinical and Experimental Research*, 31, 928–938. https://doi.org/10.1111/j.1530-0277.2007.00378.x

Hermann, D., Smolka, M.N., Klein, S., Heinz, A., Mann, K. & Braus, D.F. (2007) Reduced fMRI activation of an occipital area in recently detoxified alcohol-dependent patients in a visual and acoustic stimulation paradigm. *Addiction Biology*, 12, 117–121. https://doi.org/10.1111/j.1369-1600.2006.00039.x

Kamarajan, C., Ardekani, B.A., Pandey, A.K., Kinreich, S., Pandey, G., Chorlian, D.B. et al. (2020) Random forest classification of alcohol use disorder using fMRI functional connectivity, neuropsychological functioning, and impulsivity measures. *Brain Sciences*, 10, 115. https://doi.org/10.3390/brainsci10020115

Koob, G.F. & Volkow, N.D. (2010) Neurocircuitry of addiction. *Neuropsychopharmacology*, 35, 217–238. https://doi.org/10.1038/npp.2009.110

Kühn, S. & Gallinat, J. (2011) Common biology of craving across legal and illegal drugs—a quantitative meta-analysis of cue-reactivity brain response. *European Journal of Neuroscience*, 33, 1318–1326. https://doi.org/10.1111/j.1460-9568.2010.07590.x

Lee, S., Yu, L.Q., Lerman, C. & Kable, J.W. (2021) Subjective value, not a gridlike code, describes neural activity in ventromedial prefrontal cortex during value-based decision-making. *NeuroImage*, 237, 118159. https://doi.org/10.1016/j.neuroimage.2021.118159

Luijten, M., Schellekens, A.F., Kühn, S., Machielse, M.W. & Sescousse, G. (2017) Disruption of reward processing in addiction: an image-based meta-analysis of functional magnetic resonance imaging studies. *JAMA Psychiatry*, 74, 387–398. https://doi.org/10.1001/jamapsychiatry.2016.3084

Mollick, J.A. & Kober, H. (2020) Computational models of drug use and addiction: a review. *Journal of Abnormal Psychology*, 129, 544–555. https://doi.org/10.1037/abn0000503

Phung, Q.H., Snider, S.E., Tegge, A.N. & Bickel, W.K. (2019) Willing to work but not to wait: individuals with greater alcohol use disorder show increased delay discounting across commodities and less effort discounting for alcohol. *Alcoholism: Clinical and Experimental Research*, 43, 927–936. https://doi.org/10.1111/acer.13996

Reiter, A.M.F., Deserno, L., Kallert, T., Heinze, H.-J., Heinz, A. & Schlagenhauf, F. (2016) Behavioral and neural signatures of reduced updating of alternative options in alcohol-dependent patients during flexible decision-making. *Journal of Neuroscience*, 36, 10935–10948. https://doi.org/10.1523/JNEUROSCI.4322-15.2016

Rubio, G., Jiménez, M., Rodríguez-Jiménez, R., Martínez, I., Avila, C., Ferre, F. et al. (2008) The role of behavioral impulsivity in the development of alcohol dependence: a 4-year follow-up study. *Alcoholism, Clinical and Experimental Research*, 32, 1681–1687. https://doi.org/10.1111/j.1530-0277.2008.00746.x

Schacht, J.P., Anton, R.F. & Myrick, H. (2013) Functional neuroimaging studies of alcohol cue reactivity: a quantitative meta-analysis and systematic review. *Addiction Biology*, 18, 121–133. https://doi.org/10.1111/j.1369-1600.2012.00464.x

Schad, D.J., Garbusow, M., Friedel, E., Sommer, C., Sebold, M., Hägele, C. et al. (2019) Neural correlates of instrumental responding in the context of alcohol-related cues index disorder severity and relapse risk. *European Archives of Psychiatry and Clinical Neuroscience*, 269, 295–308. https://doi.org/10.1007/s00406-017-0860-4

Sebold, M., Deserno, L., Nebe, S., Nebe, S., Schad, D.J., Garbusow, M. et al. (2014) Model-based and model-free decisions in alcohol dependence. *Neuropsychobiology*, 70, 122–131. https://doi.org/10.1159/000362840

Sebold, M., Nebe, S., Garbusow, M., Guggenmos, M., Schad, D.J., Beck, A. et al. (2017) When habits are dangerous: alcohol expectancies and habitual decision making predict relapse in alcohol dependence. *Biological Psychiatry*, 82, 847–856. https://doi.org/10.1016/j.biopsych.2017.04.019

Seo, D., Lacadie, C.M., Tuit, K., Hong, K.-I., Constable, R.T. & Sinha, R. (2013) Disrupted ventromedial prefrontal function, alcohol craving, and subsequent relapse risk. *JAMA Psychiatry*, 70, 727–739. https://doi.org/10.1001/jamapsychiatry.2013.762

Sharpe, M.J., Stalnaker, T., Schuck, N.W., 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. https://doi.org/10.1146/annurev-psych-010418-102824

Skinner, H.A. & Sheu, W.J. (1982) Reliability of alcohol use indices. The lifetime drinking history and the MAST. *Journal of Studies on Alcohol*, 43, 1157–1170. https://doi.org/10.15288/jsa.1982.43.1157

Spanagel, R., Bartsch, D., Brors, B., Dahmen, N., Deussing, J., Eils, R. et al. (2010) An integrated genome research network for studying the genetics of alcohol addiction. *Addiction Biology*, 15, 369–379. https://doi.org/10.1111/j.1369-1600.2010.00276.x

Virkkunen, M. (1994) Personality profiles and state aggressiveness in finnish alcoholic, violent offenders, fire setters, and healthy volunteers. *Archives of General Psychiatry*, 51, 28. https://doi.org/10.1001/archpsyc.1994.03950010028004

Vollstädt-Klein, S., Wichert, S., Rabinstein, J., Bühler, M., Klein, O., Ende, G. et al. (2010) Initial, habitual and compulsive alcohol use is characterized by a shift of cue processing from ventral to dorsal striatum. *Addiction*, 105, 1741–1749. https://doi.org/10.1111/j.1360-0443.2010.03022.x

Voon, V., Reiter, A., Sebold, M. & Groman, S. (2017) Model-based control in dimensional psychiatry. *Biological Psychiatry*, 82, 391–400. https://doi.org/10.1016/j.biopsych.2017.04.006

World Health Organization. (2018) *World Health Statistics*. Geneva: WHO.

Wrase, J., Schlagenhauf, F., Kienast, T., Wüstenberg, T., Bermpohl, F., Kahnt, T. et al. (2007) Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *NeuroImage*, 35, 787–794. https://doi.org/10.1016/j.neuroimage.2006.11.043

Wunderlich, K., Dayan, P. & Dolan, R.J. (2012) Mapping value based planning and extensively trained choice in the human brain. *Nature Neuroscience*, 15, 786–791. https://doi.org/10.1038/nn.3068

Yin, H.H., Knowlton, B.J. & Balleine, B.W. (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *European Journal of Neuroscience*, 19, 181–189. https://doi.org/10.1111/j.1460-9568.2004.03095.x

Yin, H.H., Ostlund, S.B., Knowlton, B.J. & Balleine, B.W. (2005) The role of the dorsomedial striatum in instrumental conditioning. *European Journal of Neuroscience*, 22, 513–523. https://doi.org/10.1111/j.1460-9568.2005.04218.x

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Magrabi, A., Beck, A., Schad, D.J., Lett, T.A., Stoppel, C.M., Charlet, K., et al (2022) Alcohol dependence decreases functional activation of the caudate nucleus during model-based decision processes. *Alcoholism: Clinical and Experimental Research*, 46, 749–758. Available from: https://doi.org/10.1111/acer.14812

Curriculum Vitae

Mein Lebenslauf wird aus datenschutzrechtlichen Gründen in der elektronischen Version meiner Arbeit nicht veröffentlicht.

Full List of Publications

Journal Articles

Magrabi, A., Beck, A., Schad, D.J., Stoppel, C.M, Lett, T.A., Charlet, K., Kiefer, F., Heinz, A. & Walter, H. (2022). Alcohol Dependence decreases Functional Activation of the Caudate Nucleus during Model-Based Decision Processes. *Alcoholism: Clinical and Experimental Research*, *46(5)*, 749–58. http://doi.org/10.1111/acer.14812

Magrabi, A., Ludwig, V.U., Stoppel, C.M., Paschke, L.M., Wisniewski, D., Heekeren, H., & Walter, H. (2022). Dynamic Computation of Value Signals via a Common Neural Network in Multi-Attribute Decision-Making. *Social Cognitive and Affective Neuroscience*, *17(7)*, 683–93. https://doi.org/10.1093/scan/nsab125

Paschke, L.M., Dörfel, D., Steimke, R., Trempler, I., **Magrabi, A.**, Ludwig, V.U., Schubert, S., Stelzel, C. & Walter, H. (2016). Individual Differences in Self-Reported Self-Control Predict Successful Emotion Regulation. *Social Cognitive and Affective Neuroscience*, *11(8)*, 1193–1204. https://doi.org/10.1093/scan/nsw036

Ludwig, V.U., Stelzel, C., Krutiak, H., **Magrabi, A.**, Steimke, R., Paschke, L.M., Kathmann, N., & Walter, H. (2014). The Suggestible Brain: Posthypnotic Effects on Value-Based Decision-Making. *Social Cognitive and Affective Neuroscience*, *9(9)*, 1281–1288. https://doi.org/10.1093/scan/nst110

Magrabi, A. (2012). The Value of Feelings for Decision-Making. *Grazer Philosophische Studien*, *85(1)*, 279-90. https://doi.org/10.1163/9789401208338_013

Handbook Articles

Gaebler, M., Paschke, L.M., & **Magrabi, A.** (2016). Neurowissenschaft. In M. Kühler & M. Rüther (Eds.), *Handbuch Handlungstheorie* (pp. 414-421). Stuttgart: J.B. Metzler.

Magrabi, A. & Bach, J. (2013). Entscheidungsfindung. In A. Stephan & S. Walter (Eds.), *Handbuch Kognitionswissenschaft* (pp. 274-289). Stuttgart: J.B. Metzler.

Popular Science

Magrabi, A. (2015). Die Wiederentdeckung des Willens. *Gehirn & Geist, 5,* 199-205.

Acknowledgments

There are many people who supported me in my work on this dissertation. I would like to thank:

My supervisors, Henrik Walter and Hauke Heekeren, for all their help, guidance, and leadership. You helped me to design, plan, and interpret my experiments, and when I was lost in details, I could always count on pragmatic advice to bring me back on track.

Vera Ludwig, Tristram Lett, Christian Stoppel, and Anne Beck, for their mentorship and hands-on support. I especially want to thank Vera for being very committed to motivating me and helping me out of low points. This thesis would probably not have seen the light of day without you.

Lena Paschke, Carina Krause, and Patrick Plitzner, for their friendship, moral support, and for being amazing human beings. When I think about my favorite moments while I worked on these projects, I think about time spent with you.

My classmates, professors, and the staff at the Berlin School of Mind and Brain. I could not imagine a better academic environment to learn, discuss, and inspire interdisciplinary scientific curiosity.

My sister, Romana Wils, for her pioneering contributions to the family tree and thereby taking pressure off me, and for indirectly motivating me by unintentionally threatening to finish her thesis before me. I hope we can celebrate together soon!

And finally, my parents, Arooj Magrabi and Hildegard Wendker-Magrabi. I am beyond grateful that you always trusted in me, supported me, and let me pursue my passions without doubting me. I have been taking the love and stability that you always gave me for granted - looking back, I realize more and more how many sacrifices you made for me, and that you made it very easy for me to find success and happiness in my life. This dissertation is also your achievement, thank you for everything.