Predictive decision-making during exploration and social interaction in frontal cortex

Nadescha Trudel

DPhil thesis

Department of Experimental Psychology
Jesus College, University of Oxford

Trinity Term, 2020



Predictive decision-making during exploration and social interaction in frontal cortex

Nadescha Trudel

Jesus College, Oxford **Trinity term 2020**

A dissertation in partial fulfilment of the requirements for the degree of Doctor of Philosophy in the University of Oxford

Abstract

In this thesis, I investigated computational and neural mechanisms underlying adaptive decision-making in social and non-social contexts. I focused on the exploration/ exploitation dilemma, during which one has to choose between exploiting the familiar option with relatively known value or exploring the option with relatively uncertain value which might lead to a better outcome in the future **(**reviewed in **Chapter 1)**. Switching between exploration and exploitation was dependent on external factors such as the number of future choice opportunities (i.e. time horizon) and internal factors such as the belief estimates about the accuracy of a predictor ('accuracy') and the subjective uncertainty in the accuracy estimate ('uncertainty') **(Chapter 2)**. I used functional magnetic resonance imaging to investigate the neural correlates of accuracy, uncertainty and their modulation during exploration and exploitation **(Chapter 3)**. Medial and lateral frontal areas were active during exploratory decisions, while ventromedial prefrontal cortex had a unique role in representing multiple decision variables depending on the current behavioural mode. Tracking one's beliefs is not only important when selecting between predictors, but also when evaluating one's confidence in the predictor. When we estimate our confidence we take into account the identity of the predictor -- whether it is another social agent or a non-social stimulus that provides information, as well as the experience that we have with the agent or stimulus; people expressed more confidence for social compared to non-social predictors. The difference between social and non-social confidence judgments covaried with activity in right posterior temporoparietal junction **(Chapter 4)**. Not only are the estimates we form shaped by the social or non-social identity of the predictors; in addition, whether we learn through direct experience of another person's actions or vicarious observation of such actions influences how loyal or opportunistic we estimate these other people to be **(Chapter 5)**.

Approximate word count: 49.000 words

Acknowledgments

First, I would like to thank my supervisors, Matthew Rushworth and Marco Wittmann. You have guided me through all my ups and downs during these five years in Oxford. Throughout, you have offered me scientific knowledge and inspiration, support, strength, achievable goals and endurance. I am very thankful to you both; to have been guided scientifically and personally by two kind-hearted people who were always out for my best and as such offered me the best circumstances to thrive.

I would like to thank Arno Riedl who has strengthened my growing interest in decision-making during my time in Maastricht and who has supported me to pursue a masters in Neuroeconomics.

There are of course many other brilliant and kind people I have met along the way. I would like to thank the Rushworth lab for their scientific input and support. In particular, I have to thank – Caroline and Daria, for their cheeriness and companionship throughout and beyond this time. Miriam for her compassion and her scientific brilliance. Pat for her wise words about progress and self-esteem in academia. Elsa and Jacquie for being examples of strong and ambitious researchers. Lea, for her humbleness, her kayak sessions and kindness. Alizee, for being a personal and scientific inspiration to me. You all set examples of empowering women in academia. I would also like to thank Lennart and Davide, for being kind and supportive from the start. Kentaro, for including me into such a fascinating project– it gave me a break from own projects which helped me to stay sane through more complicated project phases.

Lev and Cristiana who brought me joy, continuous support and a friendship that will go beyond our Oxford time. Corinna, for being a loyal friend and fantastic housemate. Sorana, my Canadian rowing friend who I admire for her mental strength. Kevin, for being someone I could always rely upon, my sport-friend-science buddy who listened to me for many hours and (I believe) never got tired of it.

My new family, who has welcomed me as one of their own and who started to make me feel like home in the UK – Matt, Ed, Louise, Pat & Angie and Laura. You are a brilliant family and I admire you for your love, harmony and creativity.

My German support; Ann-Kathrin, who has been with me for the most important parts of my life. I could not have done it without your weekly (almost daily) voice messages, your support and your loyalty, even though we have been separated physically for such a long time. Sarah, for being a strong person and setting an

example for pursuing your journey no matter what life brings – and of course, for making me a godmother of such a wonderful small human being, Jasmine. Your continuous love has given me strength throughout this journey and all journeys.

I would have never got here without my family. My mother, my sister, Timo and our new family addition, Nola. Then, also Krisi and catch-me (aka Katha). You have created a constant support system for me. You have always welcomed me with such love and kind words when I arrived totally exhausted from Oxford (and Maastricht) and gave me new energy to continue. A special thanks goes to my mom, who has always pushed me, supported me, believed in me, made me laugh about annoying situations and who has danced with me to Supertramp and Waterboys in the bathroom. You have been truly the greatest support throughout all these years, really throughout all my life. Thank you so much.

Finally – someone who was very brave and who moved in me with me during the last year of my PhD. Someone who gave me a reality check when my mind went a bit crazy, constant support, who listened to me (every single day), cooked delicious food and made me laugh. Thank you, Matt. I believe you are at least as happy as me that this thesis is done!

I am very grateful to each one of you and I am very proud to have accomplished such an achievement with your support.

Table of Contents

TABLE OF CONTENTS **5**

1. CHAPTER: INTRODUCTION **9**

1.1. BEHAVIOURAL FLEXIBILITY AND THE EXPLORATION-EXPLOITATION TRADE-OFF 9

1.2. CONCEPTS OF THIS THESIS 14

1.3. EARLY ECONOMICS THEORIES ABOUT VALUE-BASED DECISIONS 15

1.4. COMPUTATIONAL MODELS 17

1.4.1. Decision rule *19*

1.4.2. Reinforcement learning model *21*

1.4.3. Bayesian model *23*

1.4.4. Motivators underlying belief formation *27*

1.5. BRAIN AREAS RELEVANT FOR BEHAVIOURAL FLEXIBILITY 30

1.5.1. Anterior frontal cortex *30*

1.5.2. Dorsal anterior cingulate cortex *36*

1.5.3. Ventromedial prefrontal cortex *41*

1.5.4. Summary *50*

1.6. WHAT IS UNIQUE ABOUT SOCIAL PROCESSING? 51

1.6.1. Social brain hypotheses and alternatives *52*

1.7. METHODS TO MEASURE NEURAL ACTIVITY 56

1.7.1. Functional magnetic resonance imaging *56*

2. CHAPTER: THE IMPACT OF SUBJECTIVE BELIEFS ON EXPLORATION AND EXPLOITATION ACROSS TIME 58

2.1. INTRODUCTION 58

2.2. METHODS 61

2.2.1. Participants *61*

2.2.2. Experimental procedure *61*

2.3. EXPERIMENTAL DESIGN 62

2.3.1. Trial timeline *62*

2.3.2. Design *64*

2.4. BAYESIAN MODEL 68

2.4.1. Model parameters *70*

2.5. CHOICE HYPOTHESES 74

2.6. BEHAVIOURAL ANALYSES 74

2.6.1. Decision Phase *75*

2.7. BEHAVIOURAL RESULTS 78

2.7.1. Dissociable effects of uncertainty and accuracy on predictor selections across time *78*

2.8. DISCUSSION 82

3. CHAPTER: POLARITY OF UNCERTAINTY REPRESENTATION DURING EXPLORATION AND EXPLOITATION IN VENTROMEDIAL PREFRONTAL CORTEX **84**

3.1. INTRODUCTION 84

3.2. METHODS 86

3.2.1. FMRI data acquisition and data processing *86*

3.2.2. Definition of exploration, exploitation and their transition *86*

3.2.3. FMRI whole-brain analyses *90*

3.2.4. FMRI region of interest analyses *94*

3.2.5. Correlation between behavioural and neural beta weights *98*

3.3. RESULTS 99

3.3.1. VmPFC represents multiple subjective beliefs *99*

3.3.2. Polarity of uncertainty decision signal in vmPFC changes with behavioural mode *102*

3.3.3. Uncertainty-related signals in subcortical structures during exploration and exploitation *114*

3.3.4. Uncertainty representation in vmPFC scales with predictor repetition *116*

3.3.5. The transition from uncertainty to certainty representation is accompanied by the processing of accuracy between predictors *118*

3.4. SUMMARY 123

3.5. DISCUSSION 124

3.6. SUPPLEMENTARY INFORMATION 129

3.6.1. fMRI cluster-corrected peak coordinates *129*

3.6.2. Alternative computational models *134*

4. CHAPTER: SUBJECTIVE CONFIDENCE JUDGMENTS ABOUT THE PERFORMANCE OF SOCIAL AND NON-SOCIAL PREDICTORS **143**

4.1. INTRODUCTION 143

4.2. METHODS 145

4.2.1. Social and non-social task versions *145*

4.2.2. Framing of social task version *147*

4.3. BEHAVIOURAL ANALYSES 150

4.3.1. Decision phase *150*

4.3.2. Confidence phase *151*

4.4. FMRI WHOLE-BRAIN ANALYSES 155

4.5. FMRI REGION OF INTEREST ANALYSES 157

4.5.1. Common neural correlates of confidence for social and non-social predictors *158*

4.5.2. Dissociable neural correlates of confidence for social and non-social predictors *159*

4.6. BEHAVIOURAL RESULTS 161

4.6.1. Effects of uncertainty and accuracy on predictor selections *161*

4.6.2. Confidence judgments about social and non-social predictors *164*

4.6.3. Changes of mind in confidence judgments are more frequent for non-social predictors *170*

4.7. NEURAL RESULTS 173

4.7.1. Common neural correlates of confidence for social and non-social predictors *174*

4.7.2. Unique social substrate: activation in posterior TPJ correlates with confidence judgment about social predictors *178*

4.7.3. Confidence judgments about others' performance are based on recent and distant encounters *182*

4.8. DISCUSSION 185

4.9. SUMMARY 188

4.10. SUPPLEMENTARY INFORMATION 190

5. CHAPTER: LEARNING ABOUT OTHERS THROUGH DIRECT AND VICARIOUS EXPERIENCES 191

5.1. INTRODUCTION 191

5.2. METHODS 194

5.2.1. Participants *194*

5.2.2. Experimental procedure *195*

5.2.3. Experimental design *196*

5.3. BEHAVIOURAL RESULTS 206

5.3.1. Predictions about others' behaviour vary according to contextual and social factors *206*

5.3.2. Impact of irrelevant interaction on current prediction *210*

5.3.3. Choices towards others are made prospectively *212*

5.4. SUMMARY 216

5.5. DISCUSSION 217

6. CHAPTER: GENERAL DISCUSSION 220

6.1. BEHAVIOURAL FLEXIBILITY AND THE EXPLORATION/EXPLOITATION TRADE-OFF 220

6.2. BEHAVIOURAL FLEXIBILITY DURING SOCIAL INTERACTIONS 222

6.3. NEURAL CORRELATES OF EXPLORATION AND THE UNIQUE ROLE OF VMPFC 224

6.4. COMMON AND DISTINCT REPRESENTATIONS OF CONFIDENCE ACROSS SOCIAL AND NON-SOCIAL CONTEXTS 226

BIBLIOGRAPHY 228

1. CHAPTER: Introduction

Our lives change continuously and at a high pace, whether this is due to the environment we are in, the people we are interacting with or whether it is our internal state such as our thoughts. All these changes, in turn, lead to changes in our behaviour; they lead to variations in the decisions we take across contexts and time. Being able to integrate multiple and changing sources of information in a manner that varies across contexts so as to be able to continually change the way that one makes decisions is a hallmark of behavioural flexibility. However, the brain mechanisms underlying behavioural flexibility are largely unclear. In this thesis, I present behavioural and neuroimaging studies that aim to contribute to our understanding of behavioural flexibility in both social and non-social domains.

1.1. Behavioural flexibility and the exploration-exploitation trade-off

Many decisions are choices between alternative courses of actions. Decisions can be made between a concrete set of options (Figure 1A), whether to approach or to avoid a particular option (Figure 1B) or whether to continue with the current option or to search for better alternatives (Figure 1C). Differentiating between these different types of decisions may be important, because different neural networks might have a relative specialisation for one type of decisions rather than the other. In many natural settings animals are not presented simultaneously with a set of

options, but instead have to decide whether to engage with sequentially presented options as they are encountered (Figure 1B, C). This decision problem has been modelled with foraging tasks that draw on behavioural ecology. When an animal forages for food, the animal has to decide whether to engage with the current patch ('default') or to leave the current patch and search for alternative patches ('alternative') (Hayden et al., 2011a; Kolling et al., 2012). In particular, the marginal value theorem (MVT) offers a normative account to predict the optimal time when to leave the current patch to search for alternative patches. It proposes that an animal should leave the patch when the current reward rate falls below the average reward rate of the environment (Charnov, 1976; Stephen and Krebs, 1986). Successful foraging is dependent on the integrated consideration of multiple factors: the value of the current patch (engage value) and its underling statistics, for example its depletion rate; the value of alternative patches in other words, the average richness of the environment (search value) (Kolling et al., 2012, 2016a, 2018; Le Heron et al., 2020; Wittmann et al., 2016a); and the costs that are incurred when searching for alternative patches. Costs can take different forms, such as the limited probability of finding food in a patch, the effort of getting to the different patch (Klein-Flugge et al., 2016) or the delay that is imposed when travelling to a new patch (Blanchard et al., 2015; Hayden et al., 2011a). Optimal foraging in a patchy environment requires balancing the benefits of staying in a depleting patch against the benefit of leaving for a possibly richer one (Charnov, 1976).

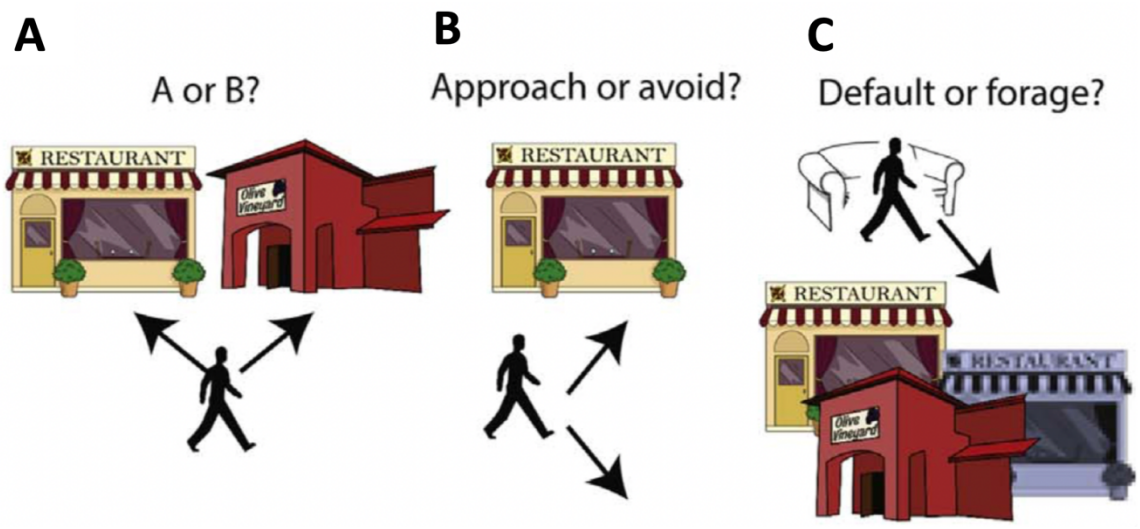


Figure 1. **Many decisions are choices between alternative courses of actions.** (A) We can make a decision between a concrete set of options, for example deciding between two restaurants. (B) Other decisions concern one option but differ in the type of action (approach/ avoid) one can make regarding this option. Yet another type of decision is whether to engage with the current choice ('default') or whether to forage and explore alternative options (C). Figure extracted and adapted from *(Scholl and Klein-Flügge, 2017)*.

Deciding to leave a patch or, more generally deciding to forego any opportunity with a known, positive value is often explained by the benefit of information sampling if the alternative options taken are less well understood. Although, information sampling incurs a cost, even if it is just an opportunity cost, receiving information is useful even if it cannot be used immediately (Loewenstein, 1994). For example, even when the action does not impact the final outcome, animals still choose to retrieve information about the reward magnitude before the outcome is revealed (Bromberg-Martin and Hikosaka, 2009; Stoll et al., 2016). This suggests that an option's subjective value is not only defined by its reward but also by the information that is predictive of future states (Bromberg-Martin and Hikosaka, 2011; Costa et al., 2019). A recent framework suggests that the value of information can be determined by multiple factors, all of which might motivate

information-seeking behaviour (Sharot and Sunstein, 2020). First, information can be of instrumental value if additional information increases the chances of making an action that leads to reward. Second, information can be of hedonic value if it predicts a positive outcome. In this case, the anticipation of reward can increase an option's expected reward (Iigaya et al., 2020). Third, information can be of cognitive value as it increases one's knowledge about the world; in this case, information-seeking behaviour increases particularly when information relates to topics of one's own interest (Sharot and Sunstein, 2020). In sum, this suggests that both reward and information determine the subjective value of a choice (Blanchard et al., 2015; Sharot and Sunstein, 2020).

The difficulty that one has deciding whether to take a choice to obtain an outcome of value such as a reward or to explore to find out information is often referred to as the explore-exploit dilemma or the explore-exploit trade-off. It refers to a situation with a given set of options whose true values are not fully known. However, according to the limited experience of the decision-maker one of the options appears as the best one. Should the decision-maker exploit this apparent best option and risk foregoing one of the other options that might, in fact, after closer inspection, turn out as the best one? Or should the decision-maker explore these alternative options at the expense of not receiving rewards from what, from all that is known so far, is the best option? In this context, exploration is the sampling of options of which little is known about, but which might turn out to be of superior value, while in contrast, exploitation refers to the pursuit of options that are currently believed to hold the highest value. Balancing the trade-off between

exploration and exploitation is one of the key problems the brain has to solve when behaving flexibly and adaptively.

Exploration is a form of information sampling as it can reduce the uncertainty of an unknown option, thereby potentially increasing long-term reward. Exploratory behaviour increases in response to changes in the internal state such as the subjective uncertainty about an option's value (Badre et al., 2012), the external state such as the novelty of an option (Costa et al., 2019) or as the time one has left (i.e. time horizon) in which to exploit the information gained by exploration increases (Wilson et al., 2014a). There are different types of exploration. For example, in volatile environments when stimulus-reward contingencies are changing dynamically across time (Behrens et al., 2007; Daw et al., 2006), exploration can reduce uncertainty about unknown options; in these cases, exploration is defined as directed exploration because it is aimed at gaining information about an uncertain option (Zajkowski et al., 2017). When choices deviate from choosing the option with the highest expected value, for example during stable reward-stimulus contingency periods, exploration might be better termed random exploration as opposed to direct exploration. In random exploration choices derive from decision noise or, more simply, mistakes in action selection (Wilson et al., 2014a, 2020; Zajkowski et al., 2017). The frequency of both types of exploratory decisions correlate across participants (Zajkowski et al., 2017), suggesting that they might derive from a common source. However, it has proven to be difficult to design a paradigm that dissociates between different types of exploratory behaviour and more importantly, that could reliably distinguish between information-seeking exploratory choices as opposed to accurate reward-based

choices or inaccurate reward-based choices that appear, but are not truly, exploratory. Partly this is because, over time, choices are directed towards rewarding options more frequently because of the information gained earlier and also because, if the time horizon that remains is short, there is less to be gained from seeking new information towards the end of a task. Nevertheless, the trade-off between exploration and exploitation captures important aspects of adaptive behaviour and decision-making. Further, exploratory behaviour can be induced by changes in the internal and external environment and therefore offers a particular interesting approach for studying behavioural adaptation as it can be dissociated when exploratory behaviour is useful and when it is not. It is, in fact, this decision trade-off that is the focus of the majority of experiments presented in this thesis. Most frequently it is discussed in the context of the broader and more general question of behavioural flexibility.

1.2. Concepts of this thesis

I have conducted several experiments that aim to disentangle behavioural and neural mechanisms of behavioural flexibility. Mostly I focused on the trade-off between exploration and exploitation and made use of computational models, in particular Bayesian models, to define different kinds of choices (exploitation, directed exploration, and random exploration) and to analyse the impact of dissociable subjective beliefs driving the trade-off between exploration and exploitation.

In this Introduction, I will cover theoretical, mathematical, and biological approaches relevant to behavioural flexibility. These include, early economic

perspectives on values-based decisions (theoretical), stochasticity in choice behaviour and its implementation into computational models (mathematical), and the brain areas involved in behavioural flexibility (biological). Finally, as multiple experiments aim to differentiate behavioural and neural mechanisms associated with social and non-social processes, I will shortly discuss the importance of behavioural flexibility in social settings and present approaches that might establish the degree to which social information processing is unique and specialized or, alternatively, that it is handled by general mechanisms identical to those used during non-social information processing.

1.3. Early economics theories about value-based decisions

Early economic theories suggested that all decisions are made to maximize reward; reward represents a beneficial outcome of any kind. This notion was first captured by the expected value theorem (EVT), which decomposes a choice into its outcome (x) and the probability of receiving this outcome p(x). The resulting expected value (EV) of a particular outcome is derived by weighting all outcomes by their respective probabilities:

$$EV(X) = \sum_{x} p(x) * x \tag{1}$$

Economic theories underlie axioms that assume specific patterns in a decision maker's behaviour that are, for example, driven by preferences between options within a set. If anyone's behaviour violated these axioms, then their

behaviour could not be described as utility maximizing. However, such violations have often been observed in empirical data. Perhaps one reason why this might be the case is that it seems unreasonable to assume that humans consider all possible outcomes and weight them respectively by their probabilities to make decisions. Another reason might be that factors other than 'utility' drive choice behaviour. Some of the first observations inconsistent with the EV theorem were made when comparing choices across contexts: for example, choosing between options of different values (expressed, for example, in units of money) were made differently across individuals. This was due to the fact that an additional unit of wealth is experienced differently depending on one's prior wealth endowment. If you are very rich already, then the prospect of earning one extra pound is of less consequence than if you are poor. Hence, the value one assigns to a particular option is subjective.

As a result of such observations, the expected utility theorem (EUT, developed by (von Neumann and Morgenstern, 1944)) assumed a degree of subjectivity between the mapping of a choice and its outcome: it is assumed that all benefits and costs could be compared on a common scale through a utility function that is both agent and context dependent. An example of context-dependency is the additional value of any level of wealth one would gain in respect to one's prior endowment, while an example of an agent-dependent factor might be the degree of their individual predisposition towards or against risk (risk aversion) in a gamble task. In common graphical representations of probabilistic decision making this is captured by the degree of concavity of their utility function.

Another proposal suggesting that rewards are valued subjectively, and that this subjectivity is further dependent on the agent and context, was made by Kahnemann and Tversky (1979) in their Prospect Theory. Their theory offers a mathematical framework which maps objective rewards onto an asymmetrical subjective value function that is steeper for losses compared to gains. In other words, the value function indicates that losses are weighted stronger compared to gains and this asymmetry is impacted by a reference point that can vary across situations and individuals. The framework had considerable impact in explaining decisions under risk and its results were recently replicated in a multinational study (Ruggeri et al., 2020).

Although this classification of economic models (EV, EU, or prospect theory) offered some of the first and still important ways to define the mapping between choices and their outcome value, it does not explain why the mapping between objective reward and subjective value occurs in this manner, nor does it explain the mechanistic process by which the computation of value itself occurs, nor the mechanism of value comparison between options during decision making. However, other mathematical formulations describing choice behaviour, -- computational models --, make further predictions about the decision process.

1.4. Computational models

Computational models are increasingly used to understand behavioural and neural mechanisms underlying cognitive, motor and perceptual processes. They provide trial-by-trial estimates of subjective decision variables and other latent variables such as uncertainty or the learning rate (Corrado and Doya, 2007),

thereby predictions about behaviour can be derived. In contrast to early economic theories, they allow testable hypotheses to be derived, data simulations can then be made on the basis of these hypotheses and tested against empirical choice data. Additionally, model-based variables can be correlated with neural activity; this makes it possible to establish whether there are neural processes that are similar to those specified in the model.

Computational models have two main parts: first, they make predictions about choice behaviour which are captured in a decision rule; second, they often consider how decision-making changes over time as the outcomes of earlier, previous decisions are witnessed. Learning rules in such models capture how observations of the outcome of one decision impacts on the next decision. More specifically, the learning rule determines the type and the weight of information that is integrated to update one's beliefs about the environment. Both rules can be adapted to fit choice behaviour in different circumstances, for example choices that might change due to fluctuations in the internal (i.e. beliefs) or external (i.e. reward rate) environment.

First, I will consider one commonly used decision rule that underlies both computational and logistic regression models used to make predictions about choice behaviour. I will show that while choices are predicted to be value-maximising as in early economic theories, nevertheless, in most instances, the choices that are made suggest a degree of stochasticity in the decision process. Next, I will review two key computational models important for this thesis, which are useful for capturing the changes in decision making that occur over time: reinforcement learning model and Bayesian model.

1.4.1. Decision rule

The decision rule makes predictions about which choice will be taken on a given instance or trial in a psychological experiment. To predict choice behaviour, one has first to define the mapping between an option and its subjective value. This mapping can be represented by the explicit formulation of a decision variable (DV) that either assigns subjective value to a single choice or calculates the difference between subjective values of multiple options, for example as follows:

$$\text{Decision variable (DV)} = V_{Option_j} - V_{Option_i} \quad (2)$$

Here, the DV is defined as the difference in value (V) between two options. To predict the probability of a choice as a function of the DV, one can fit a sigmoidal function to the choice data (Johnson et al., 2002; Mcfadden, 2001):

$$Probability\ of\ Choice\ (Option\ j) = \frac{1}{1+ e^{\beta * -DV}} \quad (3)$$

In the decision rule, the *inverse temperature* or beta determines how strictly choices are guided by the DV. Depending on the beta value, choices can either map directly onto the DV (high beta value); in this case, the decision maker is said to follow a *hardmax* function (Sutton and Barto, 1998) and the choices made are those that would be predicted by earlier notions that decision-maker maximize subjective utility. However, in other cases, a low beta value means that stochasticity occurs in choice behaviour that inevitably leads to occasions in which choices are not value-maximizing. These choices are sometimes referred to as

random exploration as they do not follow value maximization but they can still be modelled by a *softmax* function (Daw et al., 2006).

Modifications of the decision rule beyond the tuning of the inverse temperature can be made to capture choice stochasticity, for example to predict different behavioural modes of exploration and exploitation. For example, one can implement an epsilon-greedy strategy during which the highest expected value option is chosen, but occasionally a random small fraction of choices are of the non-best option (Watkins and Dayan, 1992). While this is an alternative computational method that implements random exploration, it is also possible to add an information bonus into the softmax function, which drives selections towards informative options (Costa et al., 2019; Dayan and Sejnowski, 1996). The latter constitutes a mechanism for directed exploration as it distinctly benefits from options that are currently relatively unknown but might have the potential to be rewarding.

The DV can be obtained by explicit reports such as asking participants to rate the likeability of food items (Lebreton et al., 2015) or alternatively, when variables of interest are not directly measurable (i.e. latent variables), computational models can be used to infer belief states. In this thesis, the use of a computational model, in particular Bayesian model, was crucial to differentiate between behavioural modes of exploration and exploitation by defining multiple dissociable latent variables hypothesised to drive behaviour. However, other classes of models such as reinforcement learning (RL) models have been used to make significant contributions in explaining behavioural flexibility. Both approaches can be used to estimate how humans and animals track changes in the environment (e.g. volatility

in reward rates) and how they consequently adapt their choice behaviour (e.g. exploration vs exploitation).

1.4.2. Reinforcement learning model

A problem that has to be faced by a decision maker is to learn which choices are reinforced by reward or punishment and then to, respectively, maximize or minimise them over the long run. In a simple RL framework, an agent forms a prediction about a choice's outcome and updates this estimate every time a new observation is made. This defines the delta rule:

$$\text{Prediction}_{(\text{next trial, OptionA})} = \text{prediction}_{(\text{current trial, OptionA})} + \alpha * \text{PE}_{(\text{current trial, OptionA})} \quad (4)$$

Across choices, the agent tries to minimize the discrepancy between the prediction of the outcome and the observed outcome; the difference of these two quantities defines the prediction error (PE) (Rescrola and Wagner, 1972; Sutton and Barto, 1998):

$$\text{PE}_{(\text{current trial, OptionA})} = \text{observed outcome}_{(\text{current trial, OptionA})} - \text{prediction}_{(\text{current trial, OptionA})} \quad (5)$$

As the learner is more and more able to perfectly estimate the true reward value of the option being taken, prediction errors decrease by updating the value estimate of the prediction after having observed the outcome. Prediction errors can be either positive or negative and its absolute numerical value increases with the discrepancy between prediction and true outcome. How much the estimate is

revised depends on both the prediction error and, a free parameter, the learning rate or alpha, α (see Equation (4)). Alpha can range between zero and one and determines the size of the update step across trials. In most cases, predictions are updated for the option that has been chosen and in relation to the outcome that is observed. In some formulations of RL, unchosen option values are also updated trial by trial in order, for example, to capture the effects of forgetting (Khamassi et al., 2015). In this case, predictions might return towards an initial default prediction that the learner held at the beginning of the learning process.

An analogous approach to error-driven updating is to set the next prediction of a choice's outcome as the weighted average between the outcome that was most recently actually observed and the previous, prior outcome prediction. If the learning rate is large, then the prediction will be influenced by each new reward and so each next prediction will be very similar to the most recent outcome received. By contrast, a low learning rate means that less weight is put on each new outcome and so the next prediction is still strongly influenced by many prior older estimates. Thus, the change in prediction update on each trial is minimal. As a result, the learning rate can be thought of as defining the steepness of decay of the influence of the previous observations.

Modifications to learning rules that are offered by parameters such as the learning rate are particularly useful when modelling choice behaviour in environments where the reward rates of options change across time. For example, in volatile environments value estimates have to be updated more frequently to adjust for changes in reward rates (Behrens et al., 2007; Meder et al., 2017); in these cases, the learning rate is closer to one as recent observations are more

informative than observations made in the past. An alternative approach to model the speed of value updating (i.e. alpha) in these volatile environments is through Bayesian inference (Behrens et al., 2007; Daw et al., 2006; Dayan et al., 2000; Kalman, 1960). In these approaches, the value estimate is weighted by its uncertainty. I will now look in more detail into how Bayesian models represent beliefs, update these beliefs, and how belief estimates can be estimated to provide insights into both behavioural processes and neural mechanisms.

1.4.3. Bayesian model

As with other computational models, Bayesian models can serve as cognitive models that aid thinking about the nature of internal representations of the environment and experimental task as well as making precise and testable predictions about behaviour and neural activity. Additionally, Bayesian inference can be used to evaluate hypotheses and interpret the absence or presence of an effect in terms of its probability (Keysers et al., 2020) which serves as an alternative to the commonly used frequentists approach that focuses on summary statistics (e.g. frequency, mean or expected value). There are unique elements to Bayesian models: first, beliefs are represented as probability distributions; second, estimates about value are weighted by their uncertainty; third, new observations are interpreted according to prior beliefs (O'Reilly et al., 2012).

1.4.3.1. Beliefs are represented as probability distributions

Computational models are used to estimate a person's belief about some aspect of a particular option, for example the probability that is estimated of

receiving a reward when choosing that option. Although the state of the environment takes on one single true value (e.g. the specific probability of reward), Bayesian models formulate beliefs in terms of a continuous distribution over all possible values weighted by their probability of occurrence. This defines the probability density function (pdf) or the belief distribution. Hence, beliefs can be expressed in terms of probabilities assigned to possible events (e.g. different probability values), while in frequentists statistics, events are represented by their frequency of occurrence. After an infinite number of observations, both approaches are likely to converge towards the same value estimate that is close to the true value. However, when there are only a few observations, the value estimates made by the two approaches diverge. In such cases, weighting the belief estimate by its uncertainty gives a more accurate representation of a person's belief about the state of the environment (De Martino et al., 2017; O'Reilly et al., 2013).

1.4.3.2. Learning as Bayesian update

Bayesian inference allows the integration of current and previous observations into the next updated beliefs about the environmental state. Prior and current observations are integrated according to Bayes' theorem to derive a posterior belief distribution which represents the belief after updating:

$$p(\text{true state} \mid \text{observation}) \propto p(\text{observation} \mid \text{true state}) \times p(\text{true state}) \quad (6)$$

The posterior belief distribution (left hand-side) is proportional to the prior belief distribution (right-hand side) multiplied by the likelihood of that observation (middle part). Previous observations are represented by a prior pdf, which might

assume a number of forms (i.e. uniform prior, Cauchy prior) depending on the assumptions made in the experiment. Importantly, the posterior distribution serves as the next prior distribution when the same option is encountered once again. As a result, beliefs are updated and learning evolves across observations. Successful learning from new observations can be expressed in two ways: first, the mean of the distribution, representing the estimated value, will converge towards the true value of an option. Second, the width of the distribution (i.e. uncertainty) will decrease as the amount of sampling increases. Here, two types of uncertainty can be defined: the reducible uncertainty, that decreases with information sampling and the irreducible uncertainty, that underlies the probabilistic nature of an option's reward rate. For example, an option is inherent irreducible uncertainty about what might happen on any occasion a choice is made when that choice is rewarded on 80% of occasions (Payzan-LeNestour and Bossaerts, 2011). Parallels can be drawn between reducible and irreducible uncertainties and expected and unexpected uncertainties, respectively (Yu and Dayan, 2005). These model characteristics provide a powerful tool as estimates of value and uncertainty in that value can be used to predict choice behaviour and neural mechanisms (Chapter 2,3).

1.4.3.3. A Bayesian perspective on exploration-exploitation trade-off

The key striking feature of the Bayesian perspective is that it assumes that the learner or decision-maker does not only have a single point-estimate of the most likely outcome or the average outcome of an action. Instead it assumes that the decision-maker in some way represents the entire distribution of possible

beliefs about the outcomes and each belief is associated with an estimate of the chance that it is correct. This offers a framework for accounting for both exploratory and exploitative choices. While the peak of the distribution can be linked to immediate value-maximising strategies (an exploitative choice is one made towards the option with the higher peak distribution), the relative width of the distribution is inherently linked to the uncertainty of a choice. The relative width of the distributions associated with two choices reflects the relative uncertainty of beliefs about what will ensue if those choices are taken.

It has been suggested that exploratory choices should be directed towards the more uncertain option – the option with the wider distribution (Badre et al., 2012). In fact, the study presented in this thesis uses a similar approach to identify exploratory trials. However, instead of defining exploration as a deviation from a value-maximizing strategy (i.e. exploitation) (Daw et al., 2006), or only according to the relative uncertainty between options (Badre et al., 2012), I have identified exploratory trials in a way that is consistent with both previous approaches: the belief in the value of an option and the uncertainty about the estimated value (see schematic Figure 2 for a comparison of previous approaches and the one used in Chapter 2 and 3).

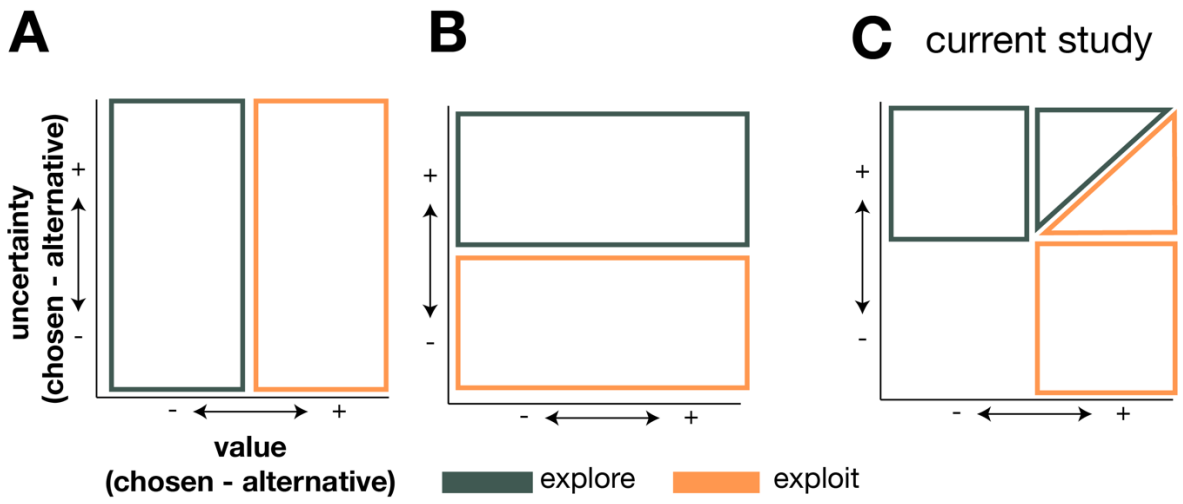


Figure 2. Schematic illustration of previous and current exploration/exploitation classification. Exploration/ exploitation has often been described as a dichotomy or a trade-off, between two distinct behavioural modes. **(A)** Previous studies have defined exploitation and exploration, respectively, as choosing options with higher or lower expected values compared to the alternative option (x-axis) (Daw et al., 2006). **(B)** Other studies have defined exploration as being directed towards options with higher relative uncertainty compared to the alternative option (y-axis) (Badre et al., 2012). **(C)** I merged all three principles (dichotomy, value, and uncertainty) for distinguishing between explore-exploit trials and define exploration and exploitation in reference to both uncertainty and value (Chapters 2,3).

1.4.4. Motivators underlying belief formation

In Bayesian terms, each time a new observation is made about a contingency in the world, such as the relationship between a predictor and a subsequent outcome, it is integrated with prior beliefs about that contingency to determine what will be the current belief about the contingency. Over the course of time and repeated observations, beliefs are updated to approximate the true value of the contingency as accurately as possible. Something similar happens in RL models that propose that the decision-maker predicts what the outcome will be every time a predictor is seen and attempts to minimize the prediction error across time, thereby approximating the true value of an option. Hence, one aim underlying

belief formation is to reflect predictor-outcome associations as accurately as possible to eventually direct choices towards rewarding outcomes (Figure 3A).

If the decision-maker aims to maximize external outcomes such as reward through learning about environmental contingencies, then this implies, first, that choices should be guided towards options that are informative of future outcomes and second, that any kind of information that facilitates effective learning about environmental contingencies should be integrated into one's belief update.

However, this does not appear always to be true. In some cases, experimental participants prefer not to receive information about future outcomes when this information predicts a loss as opposed to a gain (Charpentier et al., 2018). This pattern of behaviour is also apparent in real-life decision making, for example when patients reject medical screening or avoid receiving medical information about the possibility of an event of a negative nature such as a terminal illness. Hence, there is an asymmetry in information-seeking behaviour that is a function of the valence of the information that might be sought. Further, the preference for receiving positive information also biases belief formation. When receiving new information, participants weighted information that was 'better than expected' more compared to information that was 'worse than expected' into their belief update – this is also known as the optimism bias (Sharot et al., 2011). As a result beliefs are asymmetrically updated for positive compared to negative information (Sharot and Garrett, 2016), even though both types of information can be equally informative about environmental contingencies. It has been suggested that this type of bias reflects an adaptive mechanism (Sharot and Sunstein, 2020), as it seems to occur mainly in situations during which it is beneficial to be

optimistic. On the one hand, when facing an uncertain environment, it may be better to overestimate the novelty of the environment as this increases the rate of exploration and ultimately this leads to subsequent success in the task (Berger-Tal and Avgar, 2012). On the other hand, such bias evaporates during situations of perceived threat. Then, beliefs about positive and negative information are updated symmetrically and this ensures accurate decisions are made (Garrett et al., 2018).

Another bias in belief formation occurs in social settings; one's own beliefs are updated more strongly when one is exposed to opinions that are consistent as opposed to inconsistent with one's own opinion (Kappes et al., 2020). This constitutes a form of confirmation bias Hence, beliefs are not solely updated with the aim of representing external contingencies as accurately as possible, but instead they may be updated in a manner that means that one becomes more confident in one's own opinion (Kappes et al., 2020) or one becomes more optimistic about the future (Sharot et al., 2011) (Figure 3B).

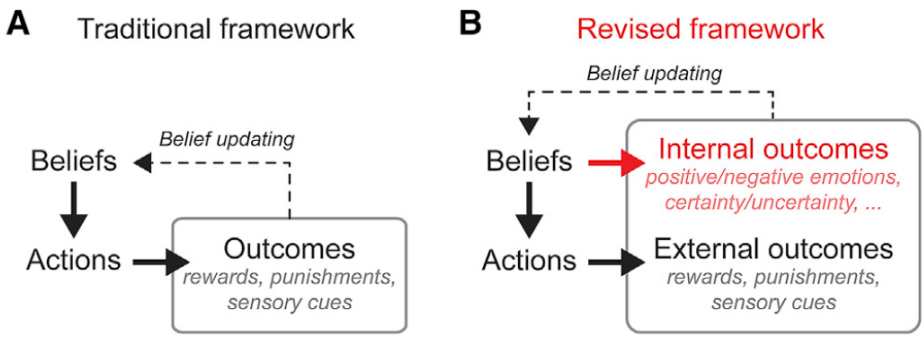


Figure 3. Motivators underlying belief formation. There are multiple motivators that impact belief formation. A traditional framework suggests that beliefs are updated to guide actions towards beneficial external outcomes such as reward **(A)**. However, studies show that actions are not only guided to maximize reward, but also to fulfil internal drives such as becoming more confident in one's beliefs or to increase positive emotions, for example by being optimistic about future outcomes **(B).** Reprinted from (Bromberg-Martin and Sharot, 2020) with permission from Elsevier.

1.5. Brain areas relevant for behavioural flexibility

This thesis focuses on a set of regions in medial frontal cortex and lateral frontal cortex relevant for value computations, behavioural flexibility and balancing exploration/exploitation. I will review the anatomy and function of the following areas: the anterior frontal cortex (aFC), dorsal anterior cingulate cortex (dACC) and ventromedial prefrontal cortex (vmPFC) and relate their functions to behavioural flexibility.

1.5.1. Anterior frontal cortex

1.5.1.1. Anatomy

According to the cytoarchitectonic map of Brodmann (1909) and as revised by Walker and others (Carmichael and Price, 1994; Petrides et al., 2012; Walker, 1940), anterior frontal cortex (aPFC) corresponds to area BA10 and comprises the most rostral part of superior and middle frontal gyri (Figure 3A). In general, subdivisions can be made with help of cytoarchitectonic features, which detect changes in laminar patterns (Amunts et al., 1999). However, using MRI to draw detailed borders in the rostral part of the brain, for example within area 10 or area 11 is difficult due to artefacts induced by air/tissue interfaces as these areas are located close to the frontal sinus (Bludau et al., 2014). Therefore, differentiating borders according to connectivity patterns taken from diffusion weighted MRI is another *in vivo* option to subdivide the anterior parts of the frontal cortex (Neubert et al., 2014). Simultaneously, this method allows us to draw cross-species comparisons in respect to their connectivity fingerprints. Using such an approach,

three areas within anterior frontal cortex could be differentiated: dorsolateral prefrontal cortex (dlPFC) which corresponds to area 46 (Figure 4B) and two areas within the frontal pole, the middle (FPm) and lateral frontal pole (FPl) (Figure 4B) (Bludau et al., 2014; Neubert et al., 2014).

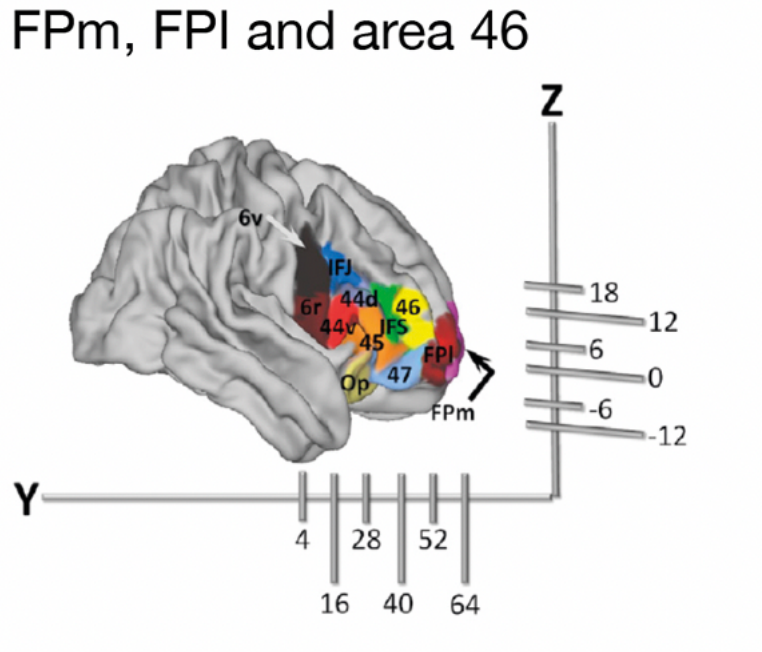


Figure 4. Subdivisions of the frontal cortex. Brodmann areas of the human prefrontal cortex. Anterior lateral frontal cortex corresponds to area 10. Further subdivisions of interest to this thesis correspond to area 46 (dorsolateral prefrontal cortex), middle (FPm) and lateral frontal pole (FPl). Reprinted from (Neubert et al., 2014) with permission from Elsevier.

Area 46 is functionally connected with frontal-parietal networks and corresponds to macaque area 46. Further distinctions can be made between FPm and area 46 in their connectivity patterns; compared to area46, FPm shows a weak connectivity to parietal areas and instead is connected to the temporal pole, posterior cingulate cortex and amygdala. In comparison to FPm, the adjacent area FPl showed a different connectivity profile with a lack of connectivity to amygdala (unlike FPm) and instead being connected to the inferior parietal lobule (like area46). Interestingly, FPm has the most direct correspondence to a homologous

area in the macaque brain; it corresponds to the macaque's frontal pole region 10. By contrast, there is no direct homologue of FPl in the macaque brain. Both factors, the difference in connectivity profiles of FPl compared to FPm and area46 and the lack of a simple comparison to macaque brain, might suggest a fundamentally distinct cognitive role for FPl (Neubert et al., 2015).

1.5.1.2. Function

Many studies report evidence for BA 10's role in executive control such as planning and task management (Mansouri et al., 2017; Passingham and Wise, 2012; Rushworth et al., 2011). The general consensus across studies shows that anterior frontal cortex maintains information that are not immediately driving behaviour but that could lead to alterations in future choice behaviour – a process underlying prospective planning. Multiple aspects need to be considered when planning prospectively: information that might lead to future goals or states such as the sequence of actions and the uncertainty about these action sequences and the way in which the choices of later steps in the sequence are dependent on the first steps that are taken (Kolling et al., 2018). Prospective planning can be tested by presenting participants with a maze that requires the planning of multiple action steps (i.e. sequence) to get to the desired goal. When humans navigated through a maze according to incomplete information about the environment, the amount of uncertainty (i.e. entropy) about the progress through the maze, that is the uncertainty associated with the action sequence, was positively correlated with activity in BA10/9 (Yoshida and Ishii, 2006) (see also: (Levy et al., 2010)).

In these studies, humans had an explicit task to navigate through a maze, inducing the necessity of representing alternative courses of actions. However,

even without explicit instruction, activity in a part of aPFC correlated with the alternative or counterfactual choice – the choice that was not currently chosen. In these cases, aPFC signal correlated positively with the reward probability of the forgone option but negative with the reward probability of the currently chosen option (Boorman et al., 2009). Additionally, the signal in aPFC accumulated across trials until a switch to the alternative option occurred. Interestingly, aPFC does not only represent the counterfactual choice *per se* but instead, when being presented with multiple options, neural activity in aPFC correlates with a ranking amongst forgone choices; neural activity correlated positive with the best forgone alternative and negatively with the worst alternative and the currently chosen option (Boorman et al., 2011). This suggests that aPFC represents the prospective value of switching to an alternative choice, and that not all alternative options are considered similarly and weighted equally – the representation of the best alternative to the current choice is special and prioritized.

Given that aPFC represents how good it might be to switch to an alternative choice it might be useful to think of aPFC as accumulating evidence for switching away from the current or 'default' option. It should, therefore, not come as a surprise that aPFC is also involved during exploratory behaviour. aPFC activity correlated with exploratory choices which were defined as choosing the lower expected value option (Daw et al., 2006), but also when exploratory choices were defined in relation to the relative uncertainty between options; then, activity in aPFC increased when choosing the more uncertain option. In particular, FPl represented the relative uncertainty, while area 46 represented the mean uncertainty of both options (Badre et al., 2012). Although this might suggest that

aPFC is important for both directed exploration (Badre et al., 2012) and random exploration (Daw et al., 2006), stimulating aPFC with transcranial magnetic stimulation (TMS) only resulted in alterations of directed exploration in long time horizons: when participants have many future opportunities to use the knowledge gained from exploration to guide subsequent decisions aPFC TMS decreased the exploration rate compared to a control stimulation site. Additionally, this difference was not observed during shorter time horizons or during random exploratory choices (Zajkowski et al., 2017). As only exploratory choices during long horizons were affected, this again supports aPFC's role in prospective planning. In other words, exploratory choices mainly occur in the beginning of long horizons as the additional knowledge gained can improve future exploitative choices – hence, the occurrence of exploratory choices in long horizons are a feature of prospective planning for future exploitative choices.

Whether or not one decides to switch away from the default option should be influenced by new evidence in respect to the decision one has just made. New evidence can influence the post-choice evaluation of one's decision; more evidence that arrives even after the decision is made and the choice is taken can lead to more or less confidence about having made the correct choice. Lateral aPFC, more specifically area 46 and FPl, mediated the impact of post-choice evidence on explicit confidence reports about a recent choice in a random dot motion task (Fleming et al., 2018). Both areas mediated changes of mind, thereby showing a possible role in metacognitive control of future behaviour. Note again that in this study, evaluating post-choice evidence in this way can be interpreted as retrospectively evaluating whether it would have been better to have made a

different choice. In sum, these studies show that aPFC has a role in prospective planning and/or metacognition as a consequence of the way in which it tracks information about alternative options that might have been taken and might still become the focus of behaviour in the future.

A summary of peak-coordinates reported in studies that have been particularly influential on this thesis is presented in Figure 5. Although not many studies differentiate between areas within aPFC, neural activity underlying exploratory behaviour can be found in both area 46 and FPl.

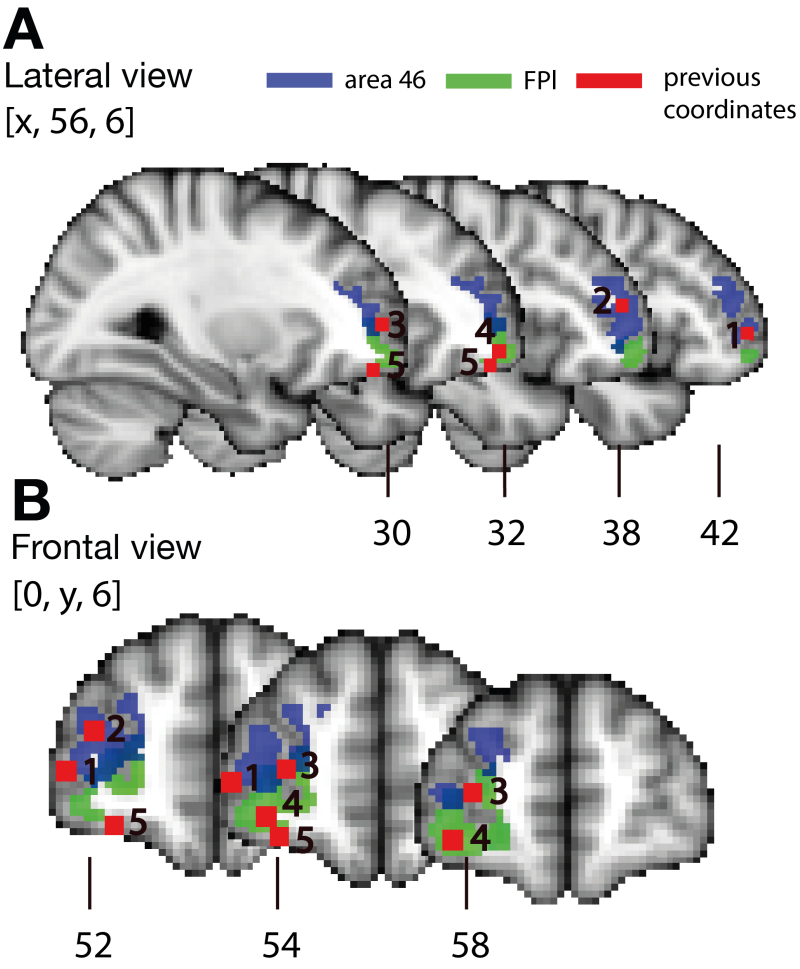


Figure 5. Previous coordinates capturing the role of aPFC in behavioural change. *The blue and green masks represent areas defined as area46 and FPl, respectively (Neubert et al., 2014). Red squares represent peak coordinates for some of the studies that have influenced*

this thesis. Often areas are classified as aPFC, however overlaying them onto masks of area 46 and FPl shows their distribution across these sub-regions of aPFC and allows for a comparison across results. [1:(Levy et al., 2010); 2: (Zajkowski et al., 2017); 3:(Daw et al., 2006); 4:(Boorman et al., 2009); 5:(Badre et al., 2012)].

1.5.2. Dorsal anterior cingulate cortex

1.5.2.1. Anatomy

Anterior cingulate cortex is part of the agranular cortex and like prefrontal cortex it can be found in humans and macaques but in addition, and unlike prefrontal cortex, a number of its subdivisions can be identified in other mammals such as rodents (Vogt and Paxinos, 2014) (Figure 6). Anterior cingulate cortex can be subdivided into areas 24, 32 and 25, which are present in all three species. Although the exact coordinates of dorsal anterior cingulate cortex (dACC) are not well-defined, it approximately corresponds to the anatomical area 24c and is sometimes referred to as the anterior rostral cingulate zone (Picard and Strick, 1996, 2001). The connectivity pattern of dACC is similar in monkeys and humans. In humans, area 24c patterns of correlation in the BOLD signal and diffusion weighted imaging suggest that it is connected with sensorimotor areas, areas in IPS, dorsolateral prefrontal cortex in particular area 46 and temporal lobes such as temporal pole and posterior temporoparietal junction (pTPJ) (Beckmann et al., 2009; Neubert et al., 2015). Although areas along the cingulate cortex share fundamental similarities in their connectivity profiles, a gradient in coupling can be detected when moving along an anterior to posterior axis: anterior parts show increasing connectivity with dlPFC, parts of vlPFC and insular cortex, while

posterior parts are more strongly connected to posterior cingulate cortex, precuneus, sensorimotor and parietal regions.

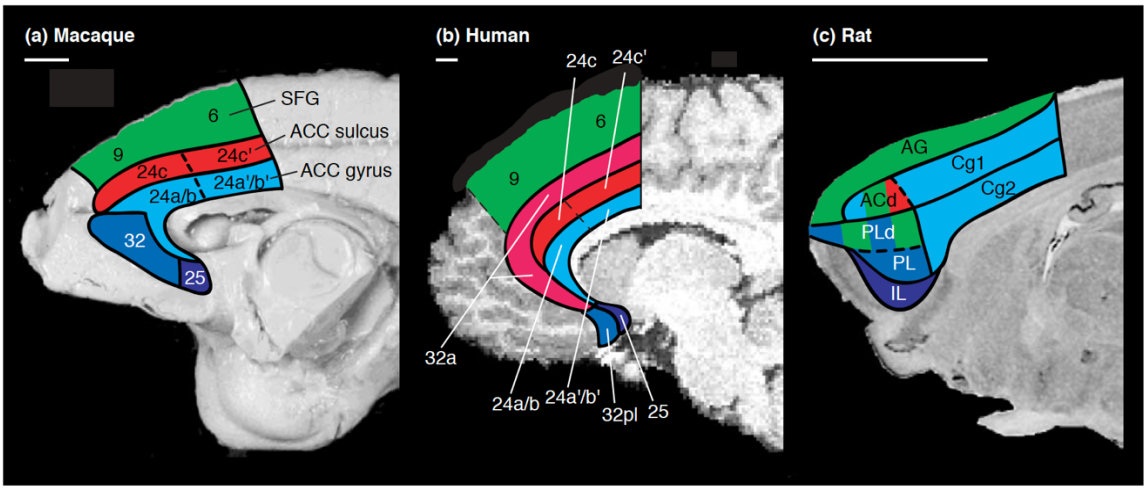


Figure 6. Anterior cingulate cortex in macaque (a), human (b) and rat (c). Dorsal anterior cingulate cortex refers to area 24c and can be compared across all three species. Reprinted from (Rushworth et al., 2004) with permission from Elsevier.

1.5.2.2. Function

The role of dorsal anterior cingulate cortex has been of interest to many researchers in cognitive neuroscience and naturally led to debate amongst them. The first studies that aimed to formalise the role of dACC had measured its activation in tasks that induce response conflict between multiple incompatible strategies such as during the Stroop task (Cohen et al., 1989). According to their observations, dACC had a role in mediating the competition between multiple response tendencies that were active at the same time (Botvinick et al., 2001, 2004). Such an account suggests that dACC's has a role in conflict monitoring, possibly mediating behavioural adjustments due to error detection in previously made responses. The conflict model (Botvinick et al., 2001; Shenhav et al., 2013)

presents dACC as a separate functional unit that monitors conflict. Hence, such a claim can be easily tested by creating tasks that induce conflict and simultaneously measuring activity in dACC. However, evidence across these studies could not clearly demonstrate dACC's role in conflict monitoring. For example, dACC activity was responsive to errors but also reward omission (Ito, 2003), neurophysiological studies did not find cells coding for response conflict *per se* (Ebitz and Platt, 2015; Nakamura et al., 2005), lesions to ACC did not result in change in conflict related behavioural change (Fellows, 2004) or obvious conflict-related impairments (Rushworth et al., 2003). Additionally, other studies were not able to demonstrate that dACC activity was explained as resulting from representing response difficulty *per se* as opposed to the possibility that was it simply representing a number of particular actions and as the evidence for these diverse actions increased so did the activity in dACC (Cai and Padoa-Schioppa, 2012; Hayden et al., 2011b; Nakamura et al., 2005).

More recent accounts suggest a role for dACC in behavioural adaptation by evaluating the appropriateness of a response (performance monitoring) and subsequently adjusting one's response (performance adjustment) (Matsumoto et al., 2007). Many studies showed dACC monitored the reward outcome of a choice (Kennerley et al., 2006; Rushworth and Behrens, 2008). Despite, an initial emphasis on the idea that dACC monitors performance predominantly during failures rather than later studies suggested that this was partly because it was only the errors that were informative for performance adjustment in these early studies; subsequent studies showed that it responded to both rewards and errors if these were informative for future behaviour (Matsumoto et al., 2007). Representing the

direction and amount of error in reward prediction could subsequently contribute to behavioural adjustment (Amiez et al., 2006; Ito, 2003; Matsumoto et al., 2007; Quilodran et al., 2008; Seo and Lee, 2007). Further refinement of this suggestion showed that dACC does not code signed reward prediction errors like dopamine neurons (Schultz et al., 1997), but instead the activity of neurons in dACC correlated with unsigned reward prediction error, or ”surprise” (Hayden et al., 2011b). In some instances, surprise might reflect the occurrence of a “one-time-only” event that might be neglected when learning about the environmental contingencies. In this case, all that is required is that the current choice might need to be corrected. In other cases, however, such an observation might suggest a long-lasting change in the environmental contingencies and should therefore result in a belief update about the environment. One study clearly differentiated between both accounts (surprise vs belief update) and showed that activity in dACC covaried with belief updates; its activity increases with increasing discrepancy between the prior belief and the posterior belief (defined as Kullback-Leibler-Divergence). In comparison, dACC activity did not correlate with surprising events that were defined as one-time-only occasions (defined as Shannon surprise) (O’Reilly et al., 2013). This suggests that dACC updates beliefs about the world, in other words the expectations one has about the contingencies in the environment. In further support of this notion, activity in dACC correlated with the prediction error, the difference between the actual and expected outcome, during outcome presentation and the degree to which future choices were consequently impacted (Behrens et al., 2007).

To detect violations in the expectations one holds about the environment, one has to first form expectations about the current contingencies. This requires the integration and representation of previous experiences. Studies showed that value signals are present at multiple time scales – both longer and shorter – in dACC; activity in some voxels of dACC covaried with expected value based on recent experiences, while other voxels represented value expectation based on more distant experiences (Meder et al., 2017). Recent studies showed that dACC additionally represents long-term expected value. For example, dACC activity reflects the value that was expected to result from a choice not just immediately but towards the end of a series of choices (Kolling et al., 2014). Another study supported this finding by showing that dACC represented a longer term future "prospective" value as well as a more immediate "myopic" value (Kolling et al., 2018). Note that dACC however, is not the only brain area that represents multiple time constants; these were also found in areas such as dlPFC and lateral intraparietal cortex (Bernacchia et al., 2011), consistent with their role in prospective planning.

These results suggest that dACC holds multiple time-linked reward representations, or in other words representations about reward contingencies that are informed by past experiences and which can be used to predict future experiences. These representations can inform behavioural adjustment dynamically as changes in the environment occur. Multiple studies have related activity in dACC to adaptive decisions, showing that dACC encodes the value of the alternative course of action relative to the current or default action and translates such counterfactual choices into actual choices (Fouragnan et al., 2019).

Moreover, dACC has been involved during foraging decisions. Foraging decisions typically have a structure that makes them amenable to computation by the types of evidence accumulation processes reported in dACC. Successful foraging requires the tracking of multiple time-linked reward rates related to the alternative choices that might be taken as opposed to the current choice that is now being taken. Animal behaviour theorists tend to think about animals making choices between foraging patches. dACC represented the average value of the environment and the cost that would be incurred to pursue an alternative choice or patch (Kolling et al., 2012). Further, separate variation in dACC activity could be related to recent rewards and rewards in the more distant past. The relative amounts of activity associated with recent versus longer term rewards reflected whether the value of the current environment was getting better or worse and predicted whether participants would switch away from it to an alternative or stay with it for longer (Wittmann et al., 2016a).

In sum, these results show that dACC has a role in monitoring and updating beliefs about the environment by representing multiple time-linked reward representations which guide behavioural change.

1.5.3. Ventromedial prefrontal cortex

1.5.3.1. Anatomy

There has been a considerable variability in the definitions that have been given of ventromedial prefrontal cortex (vmPFC). Ultimately this reflects the fact that the term vmPFC originates not in cytoarchitecture but in attempts to describe a spatial region within the human prefrontal cortex. The lack of precise consensus

about its boundaries makes many meta-analytic approaches that often used to aggregate evidence across studies into a particular brain area less efficient than usual. One perhaps successful approach to determine the location and boundaries of brain areas is according to their DTI-connectivity-based parcellations with the whole brain (Beckmann et al., 2009; Neubert et al., 2015). In such a case, a cluster dorsal to the rostral sulcus anterior and ventral to the genu of corpus callosum could be identified as the ventromedial part of the prefrontal cortex (Figure 6A, green cluster). This cluster overlapped with previous studies in which activity correlated with reward expectations (Basten et al., 2010; Boorman et al., 2009; De Martino et al., 2013). Additionally, it showed a different connectivity pattern compared to adjacent areas such as subgenual (Figure 6A, orange cluster) or perigenual ACC. Other studies have decomposed the ventromedial surface into multiple areas (Figure 6B, left panel), of which 14m resembles the DTI-based parcellation found in the previous study by Beckmann et al. (Mackey and Petrides, 2014; Öngür et al., 2003). This area maps onto a homologue version of the macaque's area 14m (Figure 6B, right panel). Other studies that based their parcellation on diffusion-weighted MRI, have identified slightly different clusters within the medial wall (Figure 6C); compare areas panels of Figure 6A-C.

To locate the vmPFC and map functional activity onto anatomical correct areas, it has been shown to be useful to take into account cross-individual variations in morphological patterns. Interindividual differences in morphological patterns are neglected in classic fMRI studies because activation patterns are mapped onto standardized templates (e.g. MNI templates). Previous studies have shown that for other brain areas, such as lateral prefrontal cortex (Amiez and

Petrides, 2017) or midcingulate cortex (Amiez et al., 2013), functional properties in this area can be predicted from individual morphological patterns. Similarly, it has been shown that vmPFC exhibits distinct morphological patterns across individuals (Lopez-Persem et al., 2019) and that these differences predict differences in resting state connectivity. In this study, Lopez-Persem et al. analysed the presence/absence of sulci, their shape and relative position and showed that the peak activation of the default mode network (DMN) was located in the anterior part of the ROS-S sulcus – hence, indicating useful mapping between this particular sulcus and the DMN peak activation (Figure 6D). This means, that peaks in activation patterns can benefit from being defined in relation to individual participant sulci morphology (Lopez-Persem et al., 2019). In addition, taking individual local sulcal morphology might explain cross-individual differences in connectivity profiles and behavioural measures reported elsewhere (Bijsterbosch et al., 2018).

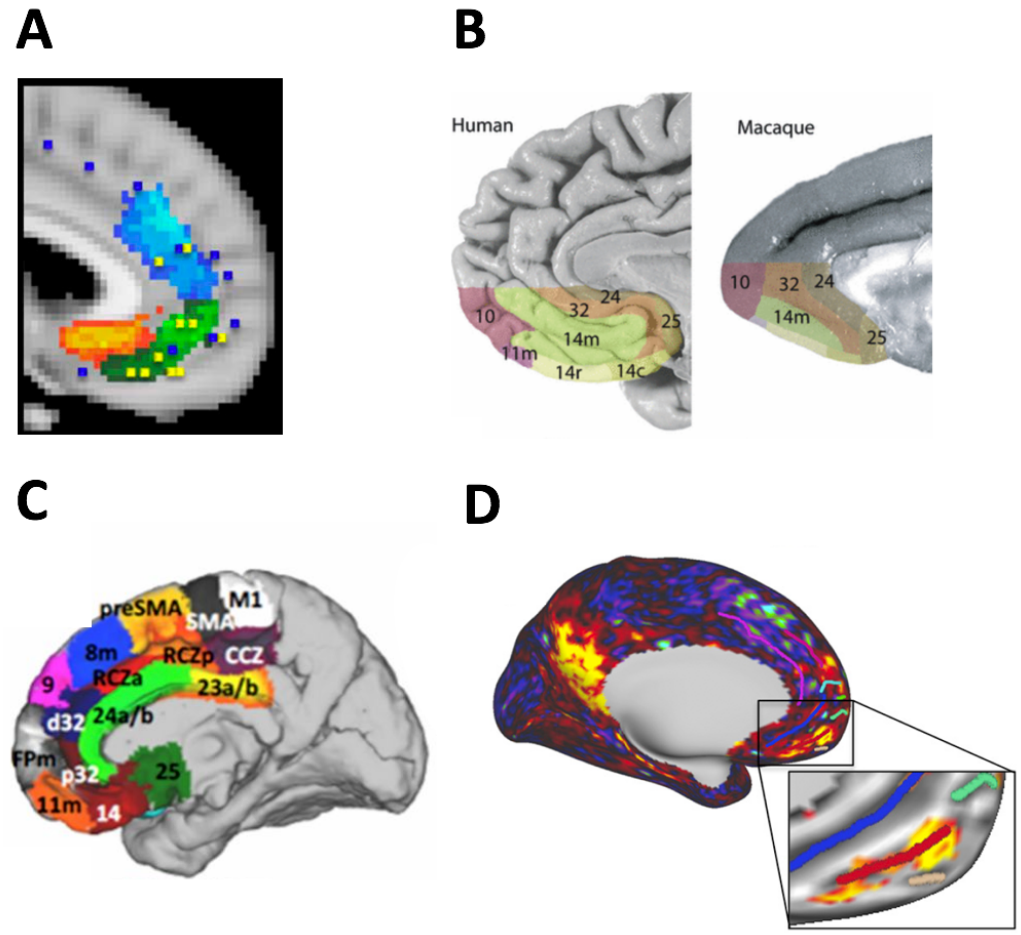


Figure 7. Anatomy and sulci morphology of vmPFC. (A) DTI-connectivity-based parcellation identified a region dorsal to the rostral sulcus anterior and ventral to the genu of corpus callosum (green cluster). Figure extracted and modified from (Beckmann et al., 2009). Copyright [2009] Society for Neuroscience. **(B)** Architectonic parcellation of ventral medial frontal surface. Reprinted from (Mackey and Petrides, 2010) with permission of John Wiley and Sons. **(C)** Diffusion-weighted MRI parcellation of the medial and orbitofrontal areas. Figure extracted and modified from (Neubert et al., 2015). Copyright [2015] National Academy of Sciences. **(D)** Subject example showing the mapping of default mode network in respect to different sulci (hot colors indicate high weights indicative of default mode network); the stripes represent the different sulci within vmPFC and show that the peak activation of the default mode network is located within the medial wall of frontal cortex and corresponds to the ROS-S sulcus (red stripe). Figure extracted and modified from (Lopez-Persem et al., 2019).

1.5.3.2. Function

Many studies identify activity reflecting the subjective values of choices considered during decision making or even, in some cases when choices are passively considered. This neural representation is observed independent of

whether it is measured with explicit reports (Lebreton et al., 2015) or with variables derived from computational models (Kable and Glimcher, 2007; Rushworth and Behrens, 2008; Wunderlich et al., 2012). One reason why it has been identified in many studies is because it represents subjective value independent of the stimulus category (object, faces, paintings, food etc.). Another reason is because activity in vmPFC is detected during a variety of value-based decisions: binary choices, sequential choices or even when simply witnessing items (Lebreton et al., 2015). Some studies have suggested that the activity reflects the value options that are available for choice (Hare et al., 2011) while other studies have argued that they reflect the value difference between options (Chapter 3; (Boorman et al., 2011)). These different views have been reconciled by high temporal resolution recording studies that have suggested that both representations occur during the decision process (Hunt et al., 2012). In line with the predictions of field summaries of cortical attractor models, activity in vmPFC activity initially reflects the value of both options at the start of a decision then, as the decision takes place, the activity reflects a competition process between the representations of the two choices and the vmPFC's aggregate activity reflects the difference in the options' values. Ultimately this leads to a choice being taken. Related patterns of activity have been reported in single unit activity recorded in adjacent orbitofrontal cortex in monkeys (Rich and Wallis, 2016) and lesions of the region compromise decision making (Noonan et al., 2010, 2017).

Generally, there is an emerging consensus that vmPFC reflects value difference between options as a result of a role that vmPFC has in guiding choice selection. However, there are different suggestions about how this value

comparison occurs. Some studies show that the value comparison is indexed by aggregate activity differences (blood oxygen level dependent [BOLD] signal differences in fMRI studies) between the value of the chosen and unchosen option. When the values of the option that is chosen and the option that is rejected are separately regressed onto the fMRI activity, activity correlates positively with the value associated with the chosen option and negatively with the value associated with the unchosen option (Basten et al., 2010; Boorman et al., 2009). Other studies suggest that the value comparison between options is constructed as the difference between a default and an alternative option (Kolling et al., 2012; Lopez-Persem et al., 2016). The default option is the preferred option and when making a binary choice between preferred and non-preferred options, vmPFC activity has been shown to be influenced by features of the preferred option even when they were ultimately irrelevant for the decision (Lebreton et al., 2009). Hence, vmPFC representation can be influenced by information learnt prior to the choice (Lopez-Persem et al., 2016; Plassmann et al., 2008). Finally, yet another account suggests that the value comparison between options is biased by attention at the time of choice (Lim et al., 2011; McGinty et al., 2016). Careful analysis of single unit activity recorded during more complex multi-faceted decision problems suggests that the value comparison-related activity may emerge in a framework determined by the direction of attention (Hunt et al., 2018) although often attention is likely to be directed towards the option that is ultimately chosen.

Accounts of vmPFC activity that focus on attention or on decisions are also not necessarily mutually exclusive for other reasons. For example, one study showed a bidirectional link between learning and attention and its consequences

on choice behaviour: attention impacted which choice features were attended and thus how values were computed during decision making and, subsequently, how these value were updated at the time of the decision outcome (Leong et al., 2017). Attentional switches amongst options were represented in frontoparietal networks. In the absence of attentional switches, the connectivity between vmPFC and these frontoparietal areas was anticorrelated which the authors argued could be interpreted as indicating that vmPFC is involved in attending to the current option. They argued that such activity patterns might modulate the likelihood of persevering with the current choice pattern as opposed to switching to the alternative option. In comparison to dACC and lateral frontal cortex, vmPFC has been associated with representing value that is relevant for current, immediate choice behaviour. For example, activity in vmPFC increased when staying with the current option compared to switching to an alternative choice in a binary choice task (Boorman et al., 2009). During sequential decisions, activity in vmPFC increased when engaging with the current option compared to foraging amongst alternative options (Kolling et al., 2012). Similarly, vmPFC is involved when exploiting known options as opposed to exploring alternatives (Daw et al., 2006).

It is, however, difficult to reconcile the fact that vmPFC is relevant for computing value comparison to guide decision making but that there is a lack of evidence for vmPFC's involvement during exploration. During exploratory choices, information such as the uncertainty (or certainty) between options is computed to guide current behaviour (Badre et al., 2012) and there is evidence that uncertainty in value estimates impact on vmPFC activity (Lebreton et al., 2015). The lack of evidence of a vmPFC role during exploration might be for two reasons. First, as

discussed earlier, it has been difficult to agree on definitions of exploratory choices; definitions have been made in relation to both their value (Daw et al., 2006) and their uncertainty (Badre et al., 2012) but both these features of choices can be correlated particularly at certain points in time during a sequence of decisions. Second, many investigators have assumed that there is a difference between information-based and value-based representations and have considered the vmPFC as only caring about value-based information. However, if information is itself of value then it might be asked whether this is reflected in vmPFC activity. This argument is consistent with the results of a recent study which showed that the subjective value of information is represented in a network that usually represents an option's expected value (Kobayashi and Hsu, 2019). VmPFC and ventral striatum represented the subjective value of information, even when controlling for the expected reward. This is consistent with previous accounts that have argued that information itself has an inherent value (Bromberg-Martin and Hikosaka, 2011; Bromberg-Martin and Sharot, 2020). In Chapter 4, I show that vmPFC in fact represents whatever decision variable is relevant for guiding choice regardless of whether this is during exploration or exploitation. Instead I show that the relevance of various decision variable changes depending on the behavioural modes of exploration and exploitation and this change in decision variable is reflected in vmPFC.

A related line of research additionally supports the notion that vmPFC not only represents the value difference between choice options, but also a second-order representation, that is one's own confidence (or certainty) in a choice. Some studies showed a representation of confidence in vmPFC (De Martino et al., 2013),

while others found correlates of confidence in lateral intra-parietal (LIP) neurons (Kiani and Shadlen, 2009) or pgACC (Bang and Fleming, 2018). Discrepancy amongst studies might be due to the nature of the task or the definition of confidence. Confidence can be measured in multiple ways, for example with explicit reports (De Martino et al., 2013; Fleming et al., 2018), the waiting time for a reward (Lak et al., 2014; Stolyarova et al., 2019), or in relation to the way that reaction time is impacted by difficulty (Kiani et al., 2014; Sanders et al., 2016). One study that showed robust representation of confidence in addition to value representation was conducted by Lebreton et al. (2015). They used multiple rating tasks that asked participants to make a judgment about a specific object, for example to judge the age of a person or the pleasantness of a food item. Across all these different types of rating tasks and even in the absence of an explicit choice, vmPFC represented the confidence in one's judgment (Lebreton et al., 2015), suggesting an automatic evaluation of an item independent of choice (Barron et al., 2015). This finding – that vmPFC encodes choice confidence in addition to choice value - is similar to observations that dACC encodes surprise or difficulty (both unsigned variables such as confidence) in addition to rewards or losses. In both cases, these brain regions encode higher-order value signals that arise as a function of the other value signals observed in these regions. It is possible that the brain computes these variables explicitly (although there is little neurophysiological evidence for this in the case of dACC, see: (Ebitz and Platt, 2015; Nakamura et al., 2005)). However, it is also possible that such fMRI signals arise as by-product of canonical computations that are similar across different brain regions (Hunt and Hayden, 2017) .

1.5.4. Summary

In sum, aPFC and dACC have been shown to track and represent currently counterfactual choices that might be taken in the future choice if behaviour changes. aPFC activity has often been recorded during exploratory choices, in particular when they are part of prospective planning (e.g. during long time horizons). dACC tracks multiple time-linked action-outcomes associations that can inform prospective planning during sequential choice behaviour such as foraging. Moreover, activity in dACC increases with violations of expectations, suggesting a role in updating beliefs about environmental contingencies which consequently leads to behavioural change. Both dlPFC and dACC increased in connectivity when representing the prospective value of an option (Kolling et al., 2018). In contrast, vmPFC has a different role in representing the features of a potential choice that should guide decision making in the present moment rather than later in an unfolding plan. Therefore, vmPFC has often been associated with exploitation of the current option with the highest value. However, I will show that its activity reflects decisions variables relevant for guiding exploration too. Further, it is important to note that vmPFC does not only hold a single representation of an option, for example its activity does not just reflect is subjective value, but instead holds multiple representations, some of which are related to confidence in the current choice.

1.6. What is unique about social processing?

It is especially important for behaviour during social interaction to be flexible because we do not only have to take into account our own perspective and motivation when choosing a course of action but also those of other people and this can have an impact on our preferences (Garvert et al., 2015; Izuma and Adolphs, 2013), learning (Behrens et al., 2008; Boorman et al., 2013a; Wittmann et al., 2016b), decisions (Hill et al., 2017) and confidence (Chapter 4; (Bang et al., 2017; Kappes et al., 2020; Will et al., 2017)). We adapt and reorient our behaviour according to our perceptions of others, and we do so in different ways depending on whether they are present at the moment or whether we have experienced interactions with them in the past.

One question which naturally arises is whether processing information that is social is different or in some way unique compared to processing information that is of non-social nature. Or, more specifically in the context of this thesis, whether exploration or exploitation of information sources are different when those sources of social information are other social agents as opposed to non-social information sources. Differences between social/non-social processing can be assessed on a variety of levels. First, one can compare behavioural parameters such as decisions or reaction times or computationally derived variables and this alone may provide insight into the mechanisms underlying behaviour. Second, one can compare the neural activity that is associated with a behaviour that is produced in a social or non-social context. Different patterns of behavioural and neural results might suggest quite different interpretations and the degree to which social and non-

social information processing is the same or different. For example, differences across both behavioural and neural levels might suggest fundamental differences between social and non-social conditions – however, these fundamental differences can be due to different sub-levels underlying the decision process: the implementation of a decision (e.g. decision rule), the information decisions were based on (e.g. learning rule/ updating rule) or a specific type of sensory input information (for example face stimuli might be especially important in social decision making contexts). In the last case, it might also be the case that despite a lack of behavioural differences there are differences in the neural networks that are associated with the behavioural output: perhaps in these cases, one might conclude that computations are similar (e.g. decision rule, learning rule), but that additional brain areas are required to represent the socially-relevant data that are an input to a shared social/non-social decision making mechanism.

Here, I will review whether, a priori, we should expect to find unique neural substrates for social processing. Here, it is not the aim to discuss different brain areas in detail and whether they are concerned with processing more or less social information, but instead present theories that give an account of how social information is processed compared to non-social information.

1.6.1. Social brain hypotheses and alternatives

Compared to vertebrates, primates have an unusual large brain compared to the rest of their body size. The social brain hypothesis (SBH) therefore puts forward the idea that the increase in brain size has to be understood in reference the evolutionary pressures imposed by complex social life (Dunbar, 1998). There is evidence that at least some parts of the brain are specifically related to negotiating

social situations (Noonan et al., 2014; Sallet et al., 2011; Sliwa and Freiwald, 2017). For example, evidence for this notion was provided by neuroimaging studies in macaques which showed a positive correlation between group size and areas in temporal lobe, such as mid-superior temporal sulcus (STS) (Sallet et al., 2011), an area that shares many network features with the human temporoparietal junction (TPJ) region (Mars et al., 2013) which has, in turn, been associated with “theory of mind” (Schurz et al., 2017). Another area that increased with group size was rostral prefrontal cortex – an area, as previously discussed involved in behavioural flexibility. Other studies showed that the correlation between group size and brain size are not only observable in macaques, but also in humans (Bickart et al., 2011; Lewis et al., 2011).

Areas in which activity is frequently reported in studies of social cognition include the posterior temporoparietal junction (pTPJ) or areas within certain parts of the dorsomedial prefrontal cortex and anterior cingulate cortex. For example, the gyrus (ACCg) and sulcus (ACCs) of anterior cingulate cortex seem relatively more concerned with social and non-social information respectively (Apps et al., 2016). Lesion studies (Rudebeck, 2006), neurophysiology studies in monkeys (Chang et al., 2013) and neuroimaging studies (Lockwood and Wittmann, 2018; Lockwood et al., 2018) showed that while ACCs is associated mainly with self-referenced decisions, ACCg has been implicated in representing attributes or rewards assigned to others.

According to Ruff et al (2014), a strong interpretation of such results is that there are networks that are associated with social processing while other networks are for non-social processing (Figure 7B). The SBH hypothesis has, however, been

challenged by alternative ecological explanations that suggest that primates have larger brains because of the evolutionary importance and complexity of foraging skills (Reader et al., 2011). Both ecological and SBH approaches are based on evolutional principles, however it is difficult to decide which of these processes determines the reason for larger brains and it is, of course, that primates are distinguished by the complexity of their social organization and the sophistication of their foraging strategies.

A potential problem for advocates of socially specific neural/behavioural processes is that there are many studies that suggest there are many common brain areas that are linked to reward and motivational processes in the context of both social and non-social behaviour (e.g. ventral striatum, amygdala, vmPFC, insula) suggesting at the very least that social information processing is not completely separate (Behrens et al., 2008; Boorman et al., 2013a; Ruff and Fehr, 2014; Will et al., 2017; Zink et al., 2008).

An intermediate position is that there might be brain areas uniquely representing social information but that these areas feed into common networks shared between social and non-social information processing (Figure 7A).

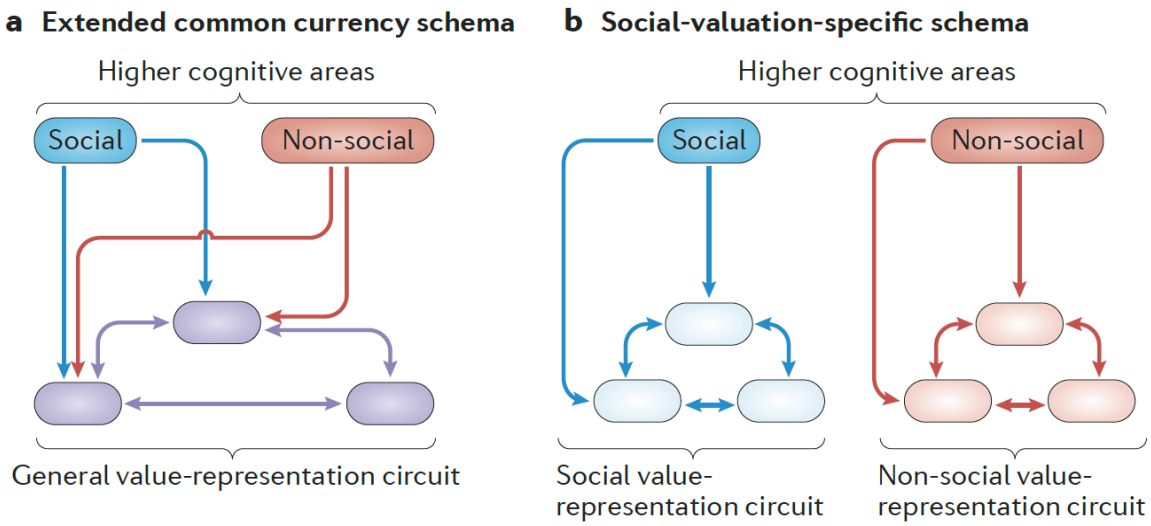


Figure 8. Two hypotheses that integrate social and non-social information to construct value representation. (A) General value representation suggests that there are brain areas representing social or non-social information that feed into a network that underlies both conditions. **(B)** Alternatively, social and non-social processes could be encoded in different networks – a hypothesis in accordance with the social brain hypothesis. Reprinted from (Ruff and Fehr, 2014) with permission from Springer Nature.

In summary, while there is evidence that there are brain areas that are mostly concerned with social processing (e.g. pTPJ, ACCg) many are equally important during both social and non-social conditions (e.g. vmPFC, striatum). However, as most comparisons are drawn from fMRI studies, it still remains somewhat unknown whether activation in 'common brain areas' reflects similar or different set of underlying neuronal populations. Different measures such as electrophysiology or multivariate-pattern analysis would be required to make such a comparison and give insight into whether a similar average activation reflects similar or different sets of neuronal populations.

1.7. Methods to measure neural activity

There are different approaches to assess brain activity. One can assess brain activity either on the level of neurons as single units, population of neurons or more broadly, on the level of a brain area and its network. In this thesis, I used functional magnetic resonance imaging (fMRI) to assess neural activity within brain areas and their networks. I will give a short description of what fMRI entails.

1.7.1. Functional magnetic resonance imaging

Functional magnetic resonance imaging (fMRI) has been one of the most frequently used techniques in cognitive neuroscience in recent years. It allows identification of activity in multiple brain regions simultaneously. Neural activity is measured indirectly by detecting changes in the BOLD signal. It relies on the fact that active neurons require more oxygen, which is then provided by increased blood flow to the active brain region, thereby changing the ratio of oxygenated and deoxygenated haemoglobin in the blood (Huettel et al., 2004). The BOLD signal follows a hemodynamic response function (HRF), which describes the temporal evolution in the difference between deoxygenated and oxygenated haemoglobin levels. The human HRF peaks around 6 seconds after the initial events that trigger neural activity to occur, however this might vary across brain area. It is suggested that activity measured with fMRI reflects changes in the post-synaptic activity in neurons dendrites – in other words their inputs from other neurons which are measurable with local field potentials (LFPs) rather than their outputs which are indexed by their spiking activity (Logothetis et al., 2001). It has been shown that

higher frequency bands such as high-gamma are especially closely related to the BOLD signal (Mukamel, 2005; Nir et al., 2007).

fMRI provides evidence of correlations between brain activity and performance of an experimental task. To assess the role of a brain area, one can use a variety of designs (block design, event-related designs) and a variety of methods to correlate activity with the task (correlation with variables, subtraction methods). In this thesis, I have used an event-related design and parametric variables derived from computational models that provide trial-by-trial values that can then be regressed onto brain activity in a univariate analysis (i.e. general linear model).

One disadvantage of fMRI is its poor temporal resolution compared to other techniques (e.g. electrophysiology) due to the sluggishness of the BOLD response and the fine time delays associated with acquiring each brain slice of fMRI data. Although its spatial resolution is better than that of other non-invasive techniques used to look at the human brain (e.g. EEG), it is still limited by the grid size of functional images (i.e. voxel size). However, to reduce some of these disadvantages, fMRI can be combined with other temporal neuroimaging techniques (e.g. EEG) to improve the temporal resolution; it can be combined with neurostimulation methods (e.g. online or offline TMS protocols) to assess the causal contribution of targeted brain areas and it can be combined with multi-variate pattern analyses to improve spatial resolution when investigating the nature cognitive representations (Kriegeskorte and Douglas, 2018).

2. Chapter: The impact of subjective beliefs on exploration and exploitation across time

2.1. Introduction

Humans and other animals are often presented with multiple information sources in the environment that can predict different outcomes such as reward. Selecting the right predictor to guide behaviour towards a particular outcome requires determining the predictors' relevance in forecasting that outcome (Akaishi et al., 2016; Leong et al., 2017). Biases in information seeking can lead to mistaken beliefs about the relationships that prevail in the world (Charpentier et al., 2018; Garrett et al., 2018). It has been argued that animals should attend either to certain predictors (Mackintosh, 1975) or, on the contrary, to uncertain predictors (Pearce and Hall, 1980). Certain predictors might be relevant as they deliver an outcome with known prediction accuracy, while attending to uncertain predictors might turn out to be more beneficial in the long-term.

We propose that which type of predictor should be considered most relevant changes during different phases of the learning process. When selecting between multiple predictors for the first time, selections should maximize information about available predictors. Selections should be explorative and directed towards uncertain predictors. The degree of exploration should also be determined by the time horizon. The time horizon is the remaining time in the current context (or block in the current experiment) (Kolling et al., 2018; Wilson et al., 2014a): exploration is

beneficial in longer compared to shorter time horizons as the knowledge gained can be used in later predictor selections. Once an estimate about a predictor's accuracy has formed, selections should be exploitative and guided by the accuracy and certainty of predictors in line with reward maximization. This perspective draws on both previously formulated hypotheses in the field of learning theory (Mackintosh, 1975; Pearce and Hall, 1980). Predictors should be selected based on the learner's uncertainty about predictors' accuracy during exploration and on the learner's certainty about predictors' accuracy during exploitation. The first aim of the study was to provide evidence that humans' predictor selections are guided by beliefs of their un(certainty) and accuracy.

Evidence for uncertainty-guided exploration has, however, recently been questioned (Findling et al., 2019). It has been argued that behaviour may sometimes appear exploratory but on closer inspection, the decisions that people make can be understood as having been guided by noisy estimates of the values of the choices that are formed during learning. In other words, when people appear exploratory they may in fact be attempting to make exploitative decisions but their exploitative decisions are informed by noisy estimates of choice values. The second aim of the study was to ascertain whether people genuinely engage in exploratory behaviour. This can be tested by comparing rates of exploratory behaviour when past experience is held constant but the length of the future time horizon is manipulated; a longer future time horizon should elicit more exploration even when previous learning opportunities are the same. Moreover, the appropriateness of computational models of exploratory behaviour can also be tested by obtaining more direct empirical indices of participants' subjective

uncertainty; we obtained such measures in our experiment. In addition, the computational model can be used to identify trials in which exploratory behaviour appears to be guided by information seeking in order to reduce uncertainty and trials in which exploratory behaviour simply reflects randomness in the response selection or learning process.

This chapter introduces a novel experimental paradigm that was used to investigate which predictors are classified as informative according to subjective beliefs about the predictors. It focuses on the computations underlying choice behaviour during the experiment (Trudel et al., 2020). The subsequent chapter (Chapter 3) presents fMRI data from the same experiment, demonstrating the neural correlates of uncertainty and accuracy estimates during different behavioural modes of exploration, exploitation, and the period of transition from exploration to exploitation.

2.2. Methods

2.2.1. Participants

Thirty participants took part in the experiment. Participants were excluded because they fell asleep repeatedly during the scan (N=2), exhibited excessive motion during the scan (N=1), or because of premature termination of an experimental session (N=3) (final sample: 24 participants; 14 female). The study was approved by the Central Research Ethics Committee (MSD-IDREC-C1-2013-13) at the University of Oxford. All participants gave informed consent.

2.2.2. Experimental procedure

Participants took part in two magnetic resonance imaging (MRI) sessions on separate days. The two task versions of the experiment differed in the type of predictor stimuli. In one version, faces represented advisors that predicted a target accurately or less accurately. In a second version, fruits represented the target that fell with a small or large distance from a tree. Differences between versions aimed to compare social and non-social conditions. Both versions only differed in their framing but were matched in statistical properties. All participants completed both versions and their order was counterbalanced across subjects. Here, the results reported in the current chapter and Chapter 3 present data averaged across conditions (as they do not differ between conditions). Chapter 4 focuses on differences between conditions during a separate aspect within the experiment.

Each session lasted approximately two hours, including one hour of scanning. Participants received £15 per hour and a bonus based on task performance (per session: £5 - £7). Before each scanning session, participants were instructed about

the task and performed seven practice trials outside the scanner. After completion of both sessions, participants filled in a questionnaire that assessed their understanding of the study.

2.3. Experimental design

2.3.1. Trial timeline

We designed an experiment to investigate which predictors are classified as informative, uncertain, or certain, as a function of time in a block. Each trial included a decision, confidence, and outcome phase (Figure 1A). On each trial of the experiment, participants made two decisions. First, they made a binary choice between two predictors to find a target's location on a circle (decision phase; duration: 2sec + reaction time (RT)). Participants knew that the target location changed constantly on every trial and could not be predicted directly from previous observations of its location. Therefore, the only way to infer the target's location was through learning how well each predictor predicted the target location. Participants learned how well a predictor predicted the target by observing the distance between the location estimated by the selected predictor and the true target location (which we refer to as angular error). Importantly, predictors differed in how well they estimated the target location. Selecting a better predictor led to more rewards at the time of a second decision in the trial. During the second decision, the predictor's estimate (indicated as black dot) of the target location was revealed and participants expressed their confidence in it (confidence phase; duration: RT). They did this by adjusting the size of an interval around the predictor's estimate such that the true target location would fall within this interval.

At the end of a trial, the true target location and points earned or not earned on that trial were revealed (outcome phase; duration: 3sec). Participants gained points when the target fell within the chosen interval and the amount of points increased when the interval size was small. The payoff scheme incentivised selecting predictors with smaller angular errors in the first place as participants could maximize their points by decreasing the interval size. In addition to being informed about whether they had won or lost, the outcome phase enabled participants to update their beliefs about how well the chosen predictor estimated the target by observing the angular error.

To decorrelate variables of interest between trial phases, short intervals were included between trials (inter-trial-intervals) and within trials randomly, but equally frequently, allocated to either the transition between decision- and confidence phase or confidence- and outcome phase. The duration of an interval was drawn from a Poisson distribution with the range of 4s to 10s and a mean of 4.5s. During these intervals, a fixation cross was shown on the screen.

Note that I will focus on the decision phase during Chapters 2 and 3, while Chapter 4 analyses behavioural and neural effects associated with the confidence phase.

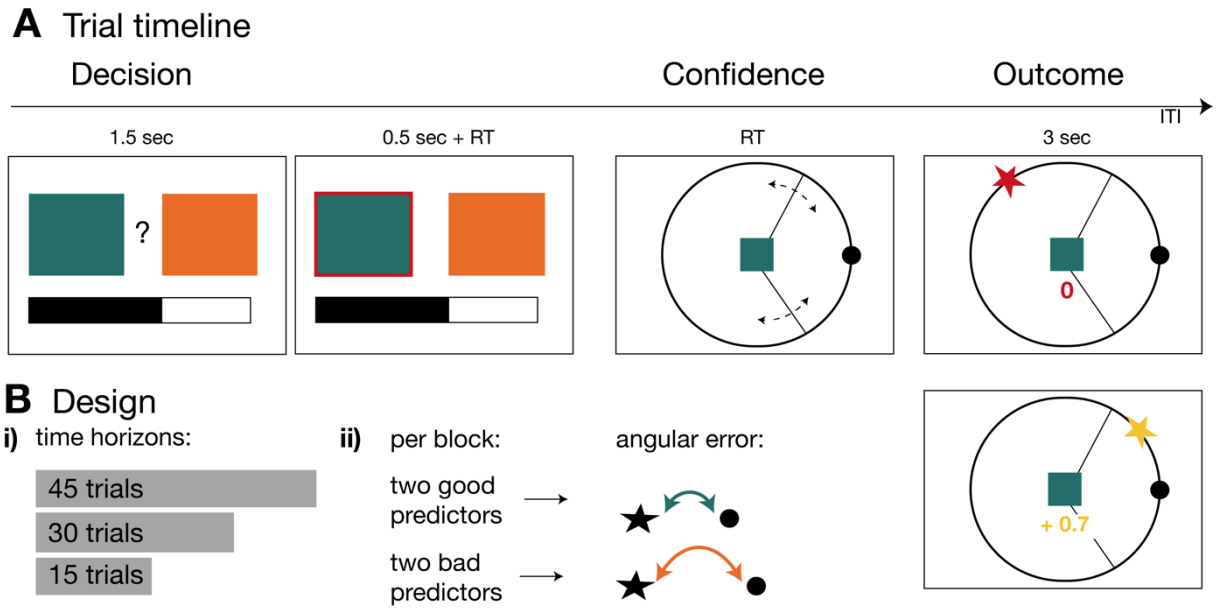


Figure 9: Experimental design: trial timeline and design. (A) *Trial timeline. In each trial, participants made two choices. First, a binary choice between two predictors (colored boxes; decision phase) to receive information about a target's location on a circle. The goal was to choose predictors that accurately predicted the target location. The length of a black bar at the bottom of the screen informed participants about the number of remaining trials in the current block. Second, participants indicated their belief in the accuracy of the chosen predictor by modifying the size (dotted lines) of an interval symmetrical around the reference point (confidence phase). In the outcome phase, the target location (star) and any points earned were indicated. Two possible example outcomes are illustrated. In the above case, the participant's prediction was incorrect as the target fell outside the interval, resulting in a null payoff. In the bottom case, the target fell within the interval, resulting in a positive payoff. Positive payoffs increase with narrower intervals as long as the target falls within the interval.* ***(B)*** *Design.* ***(B-i)*** *Participants transitioned through blocks of different numbers of trials (time horizons).* ***(B-ii)*** *Each time horizon introduced four new predictors that were categorised into two good and two bad predictors according to how well they predicted the target. The quality of predictions was determined by the angular error between target and reference location with a smaller angular error representing better target predictions (right illustration).*

2.3.2. Design

We hypothesized that the tendencies to explore or exploit a predictor should be guided by knowledge about the predictors as well as the time horizon over which a predictor is encountered. Participants should exploit and choose accurate predictors; however, they should also explore alternative predictors in case they are even better. The value of exploration lies in revealing more accurate predictors,

but this is only useful if the time horizon (the number of trials remaining) offers sufficient opportunity to exploit the newly discovered predictors (Wilson et al., 2014a). To test this idea, participants transitioned through blocks of different lengths (45, 30 and 15 trials) each with a unique set of four predictors (Figure 1B-i). This made it possible to examine the balance between exploration and exploitation as a function of time horizon. Time horizon and current progress were explicitly cued on each trial: during the decision phase, the time bar (bar below the predictors) indicated the number of trials left in the current block; it decreased after each trial until the end of a block. At the start of a new block, the type of horizon was identifiable by inspecting the time bar. The length of a block (time horizon) was either short (15 trials), medium (30 trials), or long (45 trials) (Figure 1B-i). Each time horizon was presented twice (total of 180 trials) and their order was pseudo-randomised with the constraint that blocks of the same horizon did not succeed each other directly. There was only one temporal order of predictor presentation: the orders for short and medium horizons were extracted from the long horizon such that the first 15 trials were identical across horizons.

Multiple and repeated selections of the same predictor could result in estimates of uncertainty and accuracy that might, respectively, decrease and increase across time. This would result in possible correlation between variables and with time. This, however, is only appplicable in specific circumstances: if participants were to be repeatedly offered the same pair of predictors and if they were to repeatedly choose the good predictor above the bad one. When designing the experiment, the order of predictors was carefully constructed such that variables of interest, the accuracy and uncertainty estimates, were decorrelated

statistically and across time. First, we employed multiple predictors rather than a single predictor in our task design and we asked participants to choose between different possible pairings of predictors. For example, on one trial participants might choose between predictor A and predictor B but on the next they might choose between predictors C and D then, on the next trial, between predictors B and C. The same predictor will therefore not be chosen repeatedly. Second, we introduced the four predictors associated with different sigma values at different times during the block and the net result of this is that the set of predictors available on each trial are associated with a range of accuracy and uncertainty estimates. Changing the predictor offers across time creates a variety in prediction differences during early and late phases within a block. The critical correlations between accuracy and uncertainty prediction differences are r= 0.1 across all trials (Figure 4B), r= 0.37 within exploration and r= 0.30 (Figure 6B, left and right panel respectively) within exploitation, on average across participants. This means that the maximum shared variance in these conditions is 0.14 (in exploration). Finally, we used time horizons of different lengths which also affected whether choices were of the more or less uncertain/accurate predictors. In a more general sense, one could ask how variables of accuracy and uncertainty prediction differences develop across time or whether they are correlated with a linear trend, i.e. '% remaining trials' variable. Both prediction differences of accuracy and uncertainty correlate respectively, r=0.12 and r=0.3 with '% remaining trials' (i.e. capturing time that decreases with every trial within a block) across participants. Hence, results depicted are not confounded by time .

The aim was to select predictors that gave a precise prediction of the target location. How good a predictor was, was determined by how well it estimated the target in the confidence phase. Estimations followed a Gaussian distribution centred on the true target location (Figure 1B-ii; Figure 2A-i,ii). Values, *x*, for each predictor were drawn from a Gaussian distribution and represented the difference between the true target location and the predictor's estimate (i.e. the angular error):

$$x \sim N(\mu, \sigma) \text{ with } -180 < x < 180 \quad (1)$$

where at a given trial, value *x* was derived from a normal distribution with mean of $\mu = 0$ and sigma of either $\sigma = 50$ for good predictors or $\sigma = 70$ for bad predictors. Predictors' normal distributions varied in their standard deviations (referred to here as sigma), making some predictors better in estimating the target location (lower sigma value) and other predictors worse (higher sigma value) (Figure 2A-ii). Hence, by tracking the angular errors of a predictor, participants could estimate the sigma value associated with each predictor's distribution. As participants learned about the predictor's performance through observing the angular error, they learned about the sigma of each predictor's distribution (Figure 2A-iii).

Participants maximized their points by decreasing the interval size during the confidence phase (Figure 1A). The interval covered a portion of the circle symmetrically around the reference point. The interval size was randomly initiated on each trial between a minimum of one and a maximum of 20 steps away from the predictor's estimated target location. Participants changed the interval size with a precision of up to 20 steps on each side of the reference location, that is a maximum of 40 steps as the interval was set symmetrically. A step size was

derived by dividing the circle size (6.3 radians) by the maximum number of possible steps, resulting in a step size of approximately 0.16 radians. The interval size was determined like follows:

$$\text{Interval size} = (\text{number of steps} \times 2)/ 40 \tag{2}$$

When the target fell within the interval set by the participant, the magnitude of the payoff was determined by subtracting the interval size from 1. However, if the target fell outside the confidence interval, it resulted into a null payoff. This meant that the payoff per trial ranged between 0 and 1.

$$\text{Payoff} = \begin{cases} (1 - \text{interval size}) & \text{if target is included} \\ 0 & \text{if target is excluded} \end{cases} \tag{3}$$

2.4. Bayesian Model

Exploration should not only be guided by one's belief in the predictor's accuracy, but also by one's own uncertainty in that belief. For this reason, we used a Bayesian model to capture participants' belief distribution over the angular error between the reference point and the true target location (Figure 2A-i). Sigma (σ) refers to the standard deviation of the normal distribution from which observations of the angular error were drawn, i.e. distance between target and reference location at each trial. Participants learned about how well a predictor predicts the target location across time and by doing so, they implicitly estimated the sigma value (σ) of the distribution (Figure 2A-ii). Using a Bayesian model, we derived subjects' beliefs about the sigma value (σ) of each predictor's distribution, resulting

in sigma-hat ($\hat{\sigma}$) that denotes participants' estimated sigma (Figure 2A-iii;2B). Before a belief can be formed, participants selected a predictor and then made an observation x of how good the predictor was on a given trial, defined by the angular error between the true target location and the predictor-estimated location (reference location):

$$x \text{ (angular error)} = \text{reference point} - \text{true target location}, \quad (4)$$

where the reference point indicated the predictor's prediction of the target location (represented as black dot, Figure 1A). Key features of beliefs can be captured by a probability density function (pdf) over sigma (Figure 2A-iii,iv; 2B). The parameter space comprised possible sigma values that could be estimated by the participant. The parameter space of sigma was bound between 1 and 140 degrees to allow a broad range of sigma values considering the circle shape.

Following Bayes' rule, a belief is updated by multiplication of a prior belief and a likelihood function resulting in a posterior belief, i.e. belief update (Figure 2B). Before the very first observation, participants' belief in sigma, $\hat{\sigma}$, was assumed to be uniformly distributed across parameter space, i.e. possible sigma values in parameter space were predicted to occur with equal probability:

$$p(\hat{\sigma}) = U(1,140). \quad (5)$$

A likelihood function was then calculated that described the probability of the observation x given each possible sigma value:

$$p(x \mid \hat{\sigma}) = N(0, \sigma)$$

(6)

With Bayes rule, we derived a trial-by-trial posterior distribution that was proportional to the multiplication of a prior distribution and likelihood:

$$p(\hat{\sigma}|\,x) \propto p(x \mid \hat{\sigma}\,)\, p(\hat{\sigma})$$

(7)

where,

a. p ($\hat{\sigma}$) is the prior distribution.

b. p(x |$\hat{\sigma}$) is the likelihood function.

c. p($\hat{\sigma}$| x), is the posterior pdf across parameter space.

The posterior pdf is the updated belief across sigma space and is used as prior for the next trial of the same predictor. Finally, each posterior was normalised to ensure that probabilities across all sigma values added up to one:

$$p(\hat{\sigma} \mid x) = \frac{p(\hat{\sigma} \mid x)}{\sum p(\hat{\sigma} \mid x)}$$

(8)

2.4.1. Model parameters

The belief distribution allowed us to derive two independent model-based estimates that we hypothesized to influence choice in parallel: an estimate in the accuracy of a predictor (a point-estimate derived by the mode of the belief distribution, representing the sigma believed to be the most likely of that of the chosen predictor) and an estimate of the uncertainty in that predictor (95% interval around the accuracy estimate, representing the uncertainty) (note that the use of italics indicates reference to the model-derived parameters and not to the true,

experimenter-determined distribution properties; Figure 2A-iv). We derived these measures from an option's prior distribution on every trial. Note that both variables depended on choices made by participants, because feedback was only provided for the chosen option and hence only this one could be updated.

On trial i, variables of interest were defined as follows (Figure 2A-iv):

$$\text{accuracy} = \max\left[p\left(\hat{\sigma}\right)\right] * (-1) \quad (9)$$

Note that a higher max[p ($\hat{\sigma}$)] of the pdf indicated bigger deviations of the target from the reference point. To derive an accuracy estimate that can be interpreted intuitively, the sign of max[p($\hat{\sigma}$)] is reversed (multiplication with (-1)) so that positive values can be interpreted as higher accuracy. The accuracy estimate represents a point-estimate of a subject's belief distribution in sigma-hat ($\hat{\sigma}$). This means it represents the subject's belief in the sigma value associated with the predictors' distribution.

To derive a trial-wise uncertainty estimate from the distribution, we identified a percentage (2.5%) of the lower and upper tail of the prior pdf, representing the distribution around the believed sigma value ($\hat{\sigma}$). We extracted the estimated sigma value $\hat{\sigma}_{high}$ and $\hat{\sigma}_{low}$ at each of the two positions. The difference of both sigma values constituted the estimated uncertainty variable:

$$\hat{\sigma}_{high} \leftarrow \text{cumulative}\left(p\left(\hat{\sigma}\right)\right) = 97.5\%$$

$$\hat{\sigma}_{low} \leftarrow \text{cumulative}\left(p\left(\hat{\sigma}\right)\right) = 2.5\%$$

$$\text{uncertainty} = \hat{\sigma}_{high} - \hat{\sigma}_{low}$$

(10)

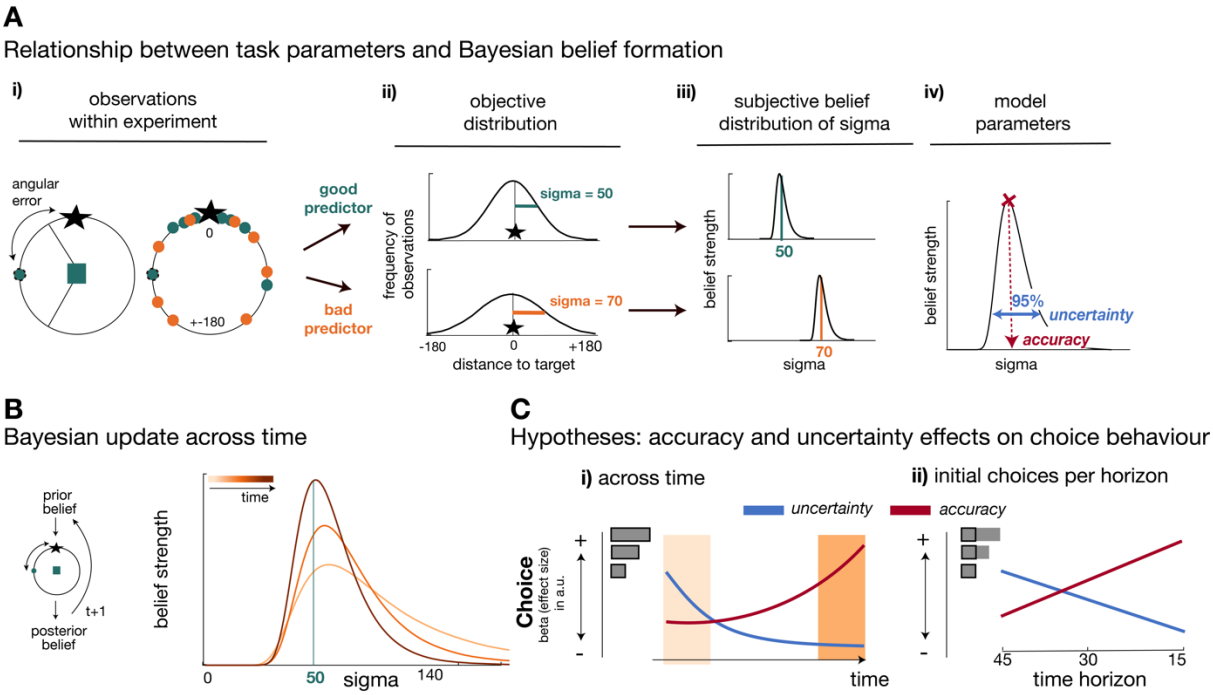


Figure 10. Task statistics, Bayesian model, and Choice hypotheses. (A) *Panels depict the mapping between observations during the task (i), their statistical properties (ii), and subjective beliefs about these properties derived with Bayes' rule (iii;iv).* ***(A-i)*** *A predictor's performance can be evaluated by the angular error at each trial (left panel), and by comparing angular errors between predictors across observations (right panel). Better predictors have on average smaller angular errors (green is better than orange).* ***(A-ii)*** *Predictors' angular errors were derived from normal distributions centred on the true target location. Critically, the normal distributions for good and bad predictors differed in their standard deviation (sigma): smaller sigma's reflected smaller angular errors, i.e. more accurate predictions of the true target location. Learning about a predictor's angular error across time corresponded to forming beliefs about a predictor's sigma value.* ***(A-iii)*** *To capture this learning process, we used Bayesian modelling and derived trial-wise belief distributions over sigma for each predictor. In other words, we estimated a probability density function that expressed the belief strength in each possible sigma over a large range of sigmas, and that was updated with each new observation via Bayes' rule. The coloured vertical lines indicate the true underlying sigmas of the predictors and the black distributions reflect the Bayesian approximation after extensive training.* ***(A-iv)*** *We captured two separable estimates about participants' beliefs concerning predictors: an estimate of the accuracy of a predictor (the mode of the distribution indicated by the position of the vertical line on the abscissa) and the uncertainty in that belief (width of the belief distribution).* ***(B)*** *In all panels, light to dark orange represents earlier and later trials, respectively, in a block. Left: Prior beliefs are updated after observing the angular error in the trial's outcome phase, resulting in a posterior belief. The posterior belief forms the prior for the next encounter with the same predictor. Right: Belief distribution when selecting the same predictor multiple times. Across time, the belief distribution will converge towards the true value of sigma (here, true sigma is 50).* ***(C)*** *Experimental hypotheses. Note that panels depict an illustration of hypothesized effect sizes of accuracy and uncertainty on choice akin to logistic GLM analyses of choice. (**C-i**) Participants' patterns of explore/exploit choices should systematically change over the course of the blocks. At the beginning of a block (light orange area), participants should pursue the more uncertain predictor, that is choices should be driven by a positive uncertainty effect, but this tendency should reverse over time. Accurate predictors should be sought out throughout (positive accuracy effect), but particularly towards the end of*

the block (dark orange area) when the value of exploration diminishes. ***(C-ii)*** *At the time of initial choices (indicated by black boxes in inset), the value of exploration should be modulated by the time horizon and choices towards uncertain predictors should systematically increase if there are more trials remaining in which to exploit the knowledge gained, i.e. in longer horizons (vice versa for accuracy-driven choices).*

2.5. Choice hypotheses

We measured the degree to which participants were exploiting accurate predictors and the degree to which they were exploring uncertain predictors. We hypothesized, first, that uncertainty drives exploration between choices at the beginning of a block and so choices might be directed to uncertain predictors. Then, over the course of a block, participants should become increasingly uncertainty avoiding; choices should be directed to certain predictors (Figure 2C-i). Second, we hypothesized that the initial choice pattern in a block should depend on how many more trials were still to be encountered in the block (effect of time horizon). Longer blocks favour more uncertainty-driven exploration and less accuracy-driven exploitation compared to shorter blocks (Figure 2C-ii).

2.6. Behavioural analyses

We applied a set of general linear models (GLM) to the behavioural data. All GLM analyses were applied to both versions (social and non-social) of the experiment separately. The resulting beta weights for each subject were first averaged across versions and then across participants. We used two-tailed statistical tests for all analyses.

2.6.1. Decision Phase

We analysed the trial-wise impact of Bayesian-derived estimates of accuracy, uncertainty, and their modulations across the time in a block on choice behaviour. Our first analysis aimed to show that the belief in the accuracy of a predictor (i.e. accuracy) and the uncertainty in that belief (i.e. uncertainty) influenced choice behaviour. Moreover, we focused on how these effects changed with the percentage of remaining trials in a block (% r. tr.), suggesting a transition between exploration and exploitation as time within a block passes. We used a logistic general linear model (GLM) to investigate these effects across all trials on choice behaviour (Choice GLM1). For all GLM analyses, regressors were normalised across all trials (mean of 0 and standard deviation of 1). The first GLM comprised the following regressors:

Choice GLM1 (Figure 4A):

accuracy,

uncertainty, % r. tr.,

accuracy x % r. tr.,

uncertainty x % r. tr. .

The dependent variable was whether or not participants made a leftward choice on the current trial. Accordingly, for each regressor (except % r. tr.), we calculated the difference in the variable for the left and right option. To calculate the interaction term, we multiplied the normalised uncertainty and accuracy variables with the normalised % r. tr. variable and then normalised this interaction term again. Note that we use the accuracy and uncertainty regressors as defined in the “Bayesian model” section.

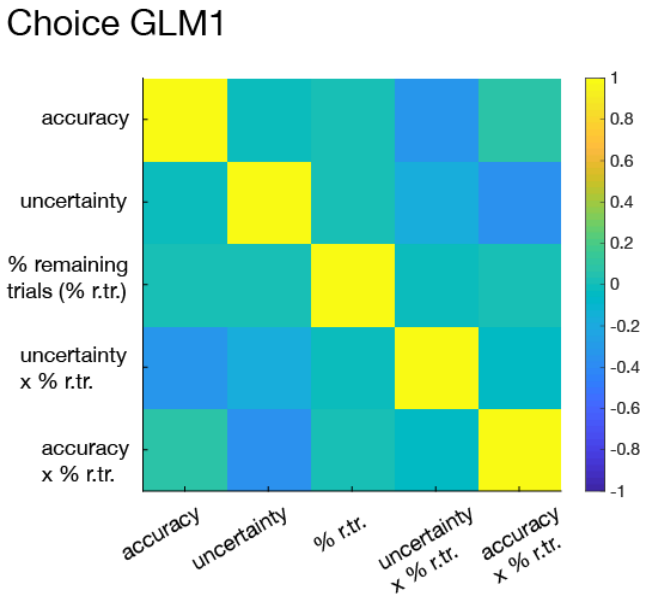


Figure 3: Correlation between variables included in main behavioural analysis. *Correlations represent an average across all participants: first, correlations were averaged for each subject across sessions and then averaged across subjects. All regressors were normalized before they were applied to the data. Interaction terms were treated identical to the behavioural analyses: they were normalized before and after multiplication before including them into the analysis.*

To further examine how the influence of uncertainty and accuracy on choice changed over time in a block, i.e. to examine the interaction terms of the previous Choice GLM1, we binned trials within a given time horizon into first and second halves (Figure 4B-i). We fitted a logistic GLM on each half with uncertainty and accuracy as regressors, irrespective of the overall time horizon length of the block:

Time GLM 1, for the first and second block halves (Figure 4B-i):

accuracy,

uncertainty.

Next, we predicted an effect of time horizon (Figure 2C-ii; 4B-ii) on the first trials of a block. The influence of accuracy and uncertainty on behaviour should scale with the number of trials still remaining in the current block. We fitted a robust linear GLM on the first 15 trials (a multiple of all horizons, which were 15, 30 and

45) with accuracy and uncertainty as regressors to investigate whether a variable's effect covaried with the amount of remaining trials.

Time GLM 2, for the first 15 trials within horizons (Figure 4B-ii):

accuracy,

uncertainty.

The first 15 trials were identical across horizons in terms of their predictor order and statistical properties (apart from the specific choice sequence taken by participants). We used a linear robust regression to better estimate effects given the small number of trials included in the analysis. Note that all significant results reported in Figure 4B-ii were also significant when basing the statistical tests on the t-stats of the effect sizes obtained from a logistic regression (reported interaction effect: 3x2 repeated measures ANOVA with horizon (long, medium, short) and variable (accuracy, uncertainty); horizon x variable interaction: $F(2,22)=27.6$, $p<0.001$; reported main effects: positive uncertainty during long horizon: $t(23)=4.7$, $p=1.0042e-04$; medium horizon: $t(23)=2.6$, $p=0.017$).

2.7. Behavioural results

2.7.1. Dissociable effects of uncertainty and accuracy on predictor selections across time

We first tested whether subjective beliefs in accuracy and uncertainty guide selections between predictors and whether their impact is modulated as a function of time. We applied a logistic general linear model (GLM, see GLM1 in Methods) to participants' selections during the decision phase and then averaged across participants (Choice GLM1, Figure 3;4A). Regressors of interest (accuracy and uncertainty) were coded as the difference between left and right predictors to predict leftward selections. As would be expected if participants were attempting to maximize payoff, participants generally sought out accurate predictors (main effect of accuracy: $t(23)=7.5$, $p<0.001$). Moreover, there was a trend that participants avoided uncertain predictors (main effect of uncertainty: $t(23)=-1.9$, $p= 0.066$). Next, to examine the time-dependent effect of uncertainty and accuracy on selection, we included the percentage of trials remaining in a block (% r. tr.) into the GLM model and examined its interaction with accuracy and uncertainty. Indeed, participants alternated between behavioural modes of exploration and exploitation by integrating information about the remaining trials into their predictor selections: a positive interaction term between uncertainty and % r. tr. ($t(23)=5.8$, $p<0.001$) showed that uncertain sources were explored when many trials remained. By contrast, a negative interaction term between accuracy and % r. tr indicated that, as time passed, choices were increasingly directed towards accurate predictors (accuracy x % r. tr interaction: $t(23)=7.5$, $p<0.001$; Figure 4A).

In a follow-up analysis, we further examined the interaction effects. We binned trials into those that occurred in the first and second halves of each time horizon (Figure 4B-i). Splitting trials in such a way allows characterizing the changes in uncertainty and accuracy effects on early versus late selections within a block. We fitted a logistic GLM with accuracy and uncertainty as regressors to both halves of each block's trials. Once again we found that decisions were influenced by both factors but in dynamically distinct ways (2x2 repeated measures ANOVA with block (half 1, half 2) and variable (accuracy, uncertainty); block x variable interaction: $F(1,23)=54.4$, $p < 0.001$, Figure 4B-i). Uncertain predictors were more likely to be sought out early compared to late in a block. In the second half of blocks, predictor selections were uncertainty-avoiding (negative uncertainty effect in half 2: $t(23)=-4.8$, $p = 7.3094e-05$). A negative uncertainty effect can be interpreted as a certainty effect; i.e., selections are directed towards certain predictors. For the sake of simplicity in further analyses, we will refer to negative uncertainty effects as certainty effects. Accurate predictors were preferred to inaccurate ones and this was increasingly the case in the second half of the blocks (paired t-test between early and late time points for each variable: uncertainty, $t(23)=5.9$, $p=4.17e-07$; accuracy: $t(23)=-3.2$, $p=0.003$).

Next, we tested our second hypothesis that the degree of exploration during initial choices should be stronger in longer time horizons, i.e. if subsequent encounters with the same predictor are expected to be more frequent. We compared choices during the first 15 trials across all time horizons by fitting a linear robust GLM to data from each time horizon. The first 15 trials in all three horizons were identical in their order presentation and importantly, their trial-by-trial target

estimates were drawn from a Gaussian distribution with the same parameters (sigma of either 50 or 70). As predicted, participants adjusted their behavioural strategy in the initial trials according to the horizon type: participants explored more in longer than shorter horizons and in a complementary manner, shorter horizons led to a rapid convergence onto accurate predictors (3x2 repeated measures ANOVA with horizon (long, medium, short) and variable (accuracy, uncertainty); horizon x variable interaction: $F(2,22)=34.4$, $p<0.001$; Figure 4B-ii). Uncertain predictors were particularly sought out during initial trials within long and medium time horizons (long horizon: $t(23)=4$, $p<0.001$; medium horizon: $t(23)=2.8$, $p=0.009$).

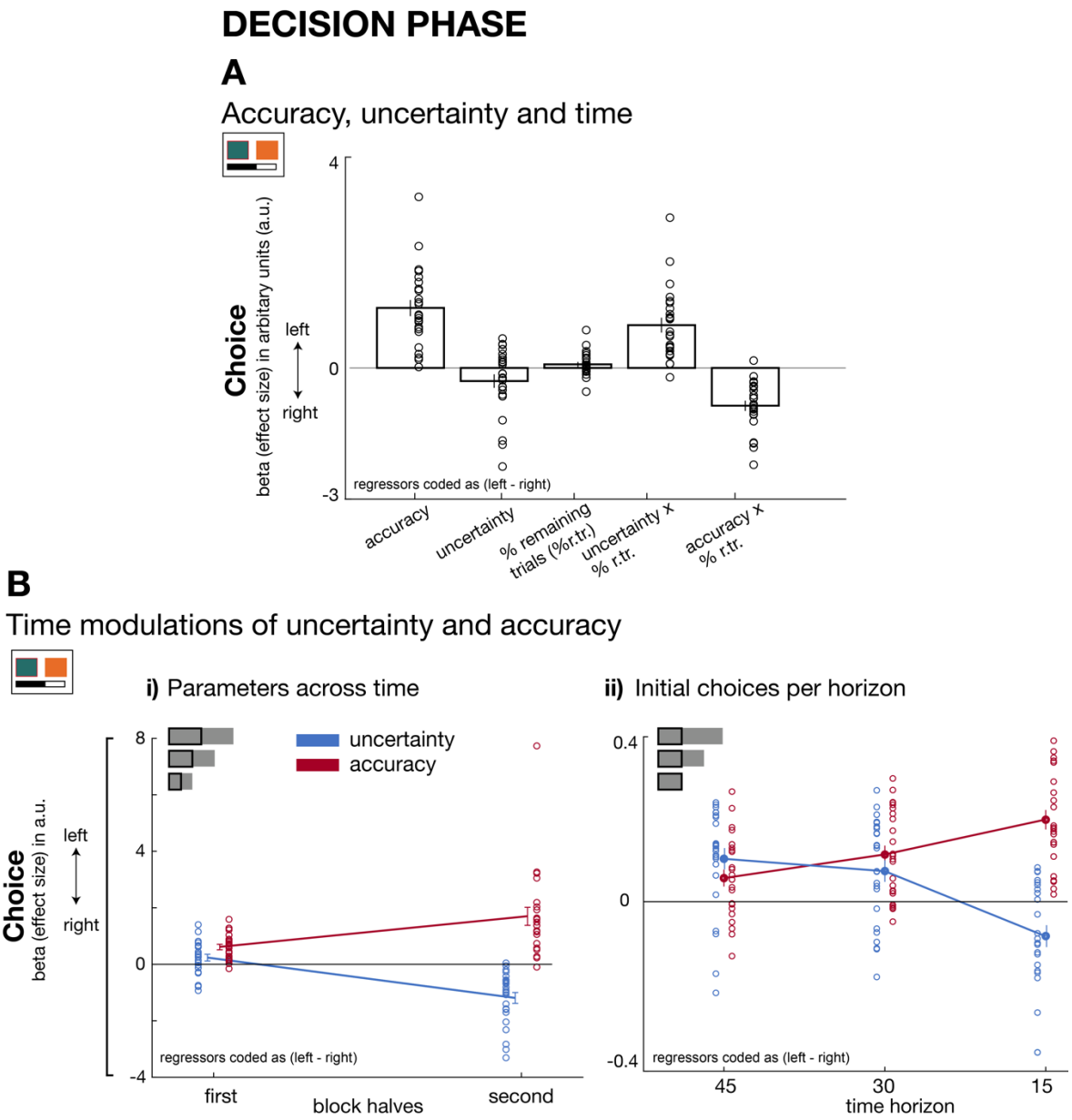


Figure 4: Dissociable effects of accuracy and uncertainty on predictor selections. *Decision phase.* ***(A)*** *Beta weights are derived from Choice-GLM 1. In general, participants preferred accurate predictors and showed a trend for avoiding uncertain ones. Uncertainty and accuracy exerted different effects depending on when choices were made: uncertain predictors were explored when many trials still remained (positive interaction term with percentage of remaining trials, % r.tr.), whereas decisions were accuracy-driven as the end of a block approached (negative interaction effect with % r.tr.).* ***(B-i)*** *Beta weights are derived from Time-GLM 1. Earlier choices (i.e. first block half) were more uncertainty-driven compared to choices towards the end of the block (i.e. second block half), then choices were uncertainty avoiding. In contrast, accuracy determined choices throughout both early and late block halves, but increasingly so in the second half.* ***(B-ii)*** *Beta weights are derived from Time-GLM 2. Accuracy and uncertainty effects on choice also varied as a function of how many trials still remained within a block: differences in the initial choice patterns (first 15 trials; see inset) across horizons showed that the exploration of uncertain predictors was more pronounced when horizons were*

longer while shorter horizons demanded more rapid exploitation of predictors estimated as most accurate. (n = 24; error bars are SEM across participants).

2.8. Discussion

Here we described how humans select in an adaptive manner between multiple information sources that enables them efficiently both to gather information about the predictors and to use that information to make choices. Using Bayesian modelling, we derived estimates of two kinds of beliefs that simultaneously influenced choice behaviour. To select between predictors, participants integrated beliefs about how accurately a predictor predicted the target (accuracy) and beliefs about the (un)certainty in that estimate (uncertainty). How much these beliefs influenced predictor selection depended on how many opportunities participants had had to learn about the predictors already (Wilson et al., 2014a). Behaviourally, participants initially gathered information about available predictors by selecting more uncertain predictors, while over time they converged towards accurate and certain predictors. However, importantly the influence of accuracy beliefs and their uncertainty depended on the future time horizon; participants explored uncertain predictors more during initial phases of a block when they knew that they had a longer time horizon remaining to exploit the knowledge gained.

Behaviour that initially appears uncertainty-directed and exploratory in nature may simply reflect noise in the decision process (Daw et al., 2006) or the learning process (Findling et al., 2019) but in the present study it is clear that behaviour is genuinely uncertainty seeking and exploratory in nature because it manifests to a

greater degree when the future time horizon is longer even when the decision context and past learning opportunities remain the same.

Our results provide a quantitative model capturing which features in the environment should be the focus of attention at any point in time as people progress from exploration to exploitation – things that you know are valuable or things that you don't know about and which might turn out to be valuable in the future. This has been a goal of learning theorists for many years(Mackintosh, 1975; Pearce and Hall, 1980) by studying attention during the learning of cue-outcome relationships. Two influential models have made opposite predictions: one model suggests that selective attention is driven by cues that are most predictive of reward (Mackintosh, 1975; Pearce and Hall, 1980), reminiscent of the accuracy-driven and certainty-driven predictor selections in the present study. The second model assumes that the uncertainty of a predictor is crucial for selective attention (Pearce and Hall, 1980). By using a Bayesian model to dissociate participants' beliefs about accuracy and uncertainty, we were able to show that in fact, both are important to determine whether a predictor will be selected to guide behaviour. Importantly, the magnitude of their influence on predictor selection depends on their relevance to the current context which varies across time.

3. CHAPTER: Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex

3.1. Introduction

The previous chapter examined how humans select between multiple predictors that predict an outcome of relevance such as reward. We showed that when selecting between predictors, participants integrated beliefs about how accurately a predictor predicted the target (accuracy) and beliefs about the (un)certainty in that estimate (uncertainty). How much these beliefs influenced predictor selection depended on how many opportunities participants had had to learn about the predictors already. In this chapter, I examine the neural correlates of accuracy and (un)certainty and their modulation during exploration and exploitation in the same experiment as described in Chapter 2.

Many previous studies have shown that ventromedial prefrontal cortex (vmPFC) activity reflects information relevant for making value-guided decisions between choices. When making a decision between choice options, vmPFC blood oxygen level dependent (BOLD) activity covaries with the decision variable that guides the decision – the difference in value between the choice taken as opposed to the choice rejected (Basten et al., 2010; Boorman et al., 2009; Chau et al., 2014; De Martino et al., 2012; FitzGerald et al., 2009; Fouragnan et al., 2019; Papageorgiou et al., 2017; Philiastides et al., 2010; Wunderlich et al., 2012). If, as

has been argued, such vmPFC activity changes reflect allocation of attention to a choice option (Hunt et al., 2018; Lim et al., 2011; Lopez-Persem et al., 2016), then it is possible that vmPFC activity also reflects selection of a predictor to guide behaviour and the reason why it is being selected to guide behaviour: either because of its predictive accuracy, because of the certainty of its prediction, or because of the uncertainty of its prediction.

Activity in vmPFC has also been associated with exploiting high-value options rather than exploring uncertain low-value options (i.e. exploratory choices) (Daw et al., 2006) or when staying with a current option rather than switching to an alternative option (Leong et al., 2017). In contrast, dorsal anterior cingulate cortex (dACC) has often been linked with more switches and changes in behaviour, for example during foraging (Kolling et al., 2012), exploratory behaviour (Daw et al., 2006), or when tracking the value of choices that are alternatives to the options currently being taken (Boorman et al., 2009). Although vmPFC has been associated with representing the decision variable that is relevant to the current behavioural mode, it remains unclear whether vmPFC is implicated during exploratory behaviour.

Here, we used fMRI in humans to investigate how subjective beliefs about accuracy and uncertainty are represented in the brain (Trudel et al., 2020). In particular, we investigated how these representations change according to the context of exploration and exploitation.

3.2. Methods

3.2.1. FMRI data acquisition and data processing

Imaging data were acquired with a Siemens Prisma 3T MRI using a multiband T2*-weighted echo planar imaging sequence with acceleration factor of two and a 32-channel head-coil. Slices were acquired with an oblique angle of 30 ◦ to the PC-AC line to reduce signal dropout in frontal pole. Other acquisition parameters included 2.4x2.4x2.4 mm voxel size, TE = 20 ms, TR = 1030 ms, 60◦ flip angle, a 240 mm field of view and 60 slices per volume. For each session, a fieldmap (2.4x2.4x2.4mm) was acquired to reduce spatial distortions. Bias correction was applied directly to the scan. A structural scan was obtained with slice thickness = 1 mm; TR = 1900 ms, TE = 3.97 ms and 1x1x1 mm voxel size.

Imaging data was analysed using FMRIB's Software Library (FSL) (Smith et al., 2004). Preprocessing stages included motion correction, correction for spatial distortion by applying the fieldmap, brain extraction, high-pass filtering and spatial smoothing using full-width half maximum of 5mm. Images were co-registered to each individual's high-resolution structural image and then nonlinearly registered to the MNI template using 12 degrees of freedom (Jenkinson and Smith, 2001).

3.2.2. Definition of exploration, exploitation and their transition

For some of the fMRI analyses, we analysed the neural data separately for different bins of trials. First, we subdivided trials into exploration and exploitation trials to compare neural signals between both behavioural modes. For each subject, we categorized trials based on the predictor selections during the decision

phase (Figure 1). On each trial, we calculated the difference between chosen and unchosen accuracy and chosen and unchosen uncertainty. Exploitative trials were defined by a positive accuracy prediction difference (chosen predictor had higher accuracy than unchosen ones) and negative uncertainty prediction differences (the chosen predictor was the predictor participants were more certain about) (approximately 52% of trials). Vice versa, exploration trials were defined by a negative accuracy prediction difference and positive uncertainty prediction differences (the more uncertain predictor is picked even though it has yielded less accurate results in the past) (approximately 18% of trials). Trials with both positive accuracy prediction difference and uncertainty prediction difference (i.e. that were both accuracy and uncertainty guided) were allocated to either the exploitative or the exploratory bin depending on the relative predominance of the accuracy prediction difference or the uncertainty prediction difference. For example, if the chosen predictor and the unchosen predictor differed more in the uncertainty of their predictions as opposed to the accuracy of their predictions (the chosen predictor was more uncertain than the unchosen predictor and the chosen predictor was, to a smaller degree, more accurate in its predictions than the unchosen predictor) then the predictor selection on that trial was labelled as exploratory. Finally, if the difference between the sizes of the decision variables was small (<5) then the trial was assigned to both exploration and exploitation bins (5% of trials). The remaining 25% of trials (white area in Figure 1-i) were not assigned to either exploitative or exploratory bins, because these choices were neither guided by uncertainty nor accuracy. As a manipulation check, we show the prediction differences for accuracy and uncertainty separated by exploration and

exploitation (Figure 1B). We find that indeed exploratory trials are characterized by a positive uncertainty prediction difference (the chosen predictor was associated with greater uncertainty than the unchosen predictor) while exploitative trials were defined by a positive accuracy prediction difference (the chosen predictor was associated with greater predictive accuracy than the unchosen predictor) and negative uncertainty, i.e. that is certainty prediction difference (the chosen predictor was associated with greater predictive certainty than the unchosen predictor).

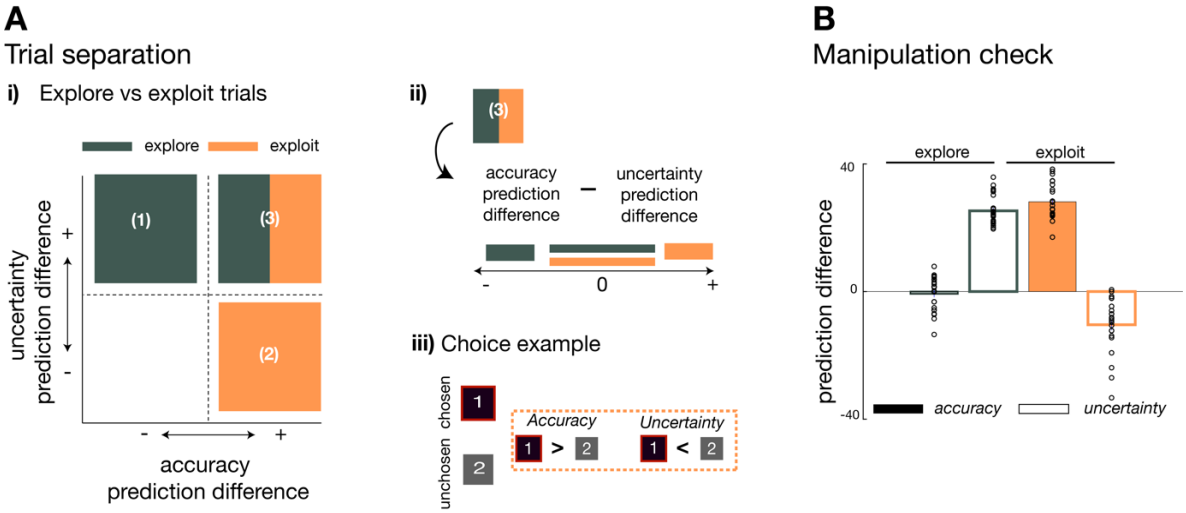


Figure 1: Trial separation into exploration/ exploitation according to individual choices. ***(A-i)*** *We classified trials into exploration and exploitation according to subjects' choices. To this end, we compared accuracy and uncertainty of the chosen and unchosen predictors, i.e. prediction difference. Explorative choices* ***(Ai-1)*** *were defined as those directed towards higher uncertainty (positive uncertainty prediction difference) and less accurate predictors (negative accuracy prediction difference). An exploitative trial (Ai-2) was defined by choices of predictors with a higher accuracy (positive accuracy prediction difference) and lower uncertainty (negative uncertainty prediction difference) than the unchosen predictor.* ***(Ai-3 & A-ii)*** *When participants chose predictors that were both more accurate and more uncertain, we compared the relative magnitudes of the accuracy prediction difference and the uncertainty prediction difference. If the difference in accuracy prediction difference was greater than the uncertainty prediction difference, then that trial was allocated to the exploitative bin and vice versa for explorative trials. If the difference between the sizes of the decision variables was small (<5) then the trial was assigned to both exploration and exploitation bins. The remaining trials (white area in panel i) were not assigned to either exploitative or exploratory bins.* ***(A-iii)*** *Example of an exploitative choice: the chosen predictor has a higher accuracy and a lower uncertainty prediction difference.* ***(B)*** *Manipulation check: Exploratory trials are characterized by a positive uncertainty prediction difference, while exploitative trials were defined by a positive accuracy prediction difference and negative uncertainty, i.e. that is certainty prediction difference.*

Furthermore, we defined a new subset of trials to understand the transition from uncertainty prediction difference signals (exploration) to a certainty prediction difference signal (exploitation) in vmPFC. Because predictor selections are not driven by (un)certainty alone, we tested whether accuracy prediction difference signals were particularly prominent in a transitional phase between exploration and exploitation in vmPFC. We defined a threshold in a range of accuracy prediction difference values between [5 20] that classified trials into the transition period. We chose this subset such that it would comprise trials that are close in accuracy values for both options and at the same time predictor selection would still be guided rationally by accuracy. Moreover, this window resulted in a sufficiently large sample for analysis (approximately 20% of the trials in the range of positive accuracy prediction difference). The threshold is arbitrary and slightly smaller or greater ranges (compromising positive values) did not alter the results. To show that the transition period was characterized by learning about predictors, and that periods outside this transition were defined by the processing of either uncertainty or certainty, we defined two separate subsets of trials (Figure 2). One subset included extreme positive accuracy-driven trials [accuracy values > 20] (Figure 2A-ii), while a second subset contained extreme negative accuracy-driven [accuracy values < -5] trials (Figure 2A-i).

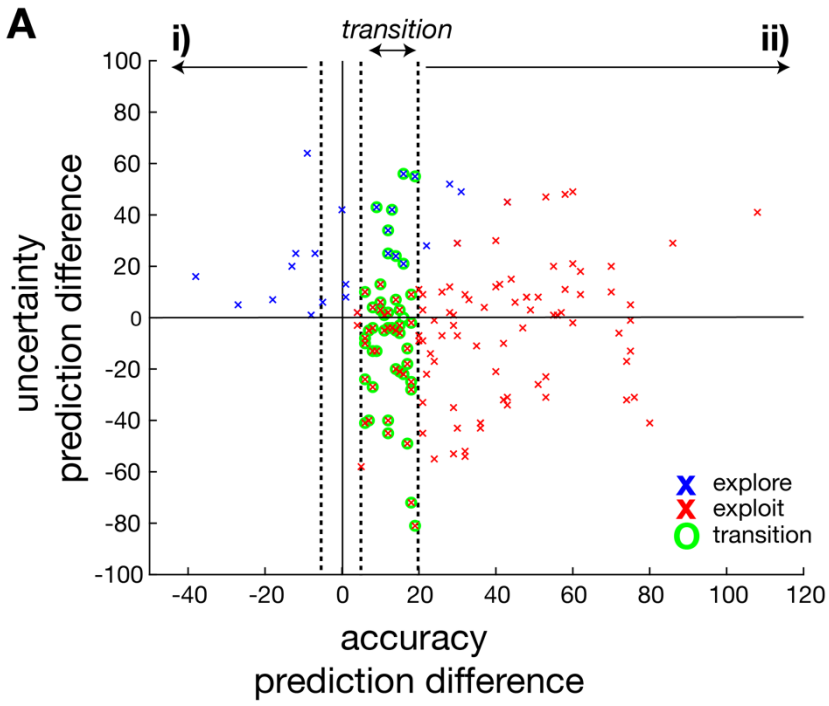


Figure 2: *Definition of* ***transition trials.*** *Example of one subject showing explore, exploit and transitional choices (one data point is one choice). Exploratory (blue crosses) and exploitative (red crosses) choices were categorised as described in Figure 1A. Transitional trials (green circles) were identified independently of the criteria used to identify explore/exploit trials; they were defined by a positive but small accuracy prediction difference in the range of [5 20], which resulted into approximately 20% of the trials (see methods for details).*

3.2.3. FMRI whole-brain analyses

We used FSL FEAT for first-level analysis (Smith et al., 2004). First, data was pre-whitened with FSL FILM to account for temporal autocorrelations. Temporal derivatives were included into the model. We used two GLMs to analyse fMRI data across the whole brain. fMRI-GLM1 was applied to all trials and fMRI-GLM2 was fitted separately to trials that had been identified as exploration and exploitation trials according to the previous section. Results were calculated using FSL's FLAME 1 with a cluster-correction threshold of $z>2.3$ and $p<0.05$, two-tailed.

To analyse BOLD changes associated with the processing of uncertainty and accuracy across participants, a second-level analysis was applied in a two-step approach: social and non-social versions were first averaged within subject (fixed-effect analysis) and then sessions were analysed across participants (FLAME1).

We included all three phases of a trial (decision, confidence and outcome) into the fMRI GLM. Each phase included a constant regressor, which was the onset of each phase as well as parametric regressors that were modelled as stick functions (i.e. duration of zero) time-locked to the relevant phase onset.

The decision phase began at the time the predictor appeared and lasted until a selection was made by the participant (Figure 1A). The decision phase was modelled as a constant and was accompanied by the following parametric regressors:

fMRI-GLM1, decision phase:

chosen uncertainty,

unchosen uncertainty,

chosen accuracy,

unchosen accuracy,

All regressors were normalised before inclusion into the analysis. We calculated the difference between chosen and unchosen predictors for both accuracy and uncertainty to derive prediction differences. To derive a domain general prediction difference, we calculated the contrast between both prediction differences of accuracy and uncertainty as follows: [(accuracy prediction difference)*1 + (uncertainty prediction difference)*(-1)] (Figure 5A) and calculated a conjunction between both cluster-corrected maps of accuracy and uncertainty

prediction differences with a cluster-correction of $z>2.3$ and $p<0.05$ (Figure 5B). For the conjunction analysis, we used the provided FSL script 'easythresh_conj' with $z>2.3$ and $p<0.05$.

The confidence phase was defined from the onset of circle and interval presentation (Figure 1A) until a decision about the interval size was made. It included a constant and the following parametric regressors:

fMRI-GLM1, confidence phase:

chosen uncertainty,

chosen accuracy,

% r. tr.,

chosen uncertainty x % r. tr.,

chosen accuracy x % r. tr.

All regressors were normalized before, and, where relevant, after building the interaction term (chosen accuracy/ uncertainty x % r. tr.). We only included the chosen predictor, as participants evaluated their uncertainty and accuracy estimates according to the predictor they have selected during the decision phase.

The outcome phase was defined by the onset of the target and payoff presentation and lasted for a fixed duration of three seconds. In addition to the constant regressor, we included the following parametric regressors:

fMRI-GLM1, outcome phase:

chosen accuracy,

chosen uncertainty,

payoff (as defined in equation 3).

In the second fMRI-GLM2, trials were binned into exploratory and exploitative trials as described above. For this purpose, we included decision, confidence and outcome phases for exploratory and exploitative trials separately. This meant that, in total, there were six phases within the fMRI-GLM2. We included the same set of regressors in the exploratory and exploitative phases. The constants for each phase was modelled as in the previous GLM, but we used separate constants for exploration and exploitation phases.

fMRI-GLM2, decision phase (for explore and exploit separately):

uncertainty prediction difference (i.e., chosen – unchosen)

accuracy prediction difference (i.e., chosen – unchosen)

fMRI-GLM2, confidence phase (for explore and exploit separately):

chosen accuracy

chosen uncertainty

fMRI-GLM2, outcome phase (for explore and exploit separately):

chosen accuracy

chosen uncertainty

payoff.

To test whether the uncertainty prediction difference significantly differed between exploration and exploitation, we built a contrast comparing uncertainty prediction differences between exploration and exploitation (Figure 9).

In addition, fMRI-GLM1 contained one regressor time-locked to all button presses, modelled as a stick function. For fMRI-GLM2, two regressors were time-locked to the button presses: one relating to the exploration phase and the other related to the exploitation phase.

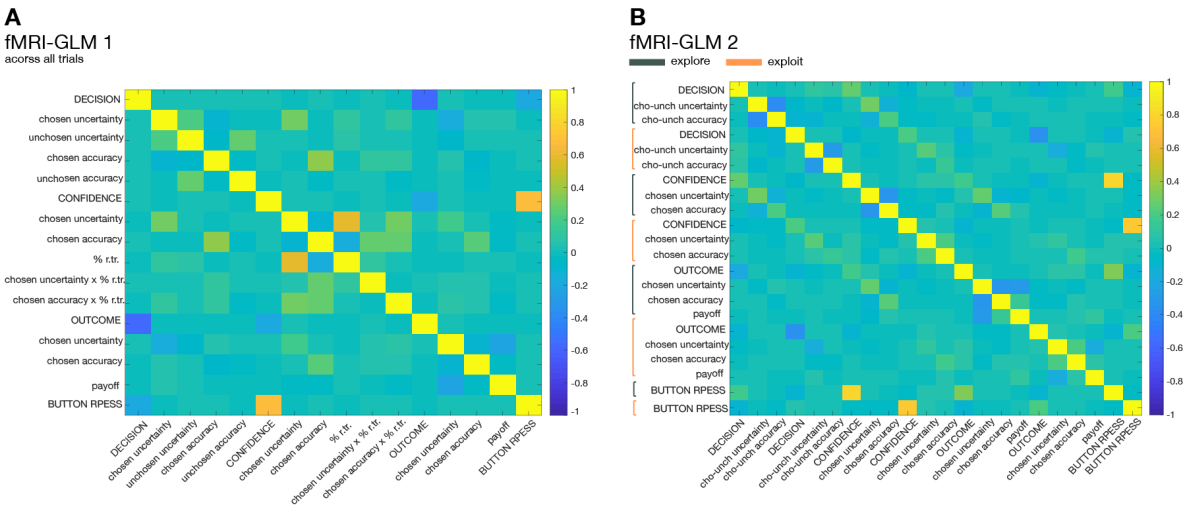


Figure 3: Correlation between variables included in neural analyses. *Correlations represent an average across all participants: first, correlations were averaged for each subject across sessions and then averaged across subjects. All regressors were normalized before they were applied to the data. Interaction terms were treated identical to the behavioural analyses: they were normalized before and after multiplication before including them into the analysis.* ***(A)*** *Regressors used in neural fMRI-GLM1 and* ***(B)*** *regressors used in neural fMRI-GLM2.*

3.2.4. FMRI region of interest analyses

We calculated ROIs with a radius of three voxels that were centred on the peak voxel of significant clusters derived from whole brain fMRI-GLM1 and fMRI-GLM2. The selected ROI was transformed from MNI space to subject space and the pre-processed BOLD time courses were extracted for each participant's session. Time courses were averaged across volumes, then normalized and oversampled by a factor of 20 for visualisation. Time courses were time-locked to the onsets of each phase consistent with timings used in whole-brain fMRI-GLMs (decision, confidence or outcome). Then, a GLM was applied to each timepoint to derive beta weights per time point for each regressor. For analyses across conditions, we used the same principle as applied to the whole-brain fMRI-GLMs and our behavioural analyses: first, we averaged the time course within a subject

across both social and non-social conditions, then we averaged across the group. For all ROI analyses, regressors were normalized (mean of zero and standard deviation of one).

We used a leave-one-out procedure to test the unbiased significance of the time courses extracted from all ROI analyses. For every participant (n = 24), we extracted the average time course based on the 23 remaining participants. We identified the peak of the group time course in a time window between 4-8 seconds. We extracted the beta value for the excluded subject at the time of the group peak. We repeated this procedure for all participants which resulted in individual peak values that were independent from the subject to be analysed. The extracted peak values were tested with a one-sample t-test against zero.

To illustrate uncertainty and certainty in exploration and exploitation phases, respectively, we included the following regressors:

ROI GLM1, decision phase (for explore and exploit separately), Figure 6C:

chosen uncertainty,

unchosen uncertainty,

chosen accuracy,

unchosen accuracy.

Effects of ROI-GLM1 was extracted from the whole-brain cluster corrected accuracy prediction difference effect in vmPFC to allow for an unbiased test.

For other analyses, we illustrate the effects of accuracy and uncertainty prediction differences during exploration, exploitation and random exploration. In these cases, we used the following regressors:

ROI GLM2, decision phase (for exploration, exploitation, random exploration), Figures 7,20,11,15:

Uncertainty prediction difference,

Accuracy prediction difference.

If the focus was on the neural representation of uncertainty, we used a ROI that was extracted according to the peak-coordinate of the cluster-corrected accuracy prediction difference (Figure 4B) and we used the opposite ROI centred on the cluster-corrected effect of uncertainty prediction difference (Figure 4A) when the focus was on the accuracy prediction difference. We obtained both clusters (accuracy prediction difference cluster and uncertainty prediction difference cluster) when applying the GLM across all trials. In this situation accuracy and uncertainty prediction differences had a very small correlation of r=0.1 (1% shared variance; manuscript Figure 4D, left panel). This degree of correlation cannot bias the assignment of variance to the respective variable. Also, because the two regressors competed for variance, the derived clusters represented variance that is uniquely explained by the associated regressor Thus, when examining either effect we could use the alternate cluster as an ROI in which to perform significance tests. Importantly, we did not select the ROI based on effects that we would then test within the ROI. ROI selection was therefore statistically independent of the effects of interest. Moreover, we used each cluster during analyses applied to a different set of trials (exploration versus exploitation) and we are not repeating the test across all trials.

Next, we tested whether the certainty effect changed when repeating the same predictor as on the last encounter. We used a ROI analysis to test for a main

effect of repetition and interaction effect between repetition and chosen certainty. We used ROI-GLM1 and additionally included the following regressors:

ROI GLM3, decision phase (across all trials), Figure 13:

additional regressors to ROI-GLM1:

repetition (1= repetition of the same predictor as during last encounter with same predictor; 0=no repetition of the same predictor)

repetition x chosen certainty,

repetition x chosen accuracy.

Then, we split trials into repetition and no-repetition categories to investigate the simple effect of chosen certainty per category (ROI-GLM4). We used ROI-GLM1, but now applied separately to repetition and no-repetition trials (Figure 13). For both ROI-GLM3 and 4, we used an unbiased ROI extracted from the whole-brain cluster corrected accuracy prediction difference effect across all trials in vmPFC.

Next, we applied a ROI analysis to show activation for accuracy prediction difference during the transitional phase (Figure 14) in vmPFC, using ROI-GLM2. We were interested whether the accuracy prediction difference effect occurred in the transition between the previously observed uncertainty and certainty prediction differences. Because we hypothesized that the accuracy prediction difference effect would occur in the same ROI as the uncertainty prediction difference effects, we used an independent ROI based on the cluster-corrected accuracy prediction difference effect across all trials (fMRI-GLM 1). We used the same ROI-GLM2 to test extreme positive and negative accuracy-driven trials (Figure 15).

3.2.5. Correlation between behavioural and neural beta weights

To calculate the correlation between the time course of neural activations and behavioural beta values, we used neural beta weights extracted from the group peak. We calculated a partial correlation between the vmPFC accuracy prediction difference effect during the transition phase and the behavioural interaction term of uncertainty x % r.tr. (Figure 14C), controlling for all other behavioural variables (main effects of accuracy, uncertainty, %r.tr. (in percentage) and the interaction between remaining trials and accuracy, see behavioural GLM1). A second partial correlation additionally included the number of individual transition trials.

3.3. Results

3.3.1. VmPFC represents multiple subjective beliefs

Our behavioural analyses show that participants relied on estimates of belief uncertainty when making a selection between two predictors. We went on to examine the coding of uncertainty in the brain during predictor selection (fMRI-GLM1). Our variable of interest was the difference in uncertainty (as captured by our model) between the chosen and unchosen predictors, i.e. uncertainty prediction difference. This is similar to studies of value-guided decision-making, where the difference in value between the option chosen and the option rejected is regressed against the BOLD signal. A value difference signal often prominently implicates the vmPFC in decision making processes (Basten et al., 2010; Boorman et al., 2009; Chau et al., 2014; De Martino et al., 2012; FitzGerald et al., 2009; Philiastides et al., 2010).

When testing for an uncertainty prediction difference signal across all trials, we found a negative effect in vmPFC. This negative effect can be interpreted as a certainty prediction difference effect (whole brain cluster-corrected; Figure 4A-i, Table1). This neural effect was in line with participants' general behavioural preference, particularly towards the end of a block, to choose certain predictors. In addition, we also found an accuracy prediction difference in a similar anatomical location in vmPFC (Figure 4A-ii, Table1). Again, this accords with participants' general preference for selecting accurate predictors to help them find the target location.

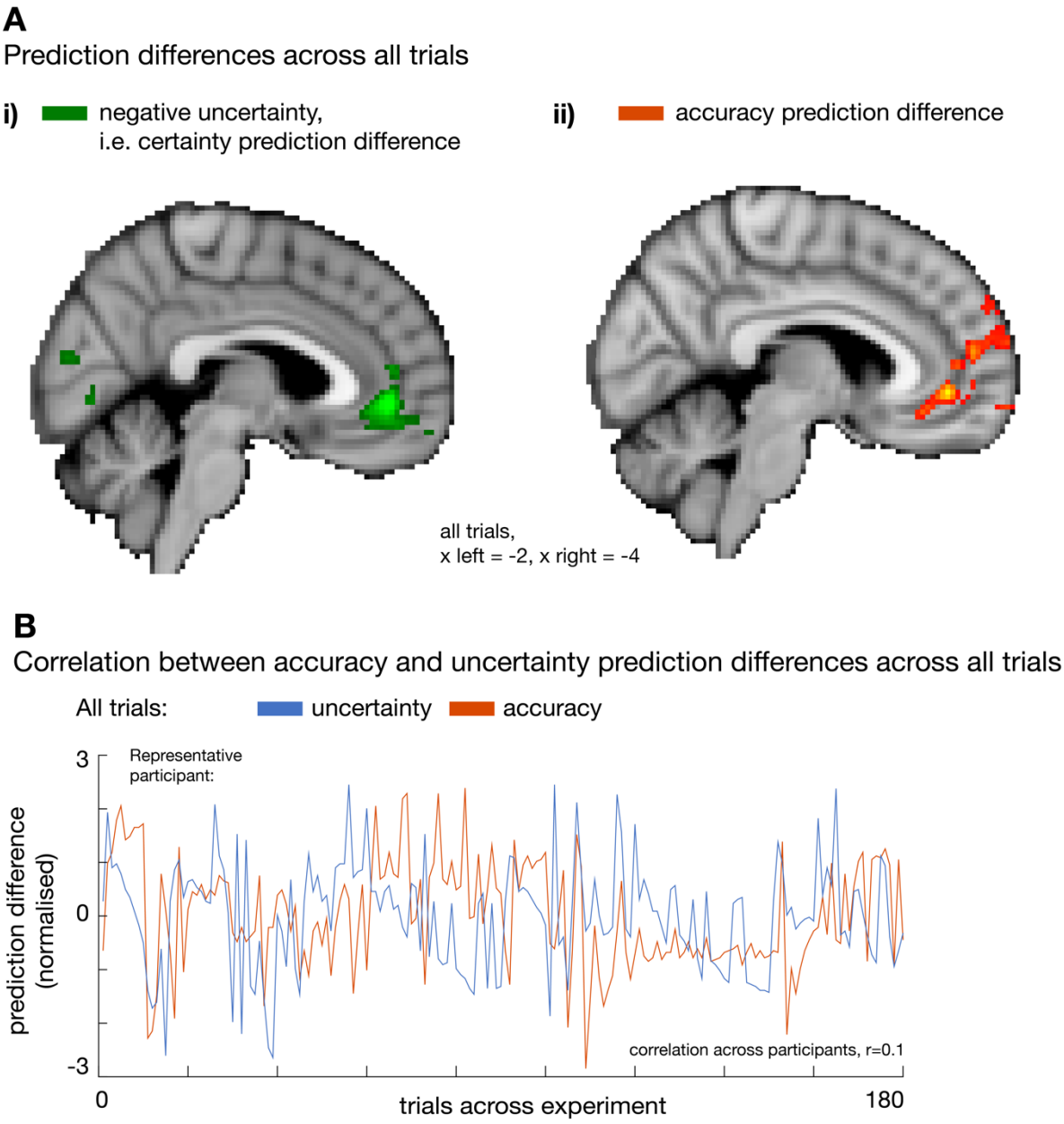


Figure 4: Accuracy and uncertainty prediction differences across all trials. *(**A**) Across all trials, a negative effect of uncertainty (i) and a positive effect of accuracy (ii) prediction differences covaried with activation in vmPFC. The variables were taken from fMRI-GLM1. Whole-brain effects were family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$. Z-thresholds indicated in figure panels are used for visualisation only. (**B**) The relationship between accuracy and uncertainty prediction differences used for all neural analyses across all trials. Average correlations between accuracy and uncertainty prediction differences across all participants are reported at the bottom of each panel, while panels show variables across time taken from a representative subject for each analysis. Accuracy and uncertainty prediction differences variable are similarly decorrelated in all other analyses. See Figure 3 for more information on correlations between variables included within each analysis type.*

To additionally show that both accuracy and uncertainty prediction differences were encoded in a similar anatomical region, we derived a domain general prediction difference by first, calculating the contrast between both prediction differences of accuracy and uncertainty as follows: [(accuracy prediction difference)*1 + (uncertainty prediction difference)*(-1)] and second, by deriving a conjunction between the same contrast (Figure 5A-i, 5A-ii, respectively, and Table3). A domain general prediction difference peaked within vmPFC. Note that accuracy and uncertainty prediction differences are independent variables sharing across all trials, on average, 0.01% of their variance (0.137% and 0.09 % of their variance is shared when exploration and exploitation trials are each considered separately; respectively Figure 4B, Figure 6B), suggesting both variables have independent effects on activity but within the same part of vmPFC. These findings underline the role of vmPFC in guiding predictor selection as a function of both the differences in accuracy and uncertainty of the predictors.

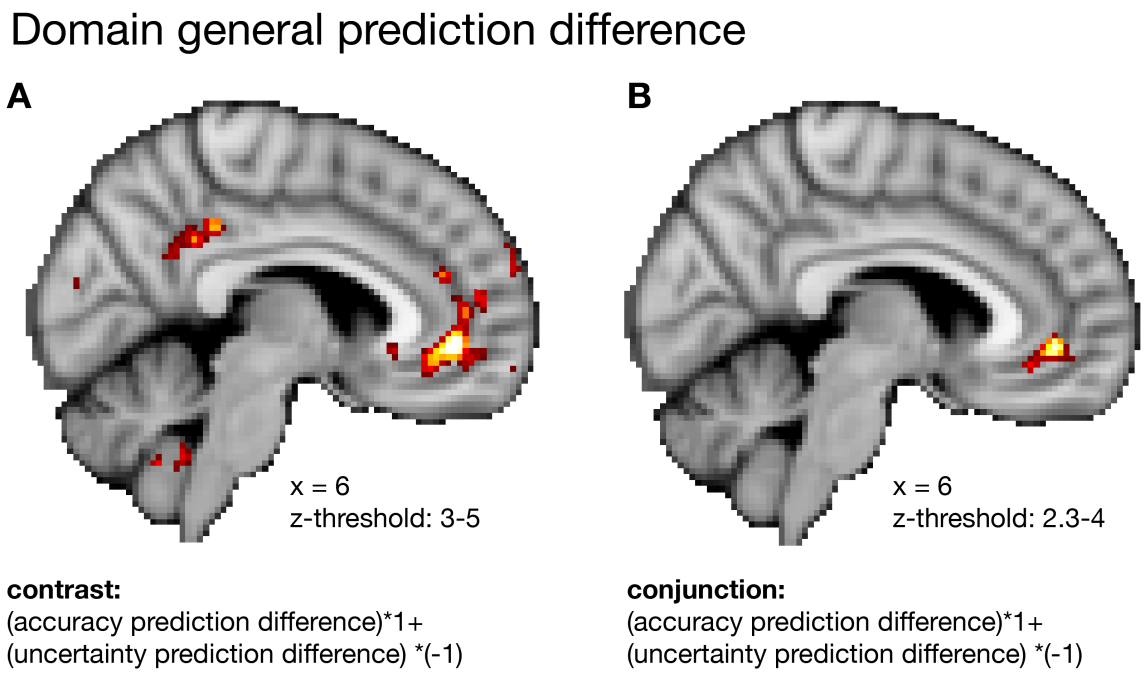


Figure 5: Domain general prediction difference across all trials. *A "domain general prediction difference" correlated with activation in vmPFC. (A) The effect was tested, firstly, by calculating the contrast between prediction differences of accuracy (Figure 4A-i) and uncertainty (Figure 4A-ii) as indicated in the panel. (B) Secondly, we tested the effect as conjunction of shared variance between these two absolute contrasts. See Methods section for more details. (n=24; whole-brain effects were family-wise error cluster corrected with z > 2.3 and p < 0.05; z-thresholds indicated in figure panels are used for visualisation only.)*

3.3.2. Polarity of uncertainty decision signal in vmPFC changes with behavioural mode

Having identified vmPFC as representing the certainty prediction difference across all trials, we then went on to test whether this signal was modulated by distinct behavioural modes of exploration and exploitation. We have shown that uncertainty tended to drive exploration of predictors at the beginnings of blocks; at that time, selections were directed to uncertain predictors (i.e. there was a positive effect of uncertainty during the first 15 trials in medium and long horizons, Chapter 2: Figure 4B-ii). Then, over the course of the block, participants became increasingly uncertainty avoiding and selections were directed to certain predictors.

We refer to this pattern of change as a “uncertainty polarity change”. We investigated whether there was a brain region with similar characteristics: transitioning from encoding an uncertainty-based to a certainty-based prediction difference as participants switched from exploration to exploitation.

In line with the behavioural results, we expected to find a positive uncertainty prediction difference signal during exploration but a positive certainty prediction difference during exploitation (Figure 6A). To test this hypothesis, we made use of the fact that our computational model allowed us to classify individual trials into exploration or exploitation according to the selection made on each trial: an exploitative selection was defined as one in which the more accurate and certain predictor was selected while a directed uncertainty-guided explorative selection was defined as the opposite: a trial in which the more inaccurate and uncertain predictor was chosen (Figure 1). Importantly this is distinct to other types of decision that might initially appear exploratory because the less accurate predictor was chosen (Daw et al., 2006). However choosing the less accurate may simply be due to noise in the learning or decision process (Daw et al., 2006; Findling et al., 2019). On such trials, selection is not just of the less accurate predictor but of the more certain predictor (Figure 7A).

To test for a neural polarity change from uncertainty prediction difference to certainty prediction difference, we extracted an independent region of interest (ROI) from the accuracy prediction difference effect across all trials. This ensured that we did not bias the analysis towards finding an effect in an area that was previously associated with the uncertainty prediction difference. First, we used a time course analysis to extract both components of the uncertainty prediction

difference signal (variance in activity related to the chosen predictor and variance in activity related to the unchosen predictor) during exploration and exploitation. Activation in vmPFC covaried with a decision signal that changed its polarity depending on the current behavioural mode: during exploitation, vmPFC carried a decision signal that reflected the certainty prediction difference (the difference between the chosen and unchosen predictor certainties; Figure 6C-ii); in exploration, when behaviour was guided by uncertainty, vmPFC activity carried an uncertainty prediction difference (the difference between chosen and unchosen uncertainty; Figure 6C-i). Note again that both prediction difference signals relate to the same variable (i.e. uncertainty) but depending on the behavioural mode, activation in vmPFC was either positively or negatively related to uncertainty.

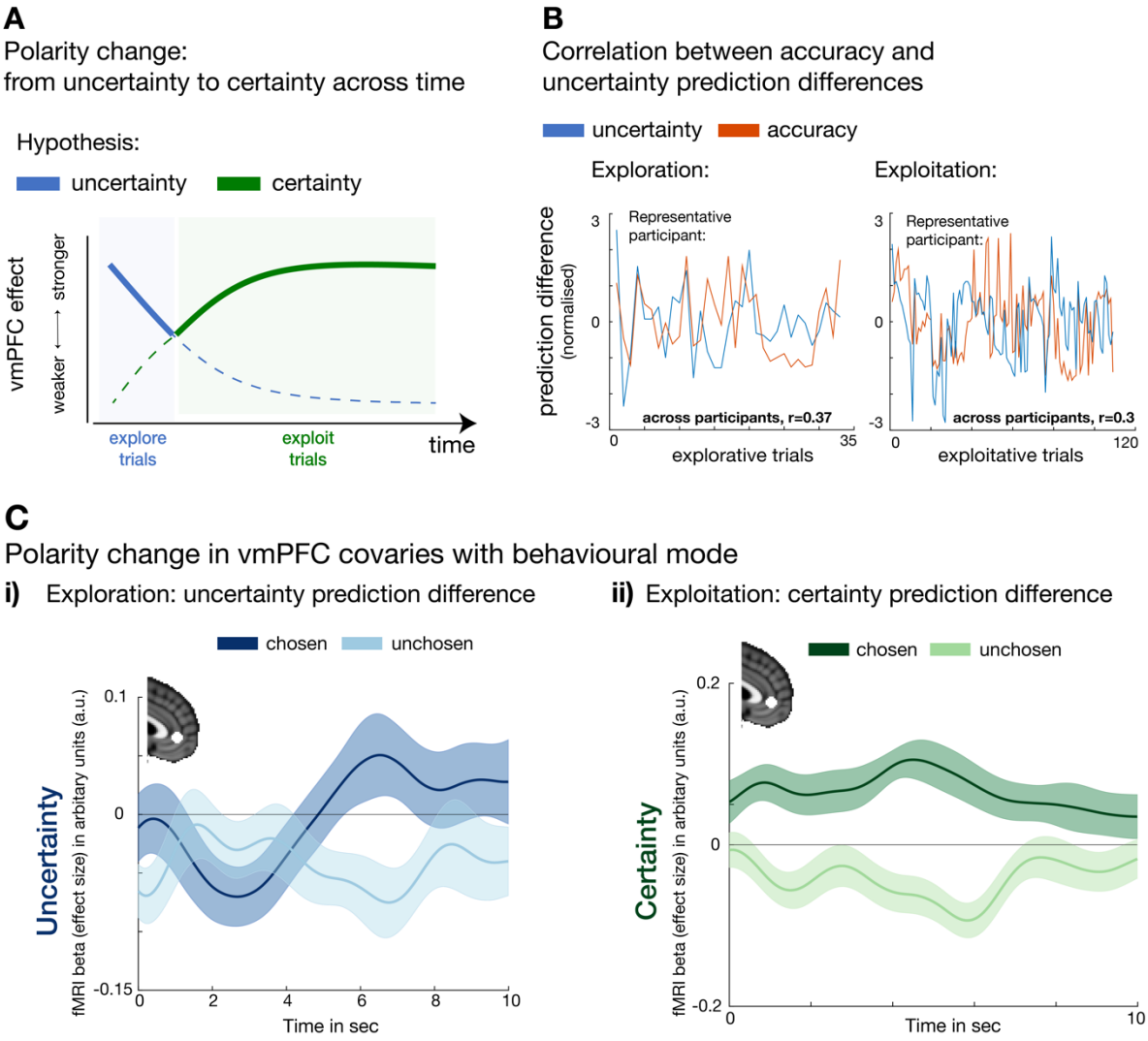


Figure 6: Polarity of uncertainty prediction difference in vmPFC changes with behavioural mode. ***(A)*** *We found a polarity change in the impact uncertainty exerted on predictor selection at a behavioural level; initial trials in longer horizons were more likely to be explorative and directed towards more uncertain predictors while behaviour in later trials was more exploitative and directed towards more certain predictors. We tested for a neural uncertainty polarity change in vmPFC comparing behavioural modes of exploration and exploitation. (**B**) The relationship between accuracy and uncertainty prediction differences used for exploration trials (left), and exploitation trials (right). Average correlations between accuracy and uncertainty prediction differences across all participants are reported at the bottom of each panel, while panels show variables across time taken from a representative subject for each analysis. Accuracy and uncertainty prediction differences variable are similarly decorrelated in all other analyses. (**C**) Time courses extracted from vmPFC for both chosen and unchosen components of an (un)certainty prediction difference signal during exploration (i) and exploitation (ii). VmPFC BOLD activity changed in accordance with the behavioural results; it transitioned from activity positively related to uncertainty prediction difference (positively encoding the uncertainty of the chosen predictor as opposed to the unchosen predictor) during initial choices to activity positively related to certainty prediction difference (negatively encoding the uncertainty of the chosen predictor as opposed to the unchosen predictor) in later trials. All effects were time-locked to the decision phase. (n = 24; error bars are SEM across participants; whole-brain effects family-wise error cluster corrected with z > 2.3 and p < 0.05).*

The trials we define as uncertainty-guided exploration trials are comparable to trials that have previously been described as directed explorative choices (Wilson et al., 2014a). They are, however, hypothesized to be distinct to apparently random choice selections that may result simply from noise in the decision process (Daw et al., 2006) or the learning process (Findling et al., 2019). In addition to examining vmPFC activity during uncertainty-guided exploration and exploitation, we therefore also tested whether vmPFC represented the (un)certainty prediction difference on random exploration trials. Random exploration trials are ones on which participants selected predictors that they were estimated to believe were inaccurate with certainty (Figure 7A). Unlike on uncertainty-guided exploratory trials, vmPFC did not represent (un)certainty prediction difference during these random exploratory selections (Figure 7B; leave-one-out test, $t(23) = -0.47$, $p= 0.64$), showing its selectivity to behavioural modes during which (un)certainty guides predictor selections: certainty prediction difference during exploitation and uncertainty prediction difference during exploration.

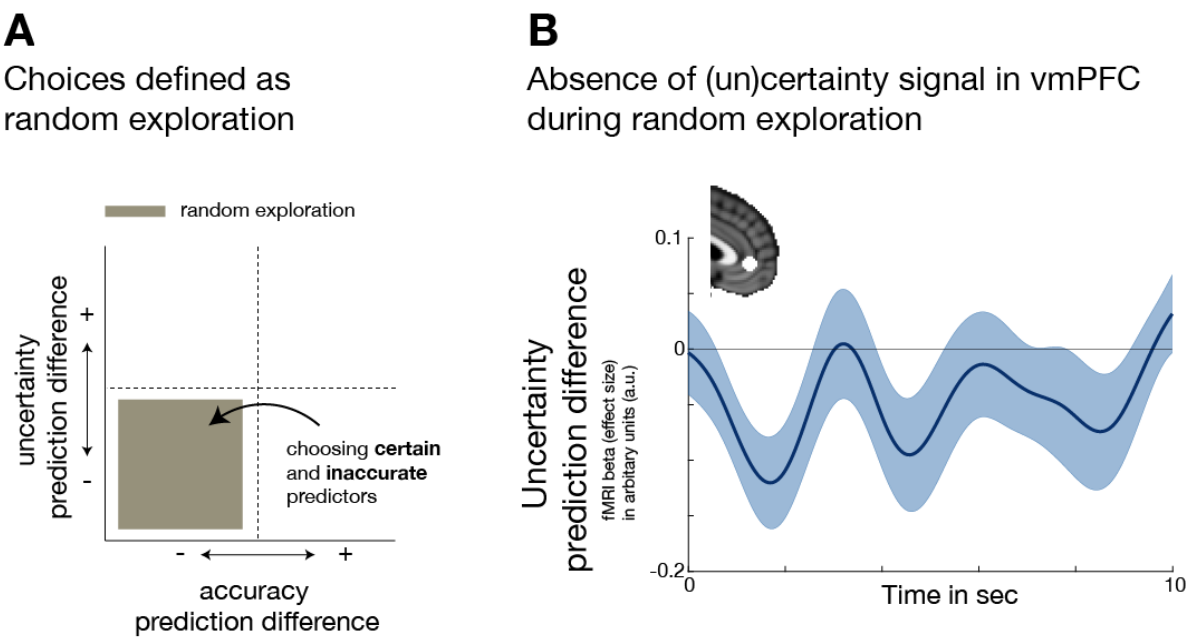


Figure 7: Absence of uncertainty prediction difference during random exploration in vmPFC. We showed that vmPFC represents uncertainty prediction difference during directed exploration (Figure 6C-i). Here, we show that this neural signature is unique to directed exploration and is absent during random explorative selections between predictors**. (A)** Selections during random explorative phases were defined as being directed towards inaccurate (negative accuracy prediction difference) and certain predictors (negative uncertainty prediction difference**). (B)** We show an absence of (un)certainty prediction differences during random exploration in vmPFC.

We have shown that vmPFC represents the uncertainty between predictors with opposing signs during exploration (positive uncertainty prediction difference) and exploitation (positive certainty prediction difference). However, it is possible to construct an alternative account of the observed sustained uncertainty activation across behavioural modes, according to which it reflects an absolute uncertainty prediction difference within both modes. To test for this alternative, we included the absolute uncertainty prediction difference as an additional regressor into the previously used ROI-analysis (fMRI-GLM2). During exploration, both the absolute uncertainty and uncertainty prediction differences are identical as in both cases, trials are defined by a positive uncertainty prediction difference, i.e. choosing the

more uncertain predictor. Therefore, we analysed the exploitation phase to dissociate between the two alternative explanations. Trials within exploitation are defined by both, positive and negative uncertainty prediction differences and therefore the polarity switch should not be affected by the additional regressor of an absolute uncertainty prediction difference. We re-ran the analysis but additionally included the absolute uncertainty prediction difference as an additional regressor. Nevertheless, we replicated the uncertainty polarity change across modes in vmPFC (t(23) = 4.7, p= 9.8814e-05; Figure 8).

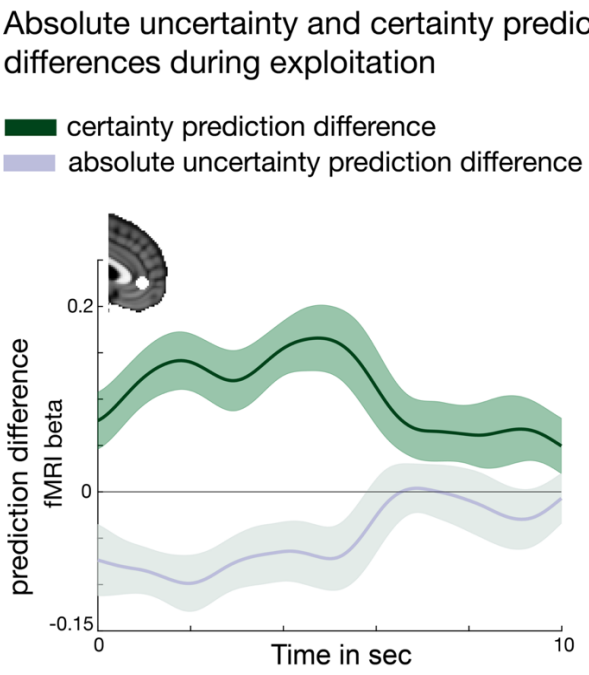


Figure 8: Polarity change of uncertainty between behavioural modes cannot be explained by an absolute uncertainty prediction difference. *We can replicate the certainty prediction difference during exploitation when additionally including the absolute uncertainty prediction difference during exploitative choices.*

We have shown that behavioural modes were associated with different polarities of uncertainty representation in vmPFC. Next, we were interested in

whether the different behavioural modes were associated with any distinct neural networks. We performed a whole-brain GLM of exploration and exploitation trials and focused again on the certainty/uncertainty prediction difference during the decision phase (fMRI-GLM2). During exploitation, we observed activity centred on vmPFC related to certainty prediction difference (Figure 9A; Table 2), confirming our previous findings. During exploration, an uncertainty prediction difference signal (difference between the uncertainty associated with chosen and unchosen predictors) was represented in vmPFC, but also across an extensive network of brain regions, including dorsomedial frontal areas (Figure 9B). A direct contrast of activation patterns in exploration and exploitation trials confirmed these differences between behavioural modes (compare panels A, B, and C of Figure 9).

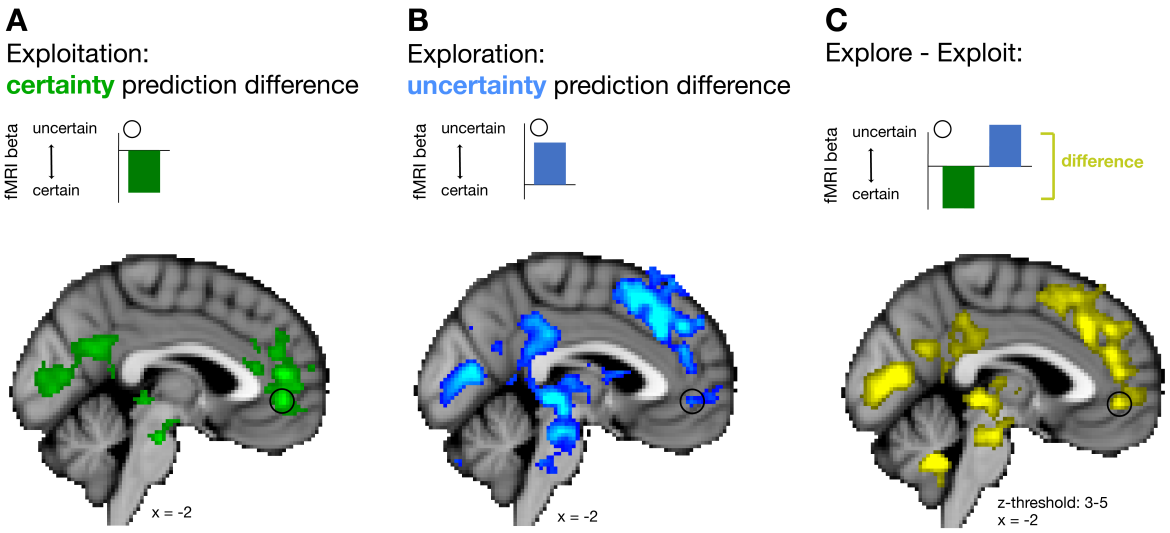


Figure 9. Whole brain maps for (un)certainty prediction difference during exploration and exploitation. *Illustrations above whole-brain images clarify the polarity (positive or negative) of the uncertainty prediction difference signal represented in vmPFC (indicated by the black circle) during exploitation, exploration and their contrast.* ***(A)*** *During exploitation, activity related to certainty prediction difference was restricted to a region centred on vmPFC (i.e. negative uncertainty prediction difference; see inset).* ***(B)*** *However, during exploration uncertainty prediction difference was associated with an extended network including vmPFC but also dorsomedial frontal areas peaking in dorsal anterior cingulate cortex (dACC) (see also Figure 10).* ***(C)*** *Difference in uncertainty prediction difference between exploration and*

exploitation. Contrasting activations between the behavioural modes of exploration and exploitation confirmed the presence of mode-specific (e.g. dACC) and mode-general (e.g. vmPFC) activations. Note that the sign of activation patterns resulting from a contrast between exploration and exploitation need to be interpreted with reference to the levels of activity found in the exploration and exploitation phases with respect to baseline (see illustration above each whole-brain map). (n = 24; whole-brain effects family-wise error cluster corrected with z > 2.3 and p < 0.05).

Dorsal ACC (dACC) in particular has been associated with exploratory (Daw et al., 2006) and foraging behaviour (Kolling et al., 2012). In line with previous findings, we show that dACC represents an uncertainty prediction differences during directed exploration (Figure 10A). Additionally, we show that dACC exhibits a distinctive neural activity pattern by representing both features of the chosen predictor as opposed to the unchosen predictor, that defined the exploration period: a positive uncertainty and negative accuracy prediction difference (Figure 10A-i).

We used ROI analyses to test for an uncertainty polarity change in dACC. ROIs were extracted according to cluster-corrected regions that were close to previous reported areas and associated with the uncertainty prediction difference during exploration. In our study, our contrast of uncertainty prediction difference during the exploration decision phase highlights similar regions in dACC as previous studies (Kolling et al., 2012) (A-iii: MNI x/y/z-peak coordinates: [0, 28,30])). dACC did not show a polarity change in the representation of uncertainty when comparing exploration to exploitation, as there was no representation of a certainty prediction difference during exploitation (Figure 10B-i: dACC: $t(23) = 0.19$, $p= 0.85$). Moreover, we show that during random exploration, dACC did not represent uncertainty prediction difference either (Figure 10B-ii: leave-one-out test, $t(23) = -1.1$, $p= 0.28$). dACC represents (un)certainty prediction differences exclusively

during directed exploration, but lacks such a representation during random exploration (Figure 9B-i) or, unlike vmPFC, exploitation (Figure 10B-ii).

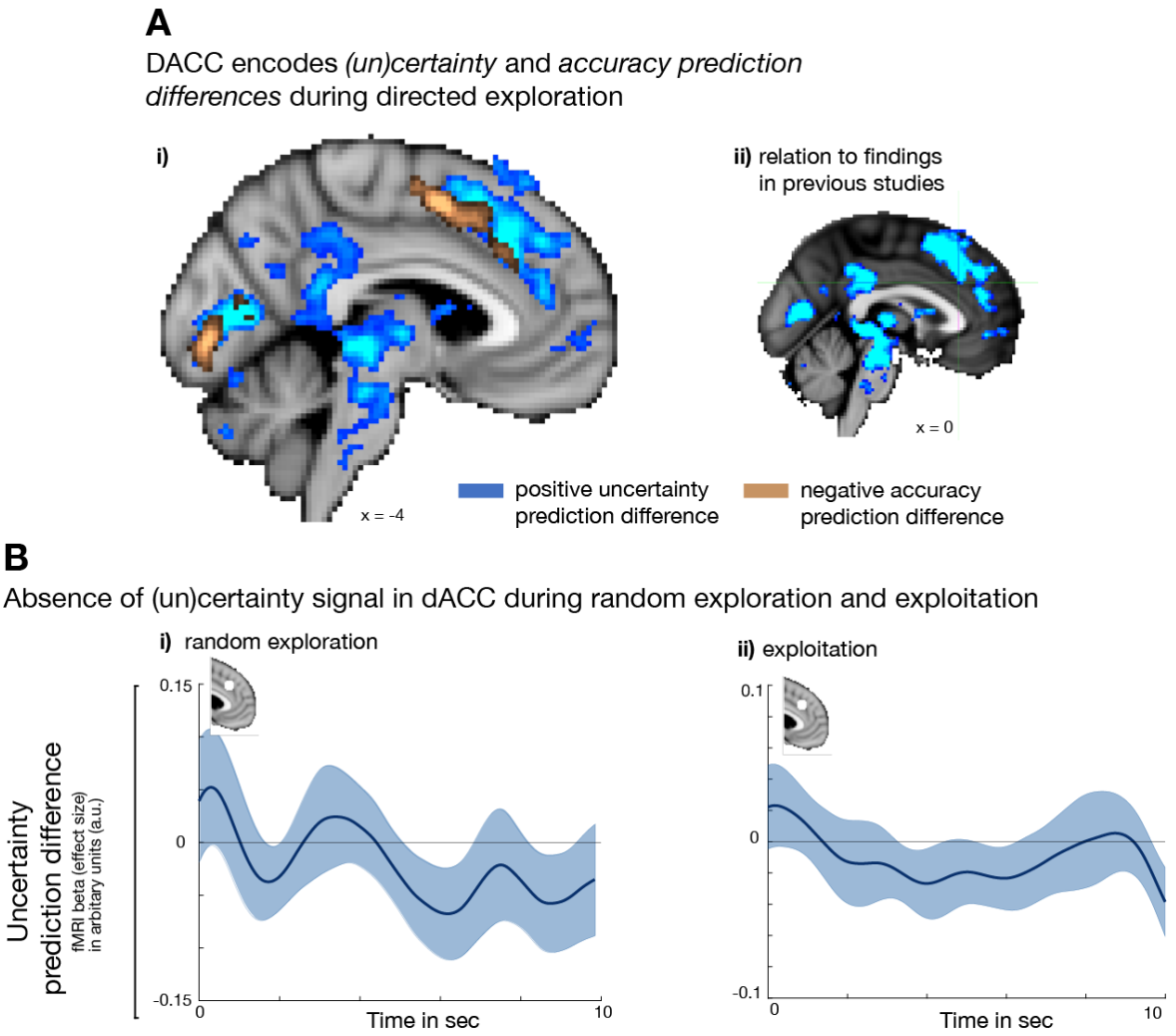


Figure 10: DACC represents uncertainty prediction differences during directed exploration, but not random exploration or exploitation. ***(A)*** *Dorsal anterior cingulate cortex (dACC) during directed exploration (compare explore/exploit trial separation in Figure 1)* ***(A-i)*** *During exploration, in the decision phase, dACC represents both components indicative of exploratory behaviour: a positive uncertainty and negative accuracy prediction difference.* ***(A-ii)*** *Activity in dACC increases during exploration (Daw et al., 2006) and choosing alternative options during foraging (Kolling et al., 2012). Fixation crosses show the peak of brain clusters reported in these previous studies. Whole-brain effects family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$.* ***(B)*** *DACC did not show a representation of uncertainty prediction difference during random exploration* ***(B-i)*** *or exploitation* ***(B-ii)****.*

Previous studies have shown the involvement of frontopolar cortex and dorsolateral prefrontal cortex (dlPFC) during exploration. Frontopolar cortex has been causally linked to the exploration of uncertain options (Zajkowski et al., 2017). Similarly, dorsolateral prefrontal cortex has been associated with the mean uncertainty between two options (Badre et al., 2012; MNI x/y/z-peak coordinates: [46 14 28]; Figure 11A). We also observed an uncertainty prediction difference in frontopolar cortex (Figure11A-i: MNI x/y/z-peak coordinates: [-40,50,14]) and dlPFC (Figure11A-i: MNI x/y/z-peak coordinates: [38 10 28]), replicating results of previous exploration studies.

Next, we tested for an uncertainty polarity change in frontopolar cortex and dlPFC. ROIs were extracted according to cluster-corrected regions that were close to previous reported areas (Zajkowski et al., 2017, Badre et al., 2012) and associated with the uncertainty prediction difference during exploration. These areas, as well as vmPFC are involved in representing uncertainty during exploration, however their profile diverges during exploitation as a polarity change is only observed in vmPFC. There was no polarity change in the representation of uncertainty when comparing exploration to exploitation, because neither area represented the certainty prediction difference during exploitation (Figure 11B-i: frontopolar cortex: $t(23) = -1.18$, $p= 0.25$; Figure 11B-ii: dlPFC: $t(23) = -1.4$, $p= 0.18$). Similar to patterns observed in dACC and other dorsomedial frontal areas, both dlPFC and frontopolar cortex have distinct profiles compared to vmPFC, as they did not represent a certainty prediction difference during exploitation and hence did not show an uncertainty polarity change across behavioural modes.

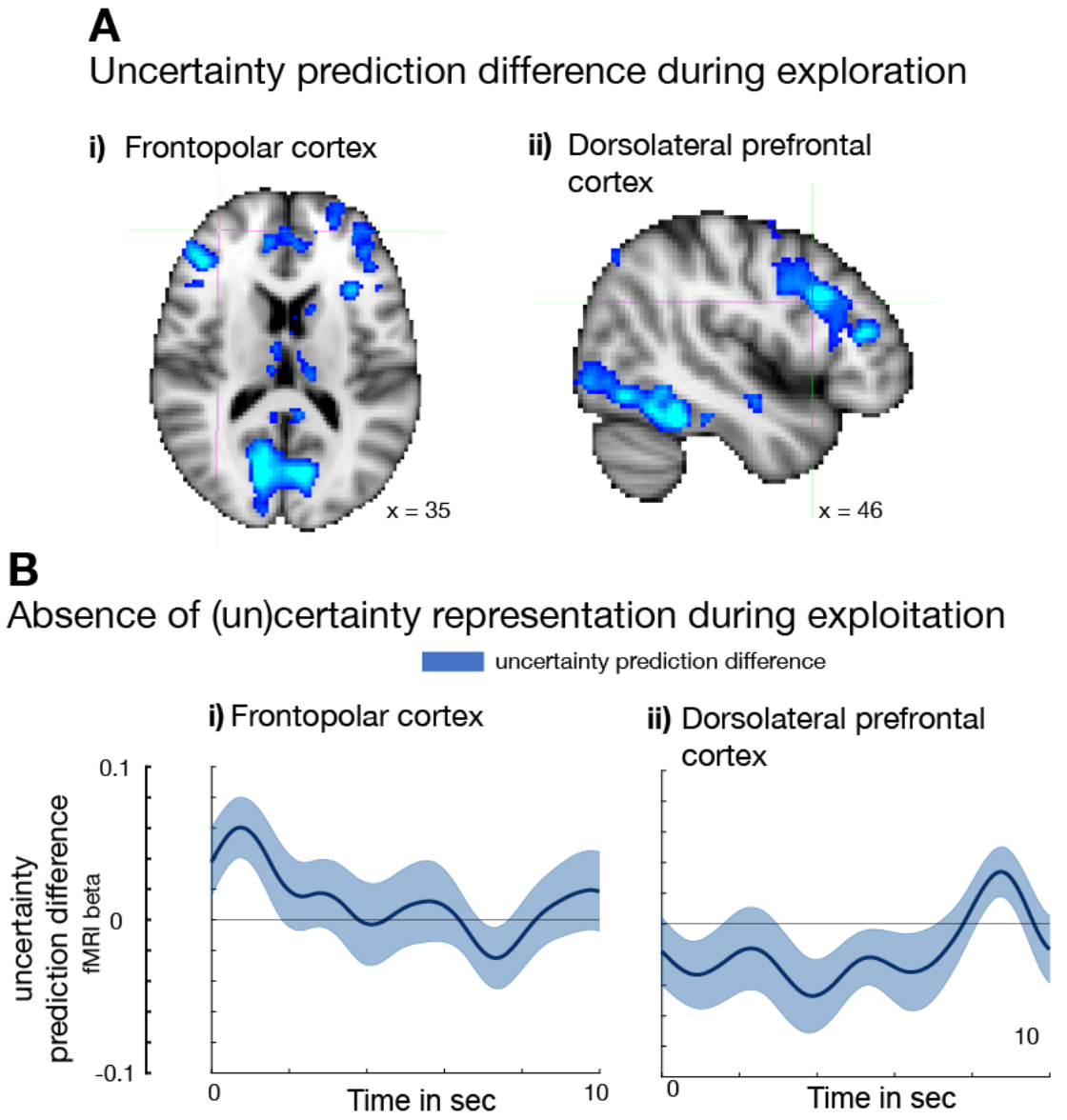


Figure 11: Frontopolar cortex and dlPFC represent uncertainty prediction difference during exploration, but not exploitation and thereby show a distinct profile to vmPFC. ***(A)*** *Our analyses also identified regions that have been previously associated with exploration, namely frontopolar cortex and dorsolateral prefrontal cortex (dlPFC). Fixation crosses show the peak of brain clusters reported in these previous studies(left: Zajkowski et al., 2017; right: Badre et al., 2012). In our study, our contrast of uncertainty prediction difference during the exploration decision phase highlights similar regions (see coordinates in main text).* ***(B)*** *There was no polarity change in the representation of uncertainty when comparing exploration to exploitation, because neither areas represented the certainty prediction difference during exploitation. Whole-brain effects were family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$.*

3.3.3. Uncertainty-related signals in subcortical structures during exploration and exploitation

We used a region-of-interest approach to test for an uncertainty prediction difference in subcortical structures during both behavioural modes. We focused on amygdala and ventral striatum as they have been previously associated with modes of exploration and exploitation(Costa et al., 2019). We also focused on ventral tegmental area (VTA) which exhibited cluster-corrected positive and negative uncertainty prediction difference during exploration and exploitation respectively (Figure 9). All three subcortical regions represented uncertainty prediction difference during at least one behavioural mode – either exploration or exploitation – but with a different pattern of activation in each case: amygdala predominantly represented uncertainty prediction difference during exploration (Figure 12A), while VS (Figure 12B) activation was most apparent during exploitative phases when it reflected a negative uncertainty prediction difference. VTA activity suggested a representation of uncertainty prediction difference during both, exploration and exploitation in the decision phase (Figure 12C). These patterns show that a network of areas including multiple cortical and subcortical areas represent uncertainty-related information during both exploration and exploitation. While it was not identical, the pattern in the VTA most closely resembled that seen in the vmPFC; it carried uncertainty signals that reversed in polarity between exploration and exploitation but there was no credible evidence for an accuracy-related signal during the transition phase between exploration and exploitation (see paragraph on transition between exploration and exploitation;

t(23) = -0.97, p=0.35, d=-0.197,95% confidence interval=[-0.07 0.026], Bayes factor$_{10}$=0.325,%error=0.037).

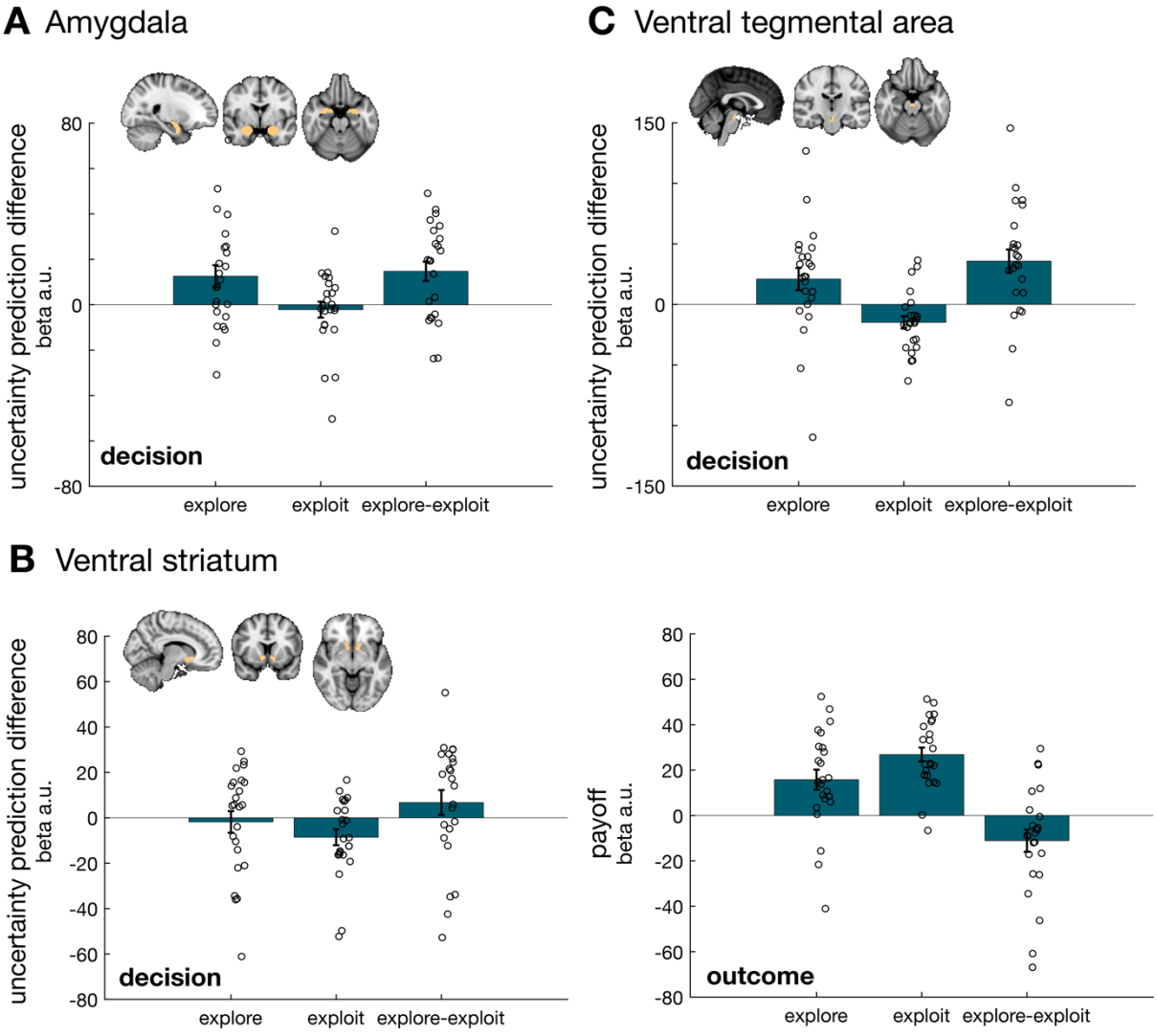


Figure 12. Uncertainty-related signals in subcortical regions during exploration and exploitation. *We show subcortical activation associated with the uncertainty prediction difference for exploration, exploitation, and their difference. Activation is shown during the decision phase, and when relevant, during the outcome phase. We used bilateral masks and averaged the results over both hemispheres for each ROI.* ***(A)*** *Amygdala represents uncertainty prediction difference during exploration more strongly than during exploitation. Activation patterns in VS during decision* ***(B-i)*** *and outcome phases* ***(B-ii)*** *suggest its primary involvement is during exploitation. VS represented both a negative uncertainty prediction difference during the decision phase during exploitation and the payoff during the outcome phase during both exploration and exploitation but it did so more strongly during exploitation.* ***(C)*** *Finally, VTA activity reflected uncertainty during both exploration and exploitation in the decision phase. (n = 24; error bars are SEM across participants).*

3.3.4. Uncertainty representation in vmPFC scales with predictor repetition

We have shown a polarity change in the effect of uncertainty/certainty on guiding behaviour and influencing vmPFC activity when comparing exploratory and exploitative behavioural modes. One possible way to interpret the certainty representation during exploitation is that vmPFC encodes a default choice(Kolling et al., 2012; Lopez-Persem et al., 2016; Noonan et al., 2012). In the context of the current task, an effective default choice is repetition of previously made choices particularly when there has been certainty about the predictor's accuracy. We therefore asked whether there was evidence of a higher frequency of choice repetition on exploitation as opposed to exploration trials. This was indeed the case, the percentage of choice repetitions during exploitation was significantly higher than during exploration ($t(23)=15$, $p = 2.5187e\text{-}19$; Figure 13A). Also note that within the two phases, this indicates a relative predominance of repetitions versus no repetitions in exploitation, but a relative predominance of no repetition choices versus repetitions in exploration.

Moreover, we examined whether activity in the same location in vmPFC reflected whether, on each trial, participants would repeat a choice they had made the last time it was offered. This was indeed the case; there was more activity in vmPFC when participants were repeating a choice made previously (repetition is coded as "repeat – no repeat": $t(23) = 4$, $p < 0.001$; Figure 13B, grey time course). In addition, the effect was greater when there was more certainty about the predictor (repetition x chosen certainty: $t(23) = 3.4216$, $p = 0.0023$; Figure 13B, red time course). We decomposed the interaction effect by binning chosen certainty

during “repetition” and “no repetition” trials at the time of the interaction effect time course peak (Figure 13B, right panel). Certainty representation increased with repetition of a previously made choice. By contrast, when switching to a new choice that had not been made on a previous trial, a behaviour more likely to occur during exploration (Figure 13A), the vmPFC had the opposite polarity (positively related to uncertainty; Figure 13B, right panel. Note that the statistical test comparing the blue and green bars was performed in the leftward panel of B by testing the interaction effect against zero).

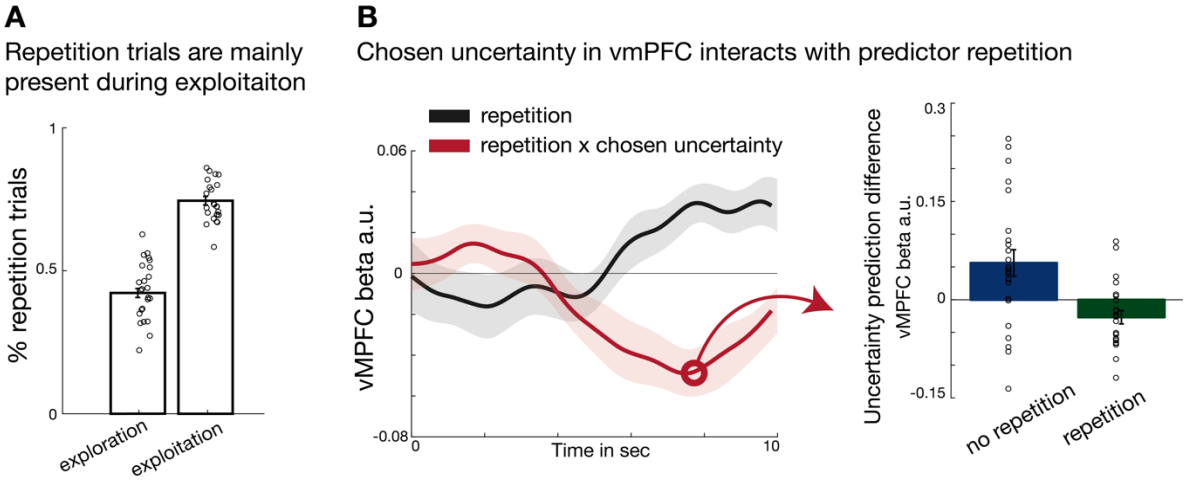


Figure 13: Interaction of repetition and uncertainty representation in vmPFC. *(**A**) The percentage of choice repetitions during exploitation was significantly higher than during exploration. Also note that within the two phases, this indicates a relative predominance of repetitions versus no repetitions in exploitation, but a relative predominance of no repetition choices versus repetitions in exploration. (**B**) VmPFC activity increased when participants repeated the same predictor selection as they had made on the last encounter with the predictor (grey time course; repetition is coded as “repeat – no repeat”). Moreover, we found a significant interaction effect of repetition x chosen uncertainty (red time course). The interaction effect is illustrated in the right panel by decomposing it into the binned effects of chosen uncertainty during “repetition” and “no repetition” trials at the time of the interaction effect time course peak. This indicates that the increase in BOLD response accompanying choice repetition was even stronger if participants were very certain about their choice (i.e. negative uncertainty during repetition; green bar in right panel); whereas in case of switching choices, the BOLD signal increased as a function of chosen uncertainty (i.e. positive uncertainty; blue bar in right panel). Note that the statistical test comparing the blue and green bars was performed in the leftward panel of B by testing the interaction effect against zero (n = 24; error bars are SEM across participants).*

3.3.5. The transition from uncertainty to certainty representation is accompanied by the processing of accuracy between predictors

So far we have shown that the transition from exploration to exploitation and choice repetition behaviour is accompanied by a change in the polarity of uncertainty/certainty signals and emergence of choice repetition signals in vmPFC. However, it remains unclear how the transition between directing behaviour towards uncertain or certain predictors occurs as the behavioural mode shifts from exploration to exploitation. It is possible that, after initial exploration but before repetitively choosing certain predictors there might be a phase in which participants focus on how well – how accurately – each predictor estimates the target's location (Figure 14A, see illustration). Such a period might naturally precede a period when the most accurate predictors are identified and continuously chosen. During such a transition period, one would expect neural activity correlating with an accuracy prediction difference, the difference between the accuracy estimates associated with the chosen and unchosen predictors. Moreover, because participants are transitioning from uncertainty- to certainty-guided behaviour, the subjective accuracy estimates held by participants for the chosen and unchosen predictors should be close in value. This would suggest that participants have no strict preference between predictors yet, as they are still learning about them. We identified a new subset of trials by selecting trials with accuracy prediction differences close in value (Figure 2). We tested whether these trials occurred at points in time in between the exploration and exploitation periods that we had previously identified. Indeed, these transition trials occurred later in time compared to previously defined explorative choices (paired t-test, explore vs. transition:

t(23)=5.9, p=4.0161e-07) and earlier in time compared to exploitative choices (paired t-test, exploit vs. transition: t(23)=-2.3, p=0.03) (Figure 14A).

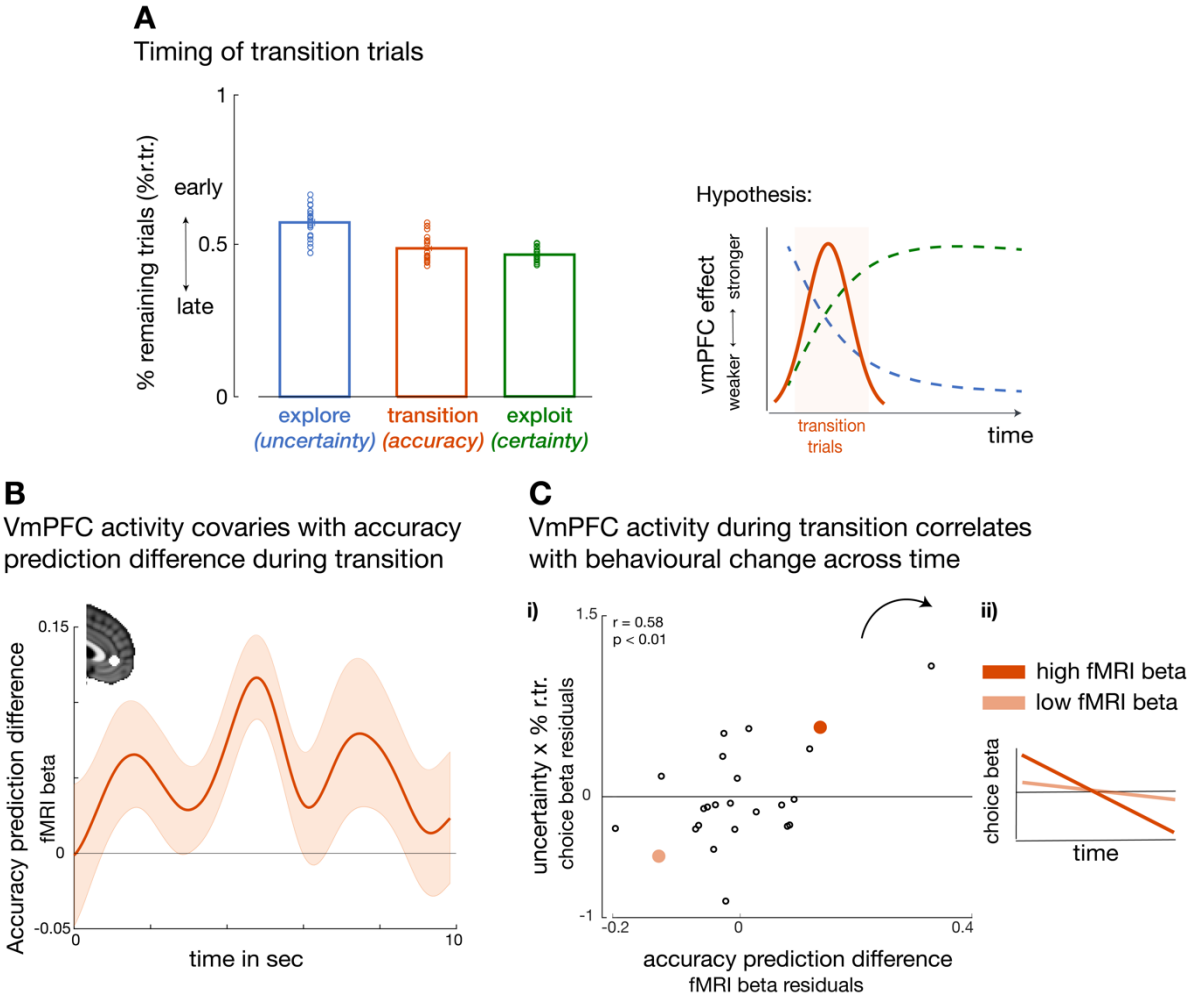


Figure 14: Accuracy processing mediates uncertainty polarity change from exploration to exploitation. (A)*Transition trials (Figure 2) occurred later than exploratory selections and earlier than exploitative selections (left panel). We hypothesized activation in vmPFC to be correlated with uncertainty, accuracy and certainty prediction differences between predictors, but at different times during the experiment (see illustration, right panel).* ***(B)*** *During transition trials, activation in vmPFC covaried with the difference in the accuracy between the chosen and unchosen predictor, i.e. accuracy prediction difference.* ***(C)*** *(i) Participants who showed a stronger vmPFC accuracy prediction difference during the transition period (variability around time course peak from panel b), also integrated more drastically the uncertainty between predictors across time into their choice behaviour (uncertainty x % r. tr.) from Chapter 2: Figure 4A). (ii) For illustration, this means that participants with stronger accuracy-related vmPFC activation had a stronger change in integrating uncertainty across time, i.e. a stronger slope in the uncertainty x % r. tr. effect. The illustration depicts two example participants, dark orange indicates a subject with both a strong vmPFC accuracy activation and pronounced behavioural change in how uncertainty was used to drive choice behaviour. By contrast, the participant indicated in light orange shows a weak vmPFC BOLD accuracy effect and only a small change*

in how uncertainty was used over time. These findings support the idea that the transition between uncertainty-driven exploration to certainty-driven exploitation is mediated by representing the accuracy between predictors. (n = 24; error bars are SEM across participants).

We then examined whether vmPFC activity reflected the accuracy prediction difference during this transitional period. To test for this effect, we chose an independent ROI in vmPFC extracted from the cluster-corrected uncertainty prediction difference effect across all trials (see Figure 4A). As predicted, activation in vmPFC correlated with an accuracy prediction difference during this transitional phase ($t(23) = 3.5$, $p= 0.002$; Figure 14B).

In further support of the suggestion that accuracy processing is especially prominent during this transition phase (in which chosen and unchosen predictors have similar accuracy values), we tested for an accuracy prediction difference in vmPFC during selections of very inaccurate predictors and very accurate predictors (Figure 2; trials on, respectively, the left (Figure 2-i) and right-hand (Figure 2-ii) side of the transition trials). These selections are likely to occur during very exploratory or exploitative periods, respectively. We extracted time courses of uncertainty (blue) and certainty (green) or accuracy (orange) prediction differences in the previously defined vmPFC region that showed a cluster-corrected effect of, respectively, accuracy or uncertainty prediction differences across all trials. We found that an accuracy prediction difference signal was absent from vmPFC when very inaccurate predictors (accuracy prediction difference: $t(23) = -0.84$, $p =0.41$; Figure 15A) or very accurate predictors were taken (accuracy prediction difference: $t(23) = -1.3$, $p =0.21$; Figure 15B). Moreover, during extreme inaccuracy-driven choices that are prominent during exploration, uncertainty prediction difference are

represented in vmPFC as a trend (t(23) = 2, p= 0.059) while during extreme accuracy-driven trials, which are prominent during exploitation, vmPFC activation represents certainty prediction difference (t(23) = 3.8, p< 0.001). Note that we used here an even stricter cut-off to define exploration/exploitation periods (compared to Figure 1), and it was still possible to show support for a change in the polarity of vmPFC signals, from uncertainty- to certainty-related during exploration and exploitation, respectively. This pattern of results suggests that the periods in which vmPFC activity reflects the uncertainty and then the certainty of the predictor chosen to guide behaviour are separated by a transition period in which vmPFC reflects the accuracy estimate of the predictor chosen to guide behaviour.

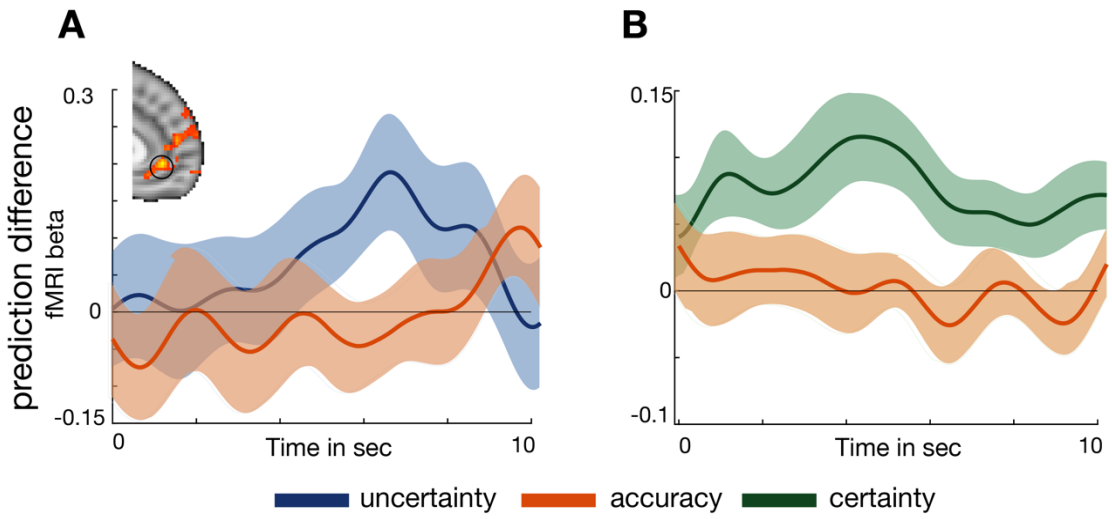


Figure 15: Differences in predictor accuracy are represented in vmPFC during a transitional period between exploration and exploitation but they are less prominent on very exploratory or very exploitative trials. *These results support the contention that vmPFC processes accuracy prediction differences during the transition period between exploration and exploitation. Accuracy prediction difference (orange timecourse) in vmPFC is specific to transitional trials; vmPFC lacks an accuracy prediction difference signal during strong inaccuracy-driven* ***(A:*** *trials on the left-hand side of the transition trials) and strong accuracy-driven* ***(B:*** *trials on the right-hand side of the transition trials) selections. However, both uncertainty (blue time course) and certainty (green time course) prediction differences can be replicated in vmPFC even with a stricter trial cut-off compared to previous analyses (Figures 1,6).*

Next, we tested whether activation during the transition phase was related to behavioural changes across time – from uncertainty-driven to certainty-driven behaviour – when selecting between predictors. As the transition phase bridges exploration (positive uncertainty) to exploitation (positive certainty), we tested whether accuracy-related vmPFC activation during the transition period was related to a behavioural effect of uncertainty that changes across time, i.e. the interaction term uncertainty x % r.tr. (see behavioural choice GLM, Chapter 2: Figure 4A). We used a partial correlation analysis to examine the relationship between each individual's accuracy-related vmPFC activity extracted from the vmPFC cluster (accuracy prediction difference effect across all trials) and the transition from uncertainty-driven to certainty-driven predictor selection behaviour. In the same analysis, we controlled for all other behavioural variables included in the previous GLM1 (see GLM1, Chapter 2: Figure 4A). We found that accuracy prediction difference-related activity in vmPFC during the transition period was positively correlated with uncertainty x % r.tr ($r= 0.58$, $p= 0.007$; Figure 14C-i). That is, the larger the vmPFC signature encoding accuracy prediction difference during the transition period, the stronger the behavioural transition from uncertainty-driven to certainty-driven behaviour over the course of a block (Figure 14C-ii). Notably, these results were not confounded by variation across participants' in the number of transition trials that were identified; a partial correlation that controlled for the number of transition trials remained significant ($r=0.43$, $p=0.04$).

This result suggests that a transition phase during which the accuracy between predictors is represented in vmPFC may facilitate a neural polarity change from first representing uncertainty when selections are exploratory to later,

representing certainty when repeatedly selecting the same predictor during exploitation (Figure 15). Participants exhibiting stronger predictor accuracy signals in vmPFC during the transition period exhibited a more drastic change from uncertainty- to certainty-driven behaviour.

3.4. Summary

The combination of computational modelling and fMRI made it possible to show that beliefs concerning the accuracy of predictors and the (un)certainty of those beliefs inform predictor selection to guide behaviour. Their influence on both behaviour and activity in vmPFC changed and transitioned in tandem. The vmPFC carried information about a multiplicity of decision variables (uncertainty, accuracy, certainty and repetition), the strength and polarity of which varied according to their relevance for the current context.

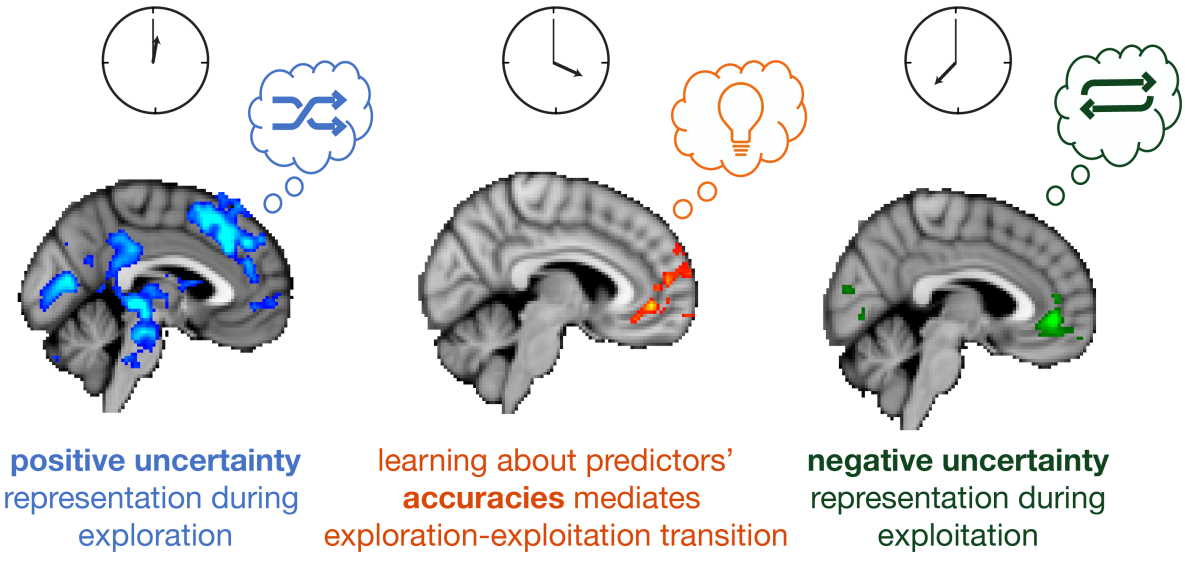


Figure 16: Summary. From exploration to exploitation: polarity of subjective uncertainty in vmPFC changes with behavioural mode. *At the beginning of a block, choices are exploratory and directed towards uncertain predictors (like a shuffle mode when playing music,*

left panel). VmPFC and an extended network centred in dACC represent the difference in uncertainty between the predictors that might be selected. With time passing, participants learn about the predictors' accuracy through observing how well they predict an outcome. A participants' belief in the accuracy of the predictors exerts the predominant influence on vmPFC activity during this transition phase (middle panel). Towards the end of a block, vmPFC activity represents the difference in certainty between the predictors. In this exploitative period, choices are repeatedly directed towards certain predictors (like a repeat mode, right panel). We show that vmPFC carries information about a multiplicity of decision variables, the strength and polarity of which vary according to their relevance for the current context: exploration, exploitation or their transition.

3.5. Discussion

VmPFC activity reflected different decision variables at different times in a manner that reflected their relevance for the current context of exploration or exploitation. We have shown a polarity change in the influence that uncertainty in one's belief exerts not just on behaviour but also on vmPFC activity. During exploitative modes, when differences in predictor certainty are the key decision variable, vmPFC reflects certainty prediction difference, but when participants are in an explorative mode, vmPFC activity reflects uncertainty prediction differences. In the latter case, vmPFC is co-active with an extensive network of regions carrying a similar uncertainty-related signal during exploration.

In accordance with the behavioural results, we found that neural activity reflected predictor differences. Activity in vmPFC reflected the difference between the selected and rejected predictor, in terms of the key feature that was currently of relevance for guiding behaviour: first uncertainty, then accuracy, and finally certainty. Previous studies have often focused on the manner in which activation in vmPFC is correlated with differences in the reward values of chosen and rejected choices (Boorman et al., 2009; Hunt et al., 2012; Rushworth et al., 2011). In such studies, differences in the reward values associated with the choices constitute the

evidence in favour of taking one choice rather than the other. Here, we show when selecting between predictors to guide behaviour, multiple types of information, rather than just a single one, can be of importance. This can be linked to the idea that vmPFC integrates a diverse set of variables that are currently choice-relevant (Wilson et al., 2014b) and to recent evidence that exploitation and exploration are not simply behaviours that are controlled by completely separate neural circuits but rather they are, at least in part, controlled by changes in mode within neural structures (Costa et al., 2019).

However, our results also suggest that vmPFC does not guide behaviour in isolation and that there are additional broader differences in the recruitment of choice-relevant brain networks between exploration and exploitation. Although activation associated with certainty prediction difference during exploitation was mainly present in vmPFC, uncertainty prediction difference during exploration was associated with a wider network including areas such as dACC, dlPFC, and frontopolar cortex that have previously been associated with exploration(Badre et al., 2012; Zajkowski et al., 2017). Activation in dACC has often been related to behavioural adaptation and the search for better alternatives, for example during foraging (Boorman et al., 2011, 2013b; Daw et al., 2006; Hayden et al., 2011a; Kolling et al., 2012, 2016b, 2016a, 2018; Quilodran et al., 2008; Stoll et al., 2016; Wittmann et al., 2016a) and to the update of internal models during environmental changes (Karlsson et al., 2012; O'Reilly et al., 2013; Tervo et al., 2014). Our results may therefore suggest that in some cases during exploration wider updates in decision networks occur that encompass both vmPFC, dACC and prefrontal areas in a similar fashion. Nevertheless, it is

important to remember that the pattern of activity in vmPFC, when considered across both behavioural modes, is different from that seen in dACC, dlPFC, and frontopolar cortex where activity only reflects uncertainty during exploration while the change in the polarity of (un)certainty-related activation, between exploration and exploitation, only occurs in vmPFC.

Anterior frontal cortex and dACC have been shown to track and represent currently counterfactual choices that might be taken in the future choice if behaviour changes. Even though both dACC and frontopolar cortex represented uncertainty-related signals during exploration, only dACC's signal represented both components of exploratory behaviour: a positive uncertainty prediction difference and a negative accuracy prediction difference. Both components define directed exploratory behaviour in this study, as they represent choices of uncertain and inaccurate predictors. This profile suggests that dACC represents information about multiple attributes that define the potential of exploration and thereby impacts the choice of engaging in further information search (Kaanders et al., 2020).

Additionally, vmPFC did not carry a clear uncertainty signal during random exploration as opposed to uncertainty-guided exploration. An important new finding is that effective exploratory behavior may simply emerge from noise in the learning process(Findling et al., 2019) and this may impact on activity in brain areas such as dACC that reflect choice value learning at multiple time scales (Bernacchia et al., 2011; Meder et al., 2017; Wittmann et al., 2016a). However, the current findings suggest that an (un)certainty signal is also carried in these areas when it is relevant for behaviour.

A related line of research supports the notion that vmPFC not only represents the value difference between choice options, but also a second-order representation, that is one's own confidence in a choice (De Martino et al., 2012; Lebreton et al., 2015). These results are compatible with our finding that both accuracy and uncertainty are represented in vmPFC. However, we show in addition that the polarity of the uncertainty representation (which is a second-order representation similar to confidence) in vmPFC changes depending on the behavioural mode. This suggests that in some cases second-order representations in vmPFC are themselves choice-guiding and highly context sensitive. The change in signal in vmPFC from signalling uncertainty prediction differences to signalling certainty prediction differences, i.e. uncertainty polarity change in vmPFC, might be related to the presence of a learning phase during which predictors' accuracies are compared. We identified a transition phase between exploration/exploitation periods, when no clear preference had yet been formed for predictors. At that point, we observed that vmPFC most prominently reflected participants' accuracy estimates for the predictors. Notably, the accuracy effect in vmPFC during the transition phase was related to the degree of change from uncertainty- to certainty-driven behaviour exhibited by each participant: participants exhibiting stronger accuracy-related vmPFC activation during the transition period also showed more drastic behavioural changes.

Although predictor selections were accuracy-guided throughout the task, we did not observe an accuracy prediction difference in vmPFC during the final exploitation stages of predictor selection. This is similar to the way in which vmPFC activity related to value comparison during choice selection has been shown to be

stronger during earlier compared to later phases of a task (Hunt et al., 2012). A predictor accuracy representation was present in vmPFC during the transition phase between exploration and exploitation when accuracy estimates between predictors were close in value, meaning that a careful comparison between predictors was required to guide predictor selections successfully. In comparison, during exploitative trials participants established which predictors were accurate resulting in repeated selections of the same predictors. At that point vmPFC activity reflected this repetitive mode of decision making and it did so in a manner that interacted with the representation of certainty about the predictor.

3.6. Supplementary information

3.6.1. fMRI cluster-corrected peak coordinates

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Certainty prediction difference: chosen - unchosen	Ventromedial prefrontal cortex	4 44 -2	4.44
	Occipital cortex	-8 -70 -4	3.59
Accuracy prediction difference: chosen - unchosen	Ventromedial prefrontal cortex	6 44 -4	3.98
	Superior temporal gyrus	-52 -2 -12	3.38

Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$

Table1: These cluster-corrected coordinates relate to results depicted in Figure 4.

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Exploration: uncertainty prediction difference	Ventromedial prefrontal cortex	0 46 -4	3
	Dorsal Anterior Cingulate Cortex	-4 30 34	5.44
	Dorsal lateral prefrontal cortex (right)	38 10 28	5.75
	Frontopolar cortex (left)	-40 50 14	3
	Superior temporal gyrus	48 -10 -14	4.38
	Angular gyrus	30 -52 42	3.5
	Fusiform gyrus	30 -76 -12	7.1
Exploitation: certainty prediction difference	Ventromedial prefrontal cortex	-4 46 -2	4.26
	Anterior insular (right)	34 18 -10	4.57
	Ventral tegmental area	0 -22 -22	4.11
	Occipital cortex	-24 -72 -12	3.79

	Occipital cortex	26 -80 -12	3.99
Exploration–Exploitation: uncertainty prediction difference	Ventromedial prefrontal cortex	-2 46 -4	5.1
	Dorsal Anterior Cingulate Cortex	-2 28 38	5.9
	Fusiform gyrus	30 -76 -12	9
	Cerebellum	-2 -54 -36	6.7
	Superior temporal gyrus	48 -8 -14	4.9

Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$

Table2: These cluster-corrected coordinates relate to results depicted in Figure 9.

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Domain general: (accuracy prediction difference(*1) + uncertainty prediction difference*(-1))	Ventromedial prefrontal cortex	6 44 -4	5.77
	Posterior cingulate cortex	-4 -38 38	4.1
	Dorsolateral prefrontal cortex	50 36 -4	4.05
	Superior temporal gyrus	-60 -4 -10	3.84
	Superior temporal junction	-64 -44 12	4.07
	Brainstem	8 -48 -42	4.19
Conjunction analysis: (accuracy prediction difference – uncertainty prediction difference)	Ventromedial prefrontal cortex	6 44 -4	3.9

Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$

Table3: These cluster-corrected coordinates relate to results depicted in Figure 5.

3.6.2. Alternative computational models

We used a Bayesian model with uniform priors at the start of each block for all four predictors, assuming participants do not have prior knowledge about the underlying distributions associated with predictors. We refer to this model as 'the original model' because it is the model used elsewhere in this study. We compared the original model to two alternative computational models: a Bayesian model with informative priors (Supplementary Figure 1) and a reinforcement learning (RL) model which tracks the payoff history of each predictor (Supplementary Figure 2). We explain in detail the rationale behind each computational model, their construction and the results below. The results demonstrate that a Bayesian model using uniform priors had a better model fit compared to a Bayesian model with informative priors or an RL model. However, a combination of the original Bayesian model with uniform priors and value-based variables derived from an RL model showed the best model fit to choice behaviour. In conclusion, RL value terms complement the Bayesian model but do not substitute for the Bayesian model terms as an explanation of behaviour; participants' beliefs in the accuracy and uncertainty of a predictor explained additional variance in choice behaviour above and beyond that explained by their choice value estimates.

3.6.2.1. Bayesian model with informative priors

There is a possibility that by selecting different predictors across time, participants form a representation of the likelihood of sigma values that can be associated with predictors. In consequence, participants' beliefs at a block start would not be represented by a uniform prior distribution but rather by a prior

distribution that had been informed by past observations. To test this, we developed a Bayesian model with priors that have been informed by past observations, we refer to these as informative priors (Supplementary Figure 1A). Ultimately, we compare the model fit between a Bayesian model with informative and uniform priors, i.e. referred to as the original model used in the current study.

We modified the Bayesian model to include block-wise priors that reflect the previous history of all observations (irrespective of predictor). First, to obtain informative priors for each block start, we constructed a separate Bayesian model that did not differentiate between predictor identity but instead learnt about all observations as derived from a single distribution ('average predictor learner'). We used a uniform distribution for the first block for all participants. The subsequent prior distributions at the end of each block differed across participants as the shape of the distribution was dependent on individual choice behaviour and on the number of observations within a block. Second, we combined these informative priors with the original Bayesian model used in the manuscript. In other words, at the beginning of each block, we set the prior distribution for each predictor to the posterior of the average predictor learner from the end of the previous block (Supplementary Figure 1A, bottom illustration). Then, however, each predictor's distribution was updated according to the observations specific to each predictor (referred to as adaptive model, green colours).

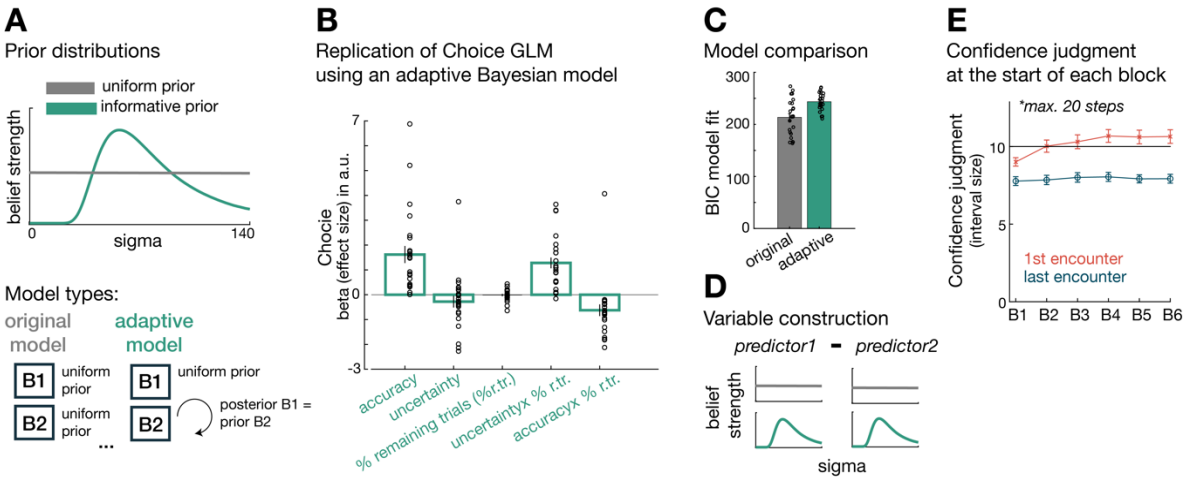
We derived estimates of accuracy and uncertainty and repeated all behavioural analyses of interest. We replicated all effects of interest even when using informative priors (accuracy, $t(23) = 4.7$, $p<0.001$, $d=0.96$, 95% confidence interval=[0.91 2.3]; uncertainty, $t(23) = 1.2$, $p= 0.25$, $d=-0.24$, 95% confidence

interval=[-0.77 0.21]; uncertainty x % r.tr., t(23) = 6, p<0.001, d=1.2, 95% confidence interval=[0.83 1.73]; accuracy x % r.tr., t(23) = 2.6, p= 0.015, d=-0.54, 95% confidence interval=[-1.1 -0.13]) (Supplementary Figure 1B). Next, we compared the model fit between the adaptive and original Bayesian model (Supplementary Figure 1C). We show that a Bayesian model using uniform priors has a better fit to choice behaviour than the adaptive Bayesian model. One reason might be that a uniform prior provides more flexibility for estimates to converge towards their true value across time.

It is important to keep in mind, that for behavioural and neural analyses, variables are constructed in relative terms, for example as the difference between left and right predictors or the difference between chosen and unchosen predictors. Hence, their relative values rather than their absolute values determine choice behaviour. When changing prior belief distributions, only the absolute values change while the relative values keep the same proportions (Supplementary Figure 1D). For this reason, our key results remain unchanged when modifying initial block-wise priors.

Next, we used a descriptive analysis to show that prior beliefs do not show a major change across blocks (Supplementary Figure 1E). We used the confidence interval size at each block start averaged across four predictors as an index of prior beliefs and compared it across all (six) blocks. A one-way ANOVA applied to the first encounters across six blocks (red line) shows that participants increase their CI size across time (Mauchly's test indicated a violation of equal variances: $x^2(14)=37.4$,p=0.001, therefore we used Greenhouse Geiser test: F(2.9,66.23) = 17.15, p<0.001, $\eta^2=0.9$). This effect however is mainly driven by the difference

between the first and the remaining blocks. Excluding the first block from the analysis did not show credible evidence for a change of CI sizes during the first encounters across blocks (Mauchly's test indicated a violation of equal variances: $x^2(9)=23, p=0.006$, therefore we used Greenhouse Geiser: $F(2.6,92) = 2.5$, $p=0.076$, $\eta^2=0.71$, Bayes $factor_{10}=0.91$, error%=0.36). Participants performed practice trials before the actual experiment during which they encountered very accurate predictors. Although participants were instructed that these observations were not reflective of the real experiment, nevertheless beliefs prior to the first block might have been impacted by observations made during these practice trials. In conclusion, analysis of the CI does not support the contention that participants narrow their initial expectation about the predictors over the course of the experiment. This suggests further that a very broad initial prior is a plausible assumption for the Bayesian modelling and suggests why a Bayesian model using a uniform prior is a better model fit for choice behaviour.



Supplementary Figure 1. Model comparison: Bayesian model with informative prior. A: *Participants might form expectations about possible sigma values across time. Then prior beliefs at block starts should reflect information gathered from past observations. We constructed a Bayesian model that incorporated block-wise priors that reflect the previous history of all observations irrespective of predictor (referred to as adaptive model; green model) and compared it to the Bayesian model using uniform priors (referred to as original model; grey model).* ***B:*** *We replicated all effects of interest when deriving belief estimates with the adaptive Bayesian model shown in Chapter 2, Figure 3A.* ***C:*** *The original Bayesian model was a better model fit for all behavioural analyses (here we only show across all trials) compared to the adaptive Bayesian model. One reason might be that a uniform prior provides more flexibility for estimates to converge towards their true value across time.* ***D:*** *Importantly, for behavioural and neural analyses, variables are constructed in relative terms (e.g. difference between left and right predictor): changing prior distributions only impact absolute values while the relative values keep the same proportions. For this reason, the key results remain unchanged when modifying initial block-wise priors.* ***E:*** *Next, we used the confidence judgment (i.e. interval size) at each block start averaged across four predictors as an index of prior beliefs and compared it across all (six) blocks. We show confidence judgments at the start of a block (red line) and at the end of a block (blue line). Participants reset their prior beliefs from one block to the next (blue line). Analysis of the CI shows no credible evidence for a change of CI sizes during the first encounters across blocks when excluding blocks that were affected by practice trials.(n = 24; error bars are SEM across participants).*

3.6.2.2. Reinforcement learning model tracking payoff history

Next, we compared the original Bayesian model to a reinforcement learning model (RL) (Supplementary Figure 2). We constructed an RL that learnt about the payoff history associated with each predictor. The payoff scheme used in this experiment reflects the participants' beliefs in the accuracies and certainties

associated with chosen predictors, however it may itself exert an additional independent effect on behaviour and neural activity.

For every subject, we fitted a standard RL model to each predictor's payoff history to estimate the expected value associated with that predictor at a given trial. On every trial t, the expected value associated with a predictor was updated using a prediction error (PE) based learning rule with a learning rate α as a free parameter:

$$Value_{(t+1,\ predictor)} = value_{(t,\ predictor)} + \alpha * PE_{(t,\ predictor)}$$

The PE was calculated according to the difference between the payoff and the value estimate for the specific predictor at the current trial t:

$$PE_{(t+1,predictor)} = payoff_{(t,\ predictor)} - value_{(t,\ predictor)}$$

For behavioural analyses, we calculated a decision variable (DV) for each trial that reflected the difference in the expected values between the left and right predictors:

$$DV_{left\text{-}right\ predictors} = value_{(left\ predictor)} - value_{(right\ predictor)}$$

We used a softmax function with an inverse temperature ß to calculate the probability of making a leftwards choice on each trial:

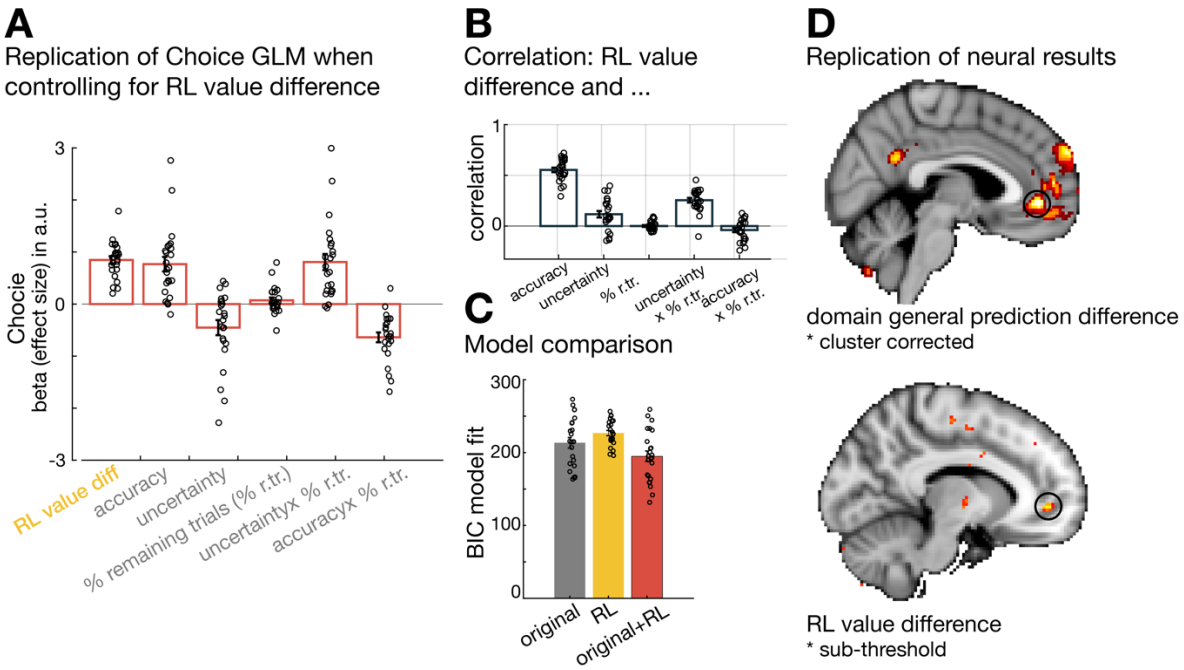
$$P(left\ predictor) = 1/\ (1+ \exp(-ß * DV_{left\text{-}right\ predictors}))$$

Note for neural analyses, the DV was calculated between the chosen and unchosen predictors in their respective expected value. First, we included the RL-derived value difference between predictors on a given trial into our main GLM to

investigate whether the other effects of interest remained stable. We replicated all effects of interest when controlling for RL value difference (accuracy, $t(23) = 5.5$, $p<0.001$, $d=1.1$, 95% confidence interval=[0.48 1.1]; uncertainty, $t(23) = -3.1$, $p= 0.0049$, $d=-0.63$, 95% confidence interval=[-0.75 -0.15]; uncertainty x % r.tr., $t(23) = 5.2$, $p<0.001$, $d=1.1$, 95% confidence interval=[0.49 1.13]; accuracy x % r.tr., $t(23) = -6.8$, $p<0.001$, $d=-1.4$, 95% confidence interval=[-0.84 -0.44] and RL value difference, $t(23) = 11.9$, $p<0.001$, $d= 2.43$, 95% confidence interval=[0.7 0.99]) (Supplementary Figure 2A). This is consistent with the relative lack of correlation between variables derived from the Bayesian model and the RL model (Supplementary Figure 2B). Next, we compared the fits between models of choice behavior for all trials (Supplementary Figure 2C). There are three possible models: a GLM applied to choice behaviour first, only including regressors from an RL model (regressor: RL value difference, yellow), second, a model only including previously used Bayesian-derived regressors (regressors: Bayesian model/original model, grey) or third, a combination of regressors of both models (regressors: RL value difference and Bayesian model regressors). A GLM model that combined (combination shown in red bars) regressors of the RL value model and the original Bayesian model was the best fit for choice behaviour, supporting the relevance of value-based and information-based variables in explaining choice behaviour.

As the combination of both RL-derived variables and Bayesian-derived variables showed the best model fit, we repeated a whole-brain analysis including regressors from both models to apply to neural data across all trials (see Methods, fMRI-GLM1 for details on the original model, to which we added RL value difference as regressor). This analysis had two aims: first, to replicate a domain general

prediction difference across trials when controlling for variance explained by the RL value difference and second, to investigate brain regions associated with RL value difference. We replicated a domain general prediction difference in vmPFC even when we incorporated the RL value difference into the GLM model (Supplementary Figure 2D; top: domain general prediction difference; cluster-corrected, MNI x/y/z-peak coordinates: [4 44 -4]; z-value: 4.05). We did not find any activation significantly cluster-corrected for RL value difference. However, the activation that was strongest was located within vmPFC (Supplementary Figure 2D; bottom panel: RL value difference; MNI x/y/z-peak coordinates: [-10 46 -2]; z-value: 2.27]. In conclusion, RL value terms complement the Bayesian model but do not substitute for the Bayesian model terms as an explanation of behaviour; participants' beliefs in the accuracy and uncertainty of a predictor explained additional variance in choice behaviour above and beyond that explained by their choice value estimates.



Supplementary Figure 2. Model comparison: reinforcement learning model tracking payoff history. A: *The payoff scheme reflects participants' beliefs in the accuracies and certainties associated with selected predictors, however it may itself exert an additional independent effect on behaviour and neural activity. We constructed a reinforcement learning (RL) model that tracked each predictor's payoff history in a recency-weighted way and compared it to the Bayesian model using uniform priors (referred to as original model). A-i: Behavioural effects of interest (Chapter 2, Figure 3A) were replicated when controlling for the RL-derived value difference (in yellow) and additionally showed an effect of RL value difference. A-ii: This is consistent with the relative lack of correlation between variables derived from the Bayesian model and the RL-derive value difference. A-iii: A combination (red bar) of the RL value model (yellow bar) and original Bayesian model (grey bar) was the best fit for choice behaviour, supporting the relevance of value-based and information-based variables for explaining choice behaviour. A-iv: We repeated the previous whole-brain analysis (fMRI-GLM1) and additionally included the RL value difference. We replicated a domain general prediction difference in vmPFC (upper panel), while there was no cluster-significant activation for RL-derived value difference (lower panel); however, the activation that was strongest was located within vmPFC. In conclusion, RL value terms complement the Bayesian model but do not substitute for the Bayesian model terms as an explanation of behaviour. (n = 24; error bars are SEM across participants; whole-brain effects family-wise error cluster corrected with $z > 2.3$ and $p < 0.05$).*

4. CHAPTER: Subjective confidence judgments about the performance of social and non-social predictors

4.1. Introduction

Most of our decisions are made in social contexts and as such we often integrate our own beliefs with those of others to make decisions (Behrens et al., 2008; Campbell-Meiklejohn et al., 2017; De Martino et al., 2017; Hertz et al., 2017; Kappes et al., 2020). To determine how much we should rely on others and the information they provide, we should consider how confident we are in the other person's advice and we can do that by forming beliefs about their expertise (Boorman et al., 2013a). However, before we receive information from others, we first need to decide whom to ask for advice. In previous chapters, we have shown that choosing between information sources requires us to consider multiple aspects of our own beliefs about those sources: how accurate is the information that they provide and how certain are we about the accuracy level that we estimate they have. Whether there are different mechanisms involved when making selections between predictors or forming beliefs about a single social or non-social predictor remains unclear. In this Chapter, I will compare and contrast how such decisions are made when choosing between social or non-social predictors.

In previous chapters, we have shown that ventromedial prefrontal cortex (vmPFC) computes the decision value relevant to the current context when making a selection between predictors (Chapter 3; see also (Basten et al., 2010; Boorman

et al., 2009; Chau et al., 2014). However, brain areas that are involved in mentalizing, such as temporoparietal junction (TPJ), dorsomedial prefrontal cortex (dmPFC) and dorsolateral prefrontal cortex (dlPFC) (Amodio and Frith, 2006; Mars et al., 2013; Schurz et al., 2014; Wittmann et al., 2016b) have also been implicated when we try to understand and predict the behaviour of others. This raises the possibility that might also be involved in deciding which information source to solicit information from. This might, however, only happen when the information is clearly another social agent.

Although brain areas such as TPJ and dmPFC seem closely associated with social cognition, there are also overlapping brain areas that underlie reward and motivational processes for social and non-social behaviour (e.g. ventral striatum, amygdala, vmPFC, insula) (Behrens et al., 2008; Boorman et al., 2013a; Ruff and Fehr, 2014; Will et al., 2017; Zink et al., 2008). However, conclusions about mechanisms that operate in common in both social and non-social domains are often based on comparisons across studies as there are only relatively few studies comparing social and non-social processing within the same participants (Ruff and Fehr, 2014).

In this Chapter, I present behavioural and neural results when participants perform a social and non-social version of the task described in previous chapters. Participants performed both versions, which were identical in their statistical features but differed in their framing. We designed the task such that participants only learnt about the predictors (social or non-social) and not the statistical properties of the task itself. In contrast to other studies, participants were not required to integrate their own beliefs with those of others, but instead only formed

beliefs associated with the (social or non-social) predictor. Further, our paradigm also allowed us to compare binary and single option decisions and whether they differed when they were being made about social or non-social predictors. Although we found no differences on either behavioural or neural levels between social and non-social conditions during binary decisions (Chapter 2,3), we show that single option decisions differed on both levels between conditions. When judging their confidence about the predictor's performance, participants were more confident about the performance of social predictors. By contrast confidence in non-social predictors was more influenced by one particular belief parameter -- its subjective uncertainty. We found common brain areas that were active when representing confidence in both social and non-social predictors, such as perigenual anterior cingulate cortex (pgACC), ventral striatum (VS) and inferior parietal lobe (IPL). However, the representation of confidence in social predictors was additionally represented in right posterior TPJ.

4.2. Methods

4.2.1. Social and non-social task versions

Participants performed a social and non-social version of the experimental task that was described in Chapter 2. Both versions differed in their framing but were matched in statistical properties. Participants performed both task versions and their order was counterbalanced across participants. First, we analysed the effect of the social vs non-social framing on selections between predictors during the initial decision phase of the trial. Next, we analysed the confidence phase to investigate behaviour and neural activity when participants' considered accuracy

and uncertainty associated with a single predictor. We examined potential differences relating to whether this predictor was a social agent (allegedly a previous participant who had knowledge of the task in the social condition) or a non-social cue (non-social condition).

In the social version, faces represented advisors that predicted a target position (a flower's position on the circumference of a circle). Some advisors predicted accurately and others less accurately. Participants could maximize the number of points they earned by identifying the most accurate advisors and following their advice about where the flower was likely to appear. Performance was defined by the size of an angular error, which represented the distance between the location of the predicted flower position (represented by a black dot) and true position (the target position – represented by a yellow dot) (Figure 1A). In the non-social version, participants were instructed to collect fruits (this was now the target position and again it was represented by a yellow dot) that could either fall a small or large distance from the tree (this was now the predictor and again it was represented by a black dot) (Figure 1B). In both versions, the aim was to select the most accurate predictor that would be the smallest distance from the target. In a second stage of each trial, participants indicated their confidence in their choice by setting a confidence interval around the predicted location. Because participants could maximize their points by setting the confidence interval narrowly during the second stage of this part of the trial there was an incentive to pick the best predictor/advisor. For more details about the experimental task, please see Chapter 2.



Figure 1: Social and non-social experimental versions. Participants performed a social and non-social version of the experimental task described in Chapter two. **(A)** Framing of social condition: participants selected between advisors that predicted the target location, represented by a flower, accurately (small angular error) or less accurately (big angular error) on the circle. **(B)** Framing of non-social condition: participants selected between fruits that either fell with a small distance (small angular error) or big distance (big angular error) from the tree. Note that the key statistical relations between predictors and targets in the social and non-social trial design were identical and only the types of pictures changed. For more information on the details of the experimental task, please see Chapter 2.

4.2.2. Framing of social task version

During the experimental task, participants could not learn about the flower location themselves, but instead had to make a binary selection between predictors (decision phase) to receive information on the target location (confidence phase). In the social condition, predictor stimuli represented advisors who had knowledge about the location of the target (represented by a flower). Participants were instructed that advisors represented previous players who had already performed a behavioural task during which they had the opportunity to learn about the

distribution of the target's location (Figure 1C). During this prior behavioural task, participants were told that the other players had learnt to predict the location of the target through trial and error across time. On each trial, these other players had first selected the target (it was always the same target) and then indicated the location in which they expected the target to appear. During the outcome phase, the other players were able to update their beliefs about the target's distribution by comparing the distance between the predicted location and the true location. This distance is defined as angular error. The aim was to predict the target as accurately as possible across trials, i.e. to minimize the angular error. Participants in the social condition were instructed that when they picked an advisor to help them predict the target location then that advisor would retain the set of trial-by-trial angular errors that they had previously made for the predictions they made. Therefore, in the main task, participants were incentivized to try and identify an advisor who was more accurate in their predictions and they were able to do this by observing the angular errors associated with their predictions (Figure 1A).

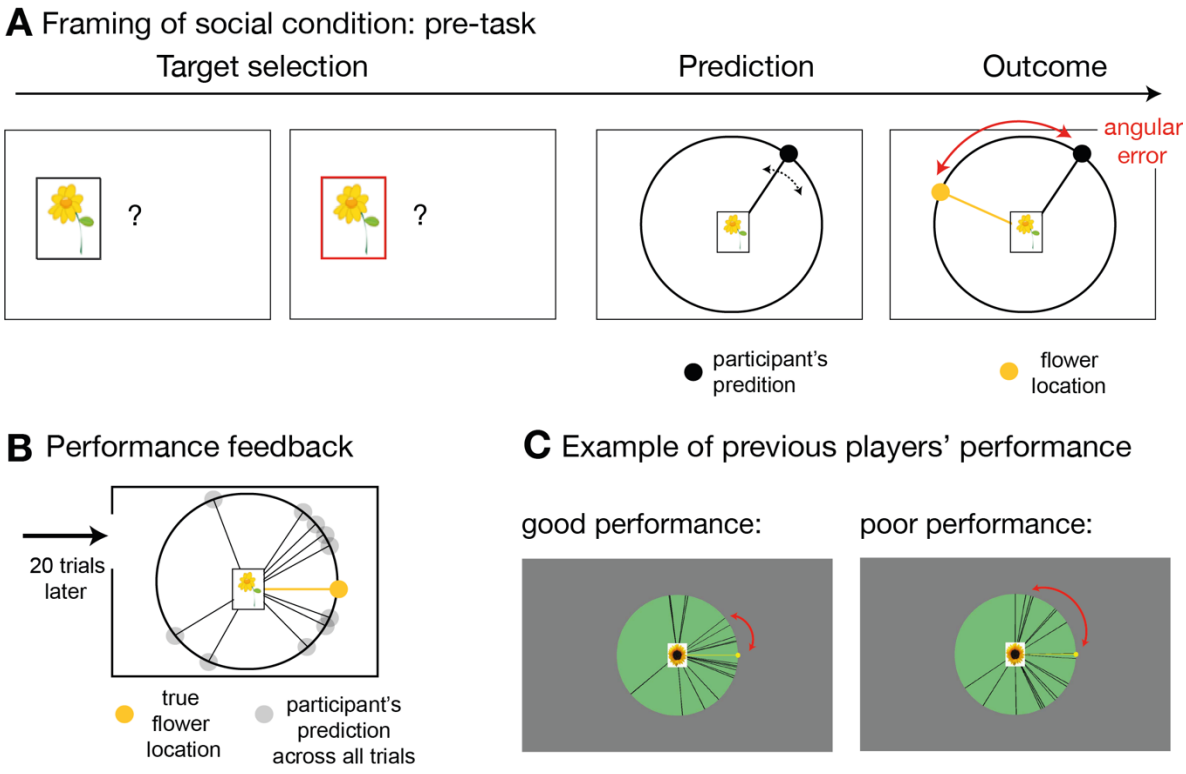


Figure 2. Framing of social condition. Instructions for the social condition which were given prior to task performance. The cover story to the social conditions asserted that predictors' estimates of the target reflected previous participants' beliefs about the target locations in a slightly modified version of the task. To make this believable, we designed such a modified task version and participants performed this task version themselves as part of the instructions to the social task condition. **(A)** Modified task version used as the pre-task in the social condition. Participants performed 20 trials of the pre-task prior to their own experiment. On every trial of the pre-task, participants made predictions about the flower location and subsequently observed the true flower location. Over time, they were able to form more precise estimates about the flower location. However, due to significant noise in the true location, estimates still remained erroneous. The current trial could for example underlie the trial shown during the main text depicted in (Figure 1A). **(B)** At the end, participants viewed their overall performance. **(C)** In addition, they were shown similar summary figures of pre-task performance from someone who performed well (left panel) or poorly (right panel) in predicting the target location. Individual differences amongst participants were believable as locating the target on each trial was a challenging task in itself. The purpose of the pre-task and subsequent explanation used in the social condition was to make it believable that previous participants, who might now be chosen as advisors in the main task, might have performed well or poorly and therefore might now be more or less accurate predictors in the main task.

Note that the actual experiment comprised sets of good and bad predictors and therefore, for the social version, it was important to make it plausible that people might differ considerably in their ability to estimate the true target location. Therefore, to emphasize the existence of individual differences in advisors'

performance, participants performed 20 trials of a different behavioural task prior to the main experimental task. We refer to this as the pre-task. The pre-task rules were identical to those of the task that participants were told that the other players (who subsequently were offered as advisors) had performed. By performing the task, we were able to ensure that participants had a clear sense of why some advisors might be better to choose than others. At the end of the pre-task, participants were presented with their own performance (Figure 1C-ii), the performance of someone who performed well (Figure 1C-iii, left panel: on average, small angular errors) and the performance of someone how had performed poorly (Figure 1C-iii, right panel: on average, big angular errors). It was then clear that it was very reasonable that there might well be individual differences in performance. Note that for the non-social condition, no comparable pre-task was used during the instruction.

4.3. Behavioural analyses

We applied a set of general linear models (GLM) to the behavioural data. All GLM analyses were applied to both social and non-social versions. We used two-tailed statistical tests for all analyses. For all GLM analyses, regressors were normalised with mean = 0 and standard deviation =1 for the set of trials used in the respective analyses.

4.3.1. Decision phase

First, we repeated previous analyses from chapter 2 to test the impact of trial-wise Bayesian-derived belief estimates of accuracy and uncertainty and their modulations across time on the selection between predictors (advisors or fruits) in

the decision phase (Figure 2). We used the same analyses as outlined in Chapter 2: Choice GLM1 (selections across all trials), Time GLM1 (selections per block halves) and Time GLM2 (selections as function of initial trials during time horizons). For details on regressors, normalisations and correlations between regressors, please see Chapter 2, section 2.2.6.1.

4.3.2. Confidence phase

Next, we were interested whether there were behavioural differences when evaluating one's own confidence about a single information source. We tested whether confidence judgments were impacted by model-based estimates of accuracy and uncertainty, and whether there were additional effects of variables that could be observed directly within the experiment. We also controlled for the starting size of the interval, which was set randomly at each trial. We used a linear GLM with the following regressor to predict participants' trial-by-trial confidence judgments:

Confidence GLM 1, across all trials (Figure 3A):

1. Chosen accuracy,
2. Chosen uncertainty,
3. Payoff at the last encounter of the same predictor, i.e. t-1
4. Confidence judgment at the last encounter of the same predictor, i.e. t-1,
5. Unchosen accuracy,
6. Unchosen uncertainty,
7. Initial size of confidence interval,

Only regressors of interest (1-6) are shown in Figure 3A. Note that confidence judgments were indicated by modifying the interval size around the chosen

predictor with a smaller interval representing higher confidence. To make this measure intuitive, we sign-reversed their relationship such that a higher confidence index represents greater confidence in the chosen predictor (see 'inverse measure' on y-axis). Note also that "chosen" and "unchosen" in the regressor description relates back to the decision phase; the chosen predictor was always the one that had been used to make the prediction that was now being considered in the confidence phase and participants were supposed to indicate how close the chosen predictor's estimate would lie near the true target location.

We observed that participants confidence judgements about both the social and non-social predictor were partly a function of the last confidence judgments they had made at the last encounter with the same predictor. We were interested whether this impact was restricted to the last encounter or whether instead, participants formed a representation that was based on multiple encounters in the past, i.e. whether they formed a confidence judgment based on a longer timescale. More specifically, we tested whether the past judgments had a recency-weighted impact on the current confidence judgment. We fitted a linear GLM to the current (t) confidence judgment as a function of judgments made on the last four encounters (t-1,t-2…) (Figure 4B):

Confidence GLM 2, Figure 7A:

Confidence judgment at t-1

Confidence judgment at t-2

Confidence judgment at t-3

Confidence judgment at t-4.

Note that we ran an additional analysis that controlled for variables in GLM1 by adding the remaining 6 regressors from that GLM into the model. The results, however, remained the same regardless of whether or not we controlled for these additional variables.

We showed that the current confidence judgment is impacted by previous judgments in a recency-weighted fashion (Figure 7A). We used a reinforcement learning (RL) model to obtain a single estimate of participant's confidence judgment at the current trial as a function of recency-weighted past judgments (Figure 7B). For every participant, we fitted a standard RL model to the history of confidence judgments to estimate the expected confidence judgment at the current trial. On every trial t, the estimated confidence judgment associated with the selected predictor was derived as follow:

Confidence estimate $_{(t+1,\ predictor)}$ = confidence estimate $_{(t,\ predictor)}$ + α * [confidence judgment $_{(t,\ predictor)}$ - confidence estimate $_{(t,\ predictor)}$]

(1)

The learning rate, α, represents the rate of change of the confidence judgements. A small learning rate indicates more stable confidence judgements in which distant confidence judgements are informative of current confidence judgements. In contrast, a large learning rate indicates that only more recent confidence judgements are informative for current ones.. Note that the RL model does not take into account the true target location and hence does not reflect

observed changes in the external environment; rather, it reflects the degree of trial-to-trial changes that ensue in the participants' own confidence judgements.

We used least square fit as an error function to fit the estimated past confidence judgment to the current confidence judgment:

Error function = sum((Confidence Estimate$_{(t-1,\ predictor)}$ - confidence judgment$_{(t,\ predictor)}$)^2)

(2)

4.4. FMRI whole-brain analyses

We used FSL FEAT for first-level analysis (Smith et al., 2004). First, data was pre-whitened with FSL FILM to account for temporal autocorrelations. Temporal derivatives were included into the model. We used one GLM to analyse fMRI data across the whole brain. fMRI-GLM1 was applied to all trials. Results were calculated using FSL's FLAME 1 with a cluster-correction threshold of $z>3.1$ and $p<0.05$, two-tailed.

We applied second-level analyses to (1) social and non-social versions separately (FLAME 1) and (2) to the difference between social and non-social versions (FLAME 1). We used a two-step approach to average neural activation across both conditions: on second-level analysis, social and non-social versions were averaged across conditions within participant (fixed effect-analysis). The resulting values were then averaged across participants on third-level analysis (FLAME 1).

We included all three phases of a trial (decision, confidence and outcome) into the fMRI-GLM1. For each phase, all regressors were normalised before inclusion into the analysis. The decision phase began at the time the predictor appeared and lasted until a selection was made by the participant (Figure 1A). The decision phase was modelled as a constant and was accompanied by the following parametric regressors:

fMRI-GLM1, decision phase:

accuracy prediction difference, i.e. chosen-unchosen accuracy

uncertainty prediction difference, i.e. chosen-unchosen uncertainty.

The confidence phase was defined from the onset of the circle presentation (Figure 1A) until a confidence judgment was made. It included a constant and the following parametric regressors:

fMRI-GLM1, confidence phase:

global confidence,

confidence shift,

contrast: global confidence + confidence shift = confidence judgment.

Note that the confidence judgment variable was sign-reversed, hence a positive neural effect can be interpreted as representing greater confidence. In the first analyses, we were interested in whether the confidence judgment could be separated into different components: the *global confidence* represents a long-term summary the confidence interval associated with a given predictor. It is calculated as an average of all previous confidence judgments for a given predictor up until the current trial. The *confidence shift* is calculated as the difference between the global confidence and the confidence judgment selected on the current trial. Hence, this allows us to derive a contrast between both variables that represents the trial-by-trial confidence judgment. For clarity, results relating to the global confidence and confidence shift will not be discussed here. The main regressor of interest is the trial-by-trial confidence judgment that can be obtained from the contrast between these two regressors.

The onset of the outcome phase was defined by the presentation of the target and payoff and lasted for a fixed duration of three seconds. In addition to the constant regressor, we included the following parametric regressors:

fMRI-GLM1, outcome phase:

chosen accuracy,

chosen uncertainty,

payoff magnitude.

4.5. FMRI region of interest analyses

We calculated ROIs with a radius of three voxels that were centred on the peak voxel of significant clusters derived from the whole brain fMRI-GLM1 averaged across social and non-social conditions or based on peak-coordinates of previous studies or based on anatomical ROIs. For all cases, the selected ROI was transformed from MNI space to subject space and then pre-processed BOLD time courses were extracted for each participant's social and non-social versions. Time courses were averaged across volumes, then normalized and oversampled by a factor of 20 for visualisation. Time courses were time-locked just prior to the confidence judgment (at 70% of the trial wise confidence phase duration; this was determined by other previous analyses). When applicable, we obtained an average across conditions by first averaging beta-weights within participant across sessions and then averaging beta-weights across participants (thus the approach was analogous to the approach used in the whole brain procedure).

A leave-one-out procedure was used to test significance of the time courses extracted from all ROI analyses without bias. For every participant (n = 24), we extracted the average time course based on the 23 remaining participants. We identified the peak of the group time course in a time window between 4-8 seconds and extracted the beta value for the excluded subject at the time of the group peak. This procedure was repeated for all participants which resulted in individual peak

values that were independent from the subject to be analysed. The extracted peak values were tested with a one-sample t-test against zero.

4.5.1. Common neural correlates of confidence for social and non-social predictors

First, we identified the brain areas in which activity correlated with participants' confidence judgment, irrespective of the nature of the stimuli. We used a time course analysis to illustrate the representation of confidence in brain areas that reached cluster-corrected significance in the whole-brain fMRI-GLM1. For all analyses, we included all three phases. When analysing the decision phase we used the regressors already listed in fMRI-GLM1. When we analysed the confidence phase, we included the following regressors, additionally controlling for the number of button presses:

ROI GLM1, confidence phase (across all trials), Figure 4C:

Confidence judgment,

Number of button presses.

Next, we were interested in whether brain areas that represented confidence judgments during the second stage of a trial had a similar representation during the outcome phase. We included the following regressors:

ROI GLM1, outcome phase (across all trials), Figure 4C:

Confidence judgment,

Binary payoff (1=rewarded, 0= not rewarded).

Effects of ROI-GLM1 were extracted from three brain regions that reached significance on the third-level analysis when social and non-social conditions were averaged in fMRI-GLM1. An lPLD ROI was based on the cluster-significant peak-

coordinate from the whole-brain analysis associated with the parametric confidence judgment effect from fMRI-GLM1 (MNI x/y/z-peak coordinates: [-48, -54 38]). In addition, we used a pgACC ROI that was based on a peak-coordinate from a previous study that investigated learning about one's own and other peoples' performances (MNI x/y/z-peak coordinates: [0, 40, 6]; (Wittmann et al., 2016b)). Finally, we used an anatomical region derived from the Harvard Subcortical Structures atlas encompassing bilateral ventral striatum.

4.5.2. Dissociable neural correlates of confidence for social and non-social predictors

Next, we tested whether there were brain areas that represented confidence judgments only for social predictors (as opposed to both social and non-social predictors). There were no whole-brain cluster corrected differences between groups. However, activation in posterior TPJ correlated with confidence judgment for social predictors on the whole-brain level. We tested whether this activation was significantly different from the non-social condition when applying a ROI-correction to right pTPJ. We used a 90% probability mask centred on right pTPJ (Mars et al., 2013). We applied fMRI-GLM1 with a volume correction for ROIs of $z>2.3$ and $p<0.05$, two-tailed (Figure 5, left panel). To illustrate the difference between social and non-social groups, we extracted a time course from right pTPJ ROI and applied the following GLM:

ROI-GLM2, across all trials (Figure 6B)

Confidence judgment,

Number of button presses.

Next, we tested whether pTPJ's confidence representation could be disentangled into individual variables that influenced the behavioural confidence setting:

ROI-GLM3, across all trials (Figure 6)

Chosen accuracy,

Chosen uncertainty,

Payoff t-1,

Confidence judgment t-1,

Number of button presses.

Note that because of the way in which the analysis was conducted, which included the use of a confidence shift regressor at an early state in the analysis, trials where there had been no previous encounter with a given predictor were excluded from the analysis because they could not be described with this approach.

In a final analysis, we tested whether pTPJ represented confidence on a longer timescale compared to the current or last encounter. We used a recency-weighted confidence judgment estimated by the RL to test whether pTPJ represents this long-term confidence judgment:

ROI-GLM4, across all trials (Figure 7)

Chosen accuracy,

Chosen uncertainty,

Payoff t-1,

Recency-weighted confidence judgment (derived from the RL model),

Number of button presses.

4.6. Behavioural results

4.6.1. Effects of uncertainty and accuracy on predictor selections

We repeated behavioural analyses of Chapter 2 which focused on the impact of Bayesian-derived model estimates of accuracy and uncertainty and their change over time in a block on predictor selection during the decision phase. In Chapter 2, we have shown that participants' predictor selections were informed by their beliefs about their accuracy and the uncertainty of those beliefs. Further, effects of accuracy and uncertainty on predictor selections were modified by time: selections were directed towards uncertain predictors in earlier compared to later trials in the block, while selections towards the end of a block were of accurate predictors (Figure 3A, B left panel). The behavioural impact of accuracy and uncertainty estimates were modulated as a function of how many trials remained within a block, i.e. the remaining time horizon: selections were uncertainty-driven when there was more time remaining (i.e. long horizon), while selections converged quickly towards accurate predictors when there was less time (i.e. short horizon) (Figure 2B, right panel). These results were replicated for both the social and non-social condition, while we did not find any behavioural differences between social and non-social conditions in the effects of accuracy and uncertainty on predictor selections across all trials, across block halves or as a function of time horizon (respectively, Figure 3A-C; Table 1).

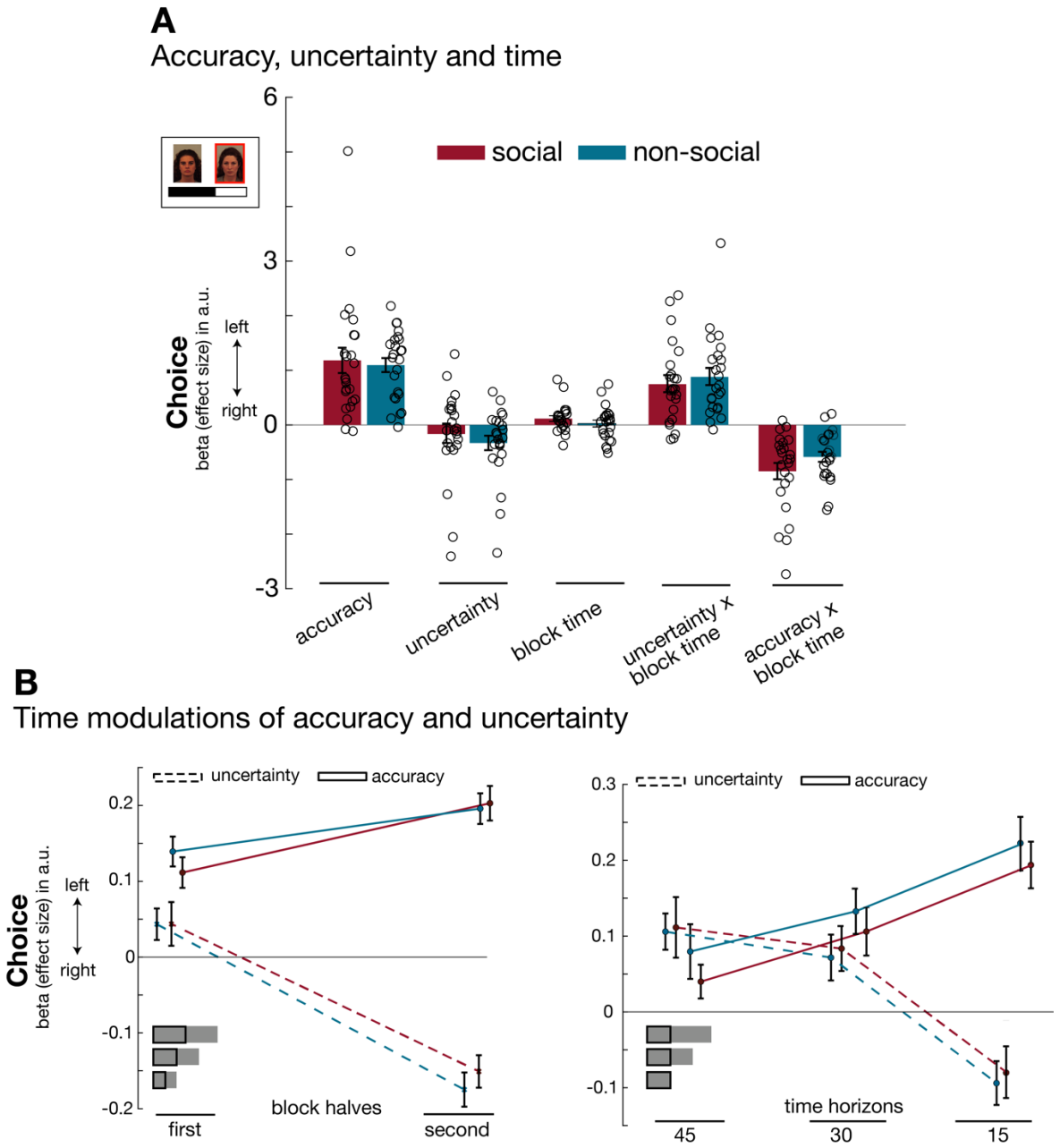


Figure 3. Decision phase: effects of Bayesian-derived estimates of accuracy and uncertainty on predictor selections during the decision phase. *There were no differences between social and non-social conditions in the impact of accuracy and uncertainty belief estimates on predictor selections across all trials* ***(A****), across block halves* ***(B, left panel)*** *or as a function of time horizon* ***(B, right panel).*** *(n=24, error bars are SEM across participants)*

Analysis	Regressors	Statistical values
All trials: paired t-test between social and non-social conditions for respective regressor (Figure 2A)	Accuracy	t(23)=0.4;p=0.7;d=0.08; 95% confidence interval = [-0.36 0.5]
	Uncertainty	t(23)=1;p=0.3;d=0.21; 95% confidence interval =[-0.16 0.5]
	Uncertainty x block time	t(23)=-1.24;p=0.22;d=-0.25; 95% confidence interval = [-0.35 0.09]
	Accuracy x block time	t(23)=-1.7;p=0.1;d=-0.34; 95% confidence interval = [-0.6 0.066]
Block halves: 2x2x2 repeated-measures ANOVA (Figure 2B)	2 (variable: accuracy, uncertainty) x 2 (block halves: first, second) x 2 (group: social vs non-social)	2x2x2 interaction: F(1,23)=0.14; p=0.7; η^2=7.254e-5

Time horizon: 2x3x2 repeated-measures ANOVA (Figure 2C)	2 (variable: accuracy, uncertainty) x 3 (time horizon: long, medium, short) x 2 (group: social vs non-social)	2x3x2 interaction: $F(1,23)=0.52$; $p=0.6$; $\eta^2=0.002$

Table 1. *Results of statistical tests applied to the decision phase, Figure 3.*

4.6.2. Confidence judgments about social and non-social predictors

Next, we were interested in assessing whether one's own confidence about the predictor's performance is different when the predictor is of social or non-social nature. This kind of decision differs from the binary choice made during the decision phase: instead of comparing both predictors and selecting between them, now participants evaluate the performance of a single predictor independent of an alternative. Arguably, the way social neuroscience experiments are conducted usually requires participants to consider only one other individual at a time, for example during mentalizing tasks when the beliefs or intentions of a single person are assessed (Saxe, 2006), for example when predicting another person's performance given their past performance (Wittmann et al., 2016b), or assessing another player's intentions in a strategic game (Hill et al., 2017).

We tested whether estimates of accuracy and uncertainty that determined participants' selections between predictors, also contributed to the confidence judgment made when evaluating a single predictor (Figure 4B). Accuracy reflects a point-estimate of the most likely angular error between target and the predictor's

estimate and should therefore have an impact on the interval size the participants use to indicate their subjective confidence during the confidence phase. Additionally, our Bayesian model suggests that participants not only form a belief about the most likely angular error – a point-estimate of the angular error, but they also form a representation of their uncertainty in this belief – a representation of the width of the distribution over angular errors. If participants are very uncertain in their point-estimate of the angular error (i.e. if the Bayesian belief distribution is very wide), then they should report a larger interval size to guarantee that the target falls within the interval. Indeed, we found that participants were more confident (i.e. selected a smaller interval size) when they believed the selected predictor was accurate, while participants were less confident (i.e. selected a bigger interval size) when they were uncertain about the predictor's accuracy (see Table 2 for statistical results of each group). There was no credible evidence that accuracy impacted confidence judgment differently between social and non-social conditions (paired t-test, $t(23)=-1.8$, $p=0.09$, $d=-0.37$, 95% confidence interval=[-0.26 0.02]). However, when making confidence judgments about non-social compared to social predictors, participants were more influenced by their uncertainty in the predictor's accuracy (paired t-test, $t(23)=2.2$, $p=0.04$, $d=0.45$, 95% confidence interval=[0.01 0.3]).

We also tested whether information about the previous encounter with the same predictor influenced participants' current confidence judgments. For example, whether or not reward is received on any trial is determined by whether the target location is included within the confidence interval the participant sets, while the reward magnitude is dependent on the interval size (larger rewards were

given for smaller interval sizes). Indeed, the reward participants received on the last encounter influenced their next confidence judgment: participants became more confident about the predictor's performance and selected a smaller interval size the larger the reward was on the last encounter. There was no evidence that this effect differed between social and non-social conditions (paired t-test: t(23)=0.16, p=0.88, d=0.03, 95% confidence interval=[-0.1 0.1]). Another predictor of participants' confidence judgment should be their judgments made on the previous encounter: if participants were confident about the predictor's performance on the last encounter, then it is likely that they are also confident on the current trial. In fact, previous judgment was the strongest predictor of the current confidence for both social and non-social conditions (compare beta weights across effects; see Table 2 for statistical results for individual groups; paired t-test: t(23)=-1.6, p=0.1, d=-0.33, 95% confidence interval=[-0.29 0.03]).

Differences became apparent when comparing the constant of the model, which represents participants 'default' confidence judgment. Participants were more confident and selected a smaller interval when making judgments about social compared to non-social predictors (Figure 4A ; paired t-test, constant: t(23)=-2.1, p=0.046, d=-0.43, 95% confidence interval=[-0.78 -0.01]).

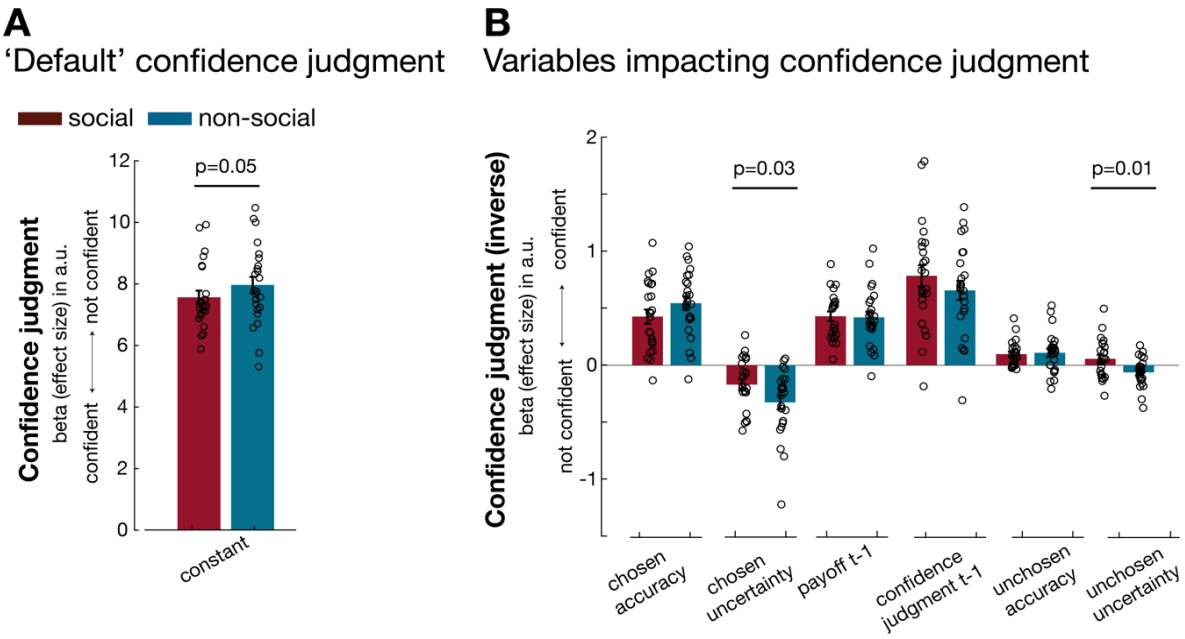


Figure 4. Evaluating a single agent: confidence judgments about the performance of social and non-social predictors. *Note that all beta weights in (A) and (B) are from the same GLM 1. Note the difference in visualisation on the y-label.* ***(A)*** *Participants are more confident (smaller confidence interval size) when judging reporting their judgments about social compared to non-social predictors.* ***(B)*** *For both social and non-social conditions, participants' set their confidence as a function of model-derived estimates of predictor accuracy and uncertainty and variables observed on the last encounter. Participants were in general more confident about the performance of social compared to non-social predictors (left panel), while confidence judgments about non-social predictors were more impacted by the uncertainty in predictor's performance. (n=24, error bars are SEM across participants)*

We tested whether participants' confidence judgments were impacted by the context in which the current confidence judgement occurred. Specifically, we examined whether information associated with the forgone predictor influenced how participants judged their confidence about the selected predictor. In terms of task requirements, information relating to the foregone predictor should be discarded after the decision phase as it is not relevant to estimating the performance of the single predictor under consideration in the confidence phase. We were in particular interested in model-derived estimates of accuracy and uncertainty that influenced predictor selections. We observed that information

about the forgone predictor influenced confidence judgments: participants were more confident about the selected predictor and selected a smaller interval if the predictor they rejected was believed to be accurate (see Fig 4B and Table 2 for statistical tests per condition; paired t-test: t(23) =-0.25, d=-0.05, p=0.8, 95% confidence interval= [-0.11 0.08]). However, only when judging confidence about non-social predictors, participants were additionally influenced by the uncertainty associated with the rejected predictor: confidence in the selected predictor decreased when participants were uncertain about the rejected predictor (see Table 2 for statistical tests per condition; paired t-test: t(23) =2.8, d=0.56, p=0.01, 95% confidence interval= [0.03 0.21]).

In summary, the reports that participants made about their confidence in the predictor's performance by setting confidence intervals could be predicted by a combination of model-derived estimates of what the participants' beliefs about their accuracy and their uncertainty about those beliefs, observations that the participants had made in previous encounters with the same predictors, and by their previous confidence judgments. Confidence judgments were not only determined by information related to the selected predictor, but also by beliefs associated with the forgone predictor. The influence uncertainty exerted on confidence judgments associated with the selected and rejected predictor was stronger when judging the performance of non-social compared to social predictors. Consistent with this observation, participants were less confident when judging non-social compared to social predictors.

Regressor	Condition	Statistical values
Chosen accuracy: one sample t-test	Social	t(23)=6.8;p<0.0001;d=1.3; 95% confidence interval = [0.29 0.55]
	Non-social	t(23)=9.1;p<0.0001;d=1.8; 95% confidence interval = [0.42 0.67]
Chosen uncertainty: one sample t-test	Social	t(23)=-3.6;p=0.001;d=-0.7; 95% confidence interval = [-0.27 -0.074]
	Non-social	t(23)=-5.4;p<0.0001;d=-1.1; 95% confidence interval = [-0.45 -0.2]
Payoff, t-1: one sample t-test	Social	t(23)=10.2;p<0.0001;d=2.1; 95% confidence interval = [0.34 0.51]
	Non-social	t(23)=8;p<0.0001;d=1.6; 95% confidence interval = [0.31 0.52]
Confidence judgment, t-1: one sample t-test	Social	t(23)=8.3;p<0.0001;d=-1.7; 95% confidence interval = [-0.1 -0.59]

	Non-social	t(23)=-7.9;p<0.0001;d=-1.6; 95% confidence interval = [-0.83 -0.48]
Unchosen accuracy: one sample t-test	Social	t(23)=4.2;p=0.0003;d=0.86; 95% confidence interval = [0.05 0.14]
	Non-social	t(23)=2.9;p=0.008;d=0.39; 95% confidence interval = [0.03 0.18]
Unchosen uncertainty: one sample t-test	Social	t(23)=1.5;p=0.14;d=0.31; 95% confidence interval = [-0.02 0.13]
	Non-social	t(23)=-2.5;p=0.02;d=-0.51; 95% confidence interval = [-0.12 -0.01]

Table 2. *Results for linear GLM applied separately to social and non-social confidence judgments displayed in Figure 4A. All effects derived from confidence GLM1.*

4.6.3. Changes of mind in confidence judgments are more frequent for non-social predictors

As a final analysis of the confidence phase, we examined the series of button presses participants made to perform the confidence judgement. On each trial, the confidence interval started at a random position and by pressing two buttons participants could increase or decrease the interval until it reflected their

final confidence estimate. This allowed us to analyse whether there were trials on which participants changed their minds just prior to the final confidence judgment. Changes of mind (CoM) were defined by a specific two-button-press sequence before the final confirmatory button press (see inset in Figure 5A): participants changed the interval size from first being more confident (i.e. decrease in interval) to less confident (i.e. increase in interval) prior to the confirmatory button press. First, we tested whether there was an increase in reaction time for the last button press in trials (see inset; duration between the end of button press 1 to the end of button press 2) where CoM occurred in comparison to trials where there was no CoM. This was indeed the case: changes in confidence required longer deliberation time (Figure 5A; paired t-test for CoM vs no-CoM trials: social: $t(23) = 7.1$, $p<0.0001$, $d=1.45$, 95% confidence interval =[0.24 0.44]; non-social: $t(23) = 6.5$, $p<0.0001$, $d=1.3$, 95% confidence interval =[0.23 0.45]; social vs non-social paired t-test between differences: $t(23)=0.07$, $p=0.95$, $d=$, 95% confidence interval=[-0.12 0.12]). Next, we tested whether changes in confidence throughout the trajectory of the phase led to differences in the final confidence judgment. We did not observe differences in the final size of the interval compared to other trials (social: $t(23) = -0.81$, $p=0.43$, $d=-0.17$, 95% confidence interval =[-0.47 0.21]; non-social: $t(23) =0.43$, $p=0.67$, $d=0.09$, 95% confidence interval =[-0.281 0.43]; social vs non-social paired t-test between differences: $t(23)=-0.8$, $p=0.43$, $d=-0.17$, 95% confidence interval=[-0.74 0.32]).

In the previous analysis of confidence, we showed that confidence was lower when participants were uncertain about the chosen predictor (Figure 4B) and this was particularly the case in the non-social condition. If indeed uncertainty

affects confidence more in the non-social condition, this might also be reflected by an increase in the number of changes of mind in the time period leading up to the confidence judgement in the non-social condition. Therefore, we tested whether the impact of uncertainty was related to more within-trial changes of mind for non-social predictors. We found that participants indeed changed their mind more frequently when judging the performance of a non-social compared to a social predictor in the period before entering their final judgment (Figure 5B, left panel: paired t-test, t(23) = -2.4, p=0.025, d=-0.49, 95% confidence interval = [-6.3 -0.46]). This frequency was correlated with the impact uncertainty exerted on the final confidence judgment: participants who had a stronger impact of uncertainty on confidence judgment, also changed their mind more frequently (Figure 5B, right panel: non-social: r=0.46, p=0.024). This relationship was absent for social predictors (social: r=-0.0004, p=0.99).

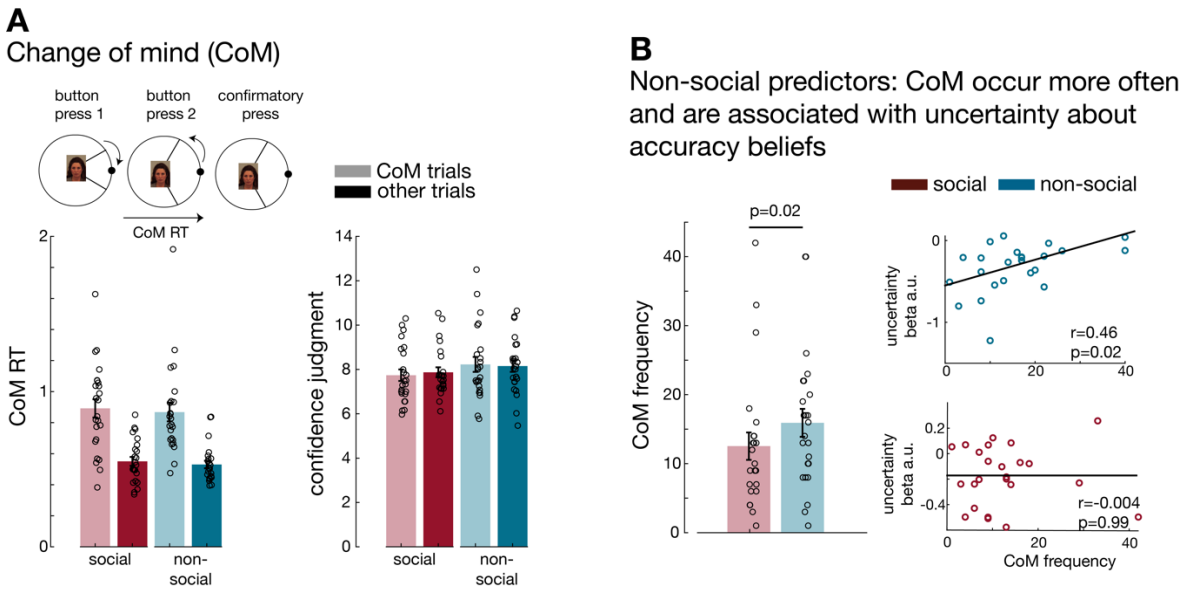


Figure 5. Changes of mind: within-trial changes in confidence judgment. (A) *Upper illustration: a change of mind was captured by a specific sequence of button presses prior to the confirmatory button press. Change of minds (CoM) were defined by a specific sequence of the two last button presses (from button press 1 to press 2) before confirming the judgment: participants first decreased the interval (button press 1: more confident) and then again*

increased the interval (button press 2: less confident). The time participants took from the end of button press 1 to the end of button press 2 is defined as the CoM reaction time (RT). Left panel: we compared the reaction time between button press 1 and button press (see inset) on trials where there was a change of mind between these two button presses compared to trials where there was no change of mind. Participant's reaction time was slower on trials where they changed their mind for both social and non-social conditions (light bar). Right panel: the final confidence judgment did not differ between CoM trials and trials where there were no CoM. ***(B)*** *Left panel: CoM trials occurred more often when judging one's confidence about non-social compared to social predictors. Right panel: the frequency of CoM trials correlated with the impact uncertainty exerted on the final confidence judgment (regression weight for chosen uncertainty from GLM shown in Figure 3) for non-social but not social predictors. (n=24, error bars are SEM across participants)*

4.7. Neural results

In the task, we are differentiating between two kinds of decisions: a binary selection between predictors and a confidence judgment about the selected predictor. In previous chapters, we have shown that there were no differences between social and non-social conditions at a neural level when making the first type of decisions in our experiment. Value-based decisions and confidence judgments are fundamentally different processes and can be differentiated as such at a neural level (De Martino et al., 2012). Here, I will present findings that are related to the coding of confidence judgments at the second stage of the trial (fMRI-GLM1). The variable of interest was the confidence judgment participants indicated at each trial. Although we were interested in differences in the coding of confidence between social and non-social predictors, we expected to find neural substrates that were common to both conditions. We first analysed neural correlates of confidence irrespective of the nature of the stimuli by averaging across social and non-social conditions. Second, we examined brain areas that uniquely coded confidence judgments for social predictors.

4.7.1. Common neural correlates of confidence for social and non-social predictors

We tested which brain area represented the coding of confidence judgment irrespective of social and non-social conditions. We applied a whole-brain GLM averaged across both social and non-social predictors (fMRI-GLM1) and found activity in three brain areas that covaried parametrically with participant's confidence judgments: dorsal inferior parietal lobe (IPLd), perigenual anterior cingulate cortex (pgACC) and ventral striatum (VS) (Figure 6A). Activation in IPLd decreased with increasing confidence about the predictor's performance, while activation patterns in pgACC and VS showed the opposite pattern: their activation increased when participants were more confident in the predictor's performance (Figure 6A).



***Figure 6. Neural substrates common to both social and non-social confidence judgments**. **(A)** Activation in pgACC and ventral striatum (VS) increased (yellow colours) when participants became more confident and selected smaller confidence intervals, while signals in bilateral IPLd decreased with increasing confidence (blue colours). **(B)** In the outcome phase, both pgACC and VS represent payoff magnitude.*

In this task, the confidence interval set in the second stage of each trial determines whether a payoff will be received and what its magnitude might be during the outcome stage. We asked whether areas represented the confidence interval during the confidence phase because it was predictive of reward in the subsequent outcome phase. PgACC has been associated with the tracking of performance and its relation to predicting reward success (Bang and Fleming, 2018; Lockwood and Wittmann, 2018; Will et al., 2017; Wittmann et al., 2016b), while activity in VS covaries with reward and motivational processes in social and non-social contexts (Lockwood et al., 2018; Rilling et al., 2002; Will et al., 2017; Zhu et al., 2012). We examined brain activity correlated with the payoff magnitude during the outcome phase and found that both pgACC and VS activity increased with increasing payoff magnitude (Figure 6B). Next, we tested whether they covaried with different aspects of the payoff magnitude: either with whether or not any payoff was made, i.e. binary payoff variable (we refer to this as binary payoff) or confidence interval that determined the reward magnitude (Figure 7). We extracted an independent ROI for pgACC (Wittmann et al., 2016b) and an anatomical ROI derived from the Harvard Subcortical Structure atlas for bilateral VS. PgACC and VS showed different profiles in representing payoff magnitude and confidence judgment at the time of outcome (Figure 5C-i: middle and left panel; 2 (variable) x 2 (area) repeated measures ANOVA; interaction: $F(1,23)=72$, $p<0.0001$). In a follow-up test, we show that pgACC represented confidence judgment more strongly than the binary payoff variable (paired t-test: $t(23)=-2.4$, $d=-0.49$, $p=0.025$, 95% confidence interval = [-0.06 -0.005]), while VS represented the received the binary payoff more strongly than the confidence judgment (paired

t-test: $t(23)=-6.9$, $p<0.0001$, $d=-1.4$, 95% confidence interval=[-0.1 -0.06]. Both pgACC and VS might track confidence judgment due to its predictability of reward: both areas represent confidence judgment during the second phase and the (binary) payoff representation during the outcome phase. While pgACC additionally also represented one's confidence judgment during the allocation of outcome.

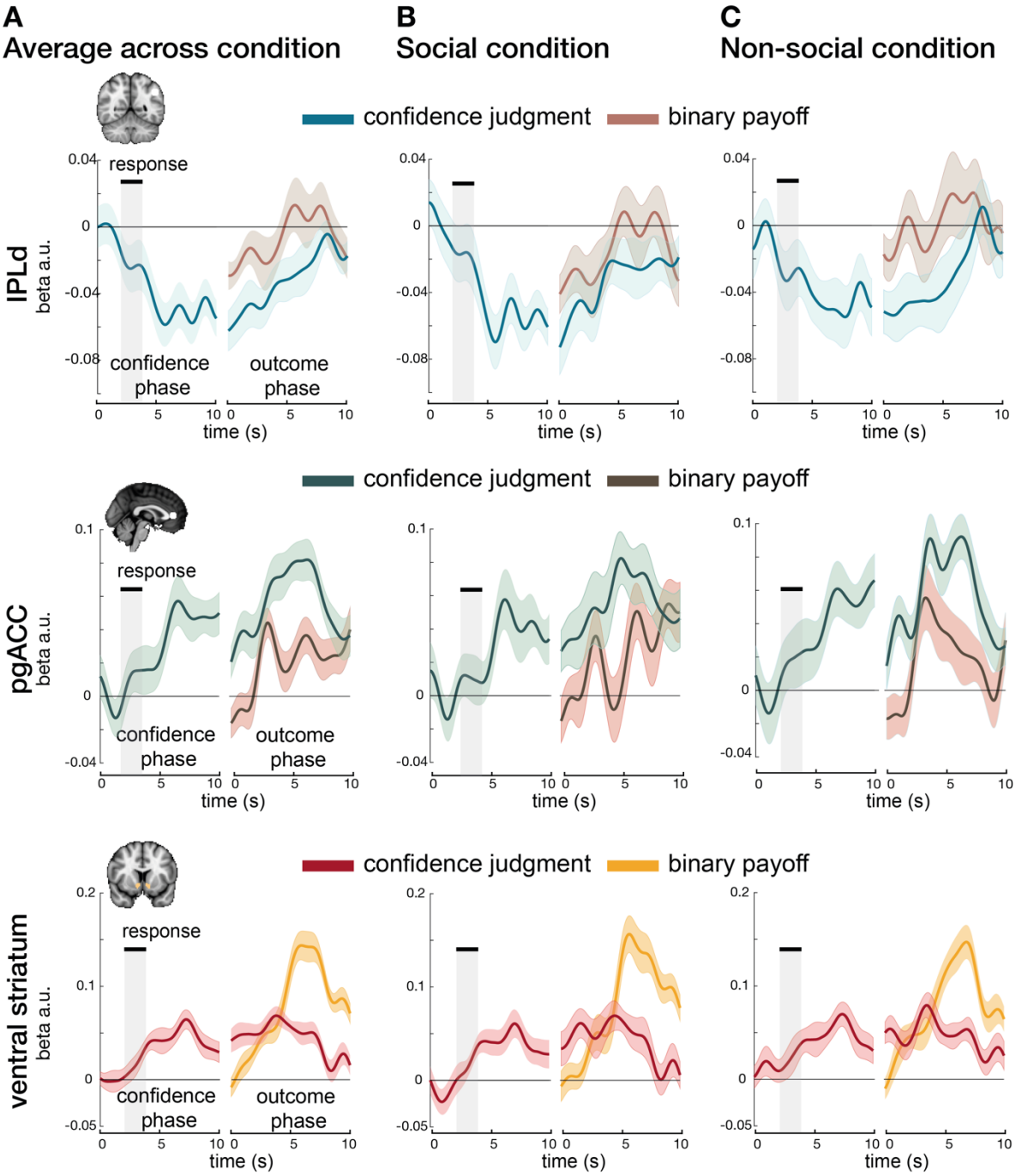


Figure 7. Time courses of common substrates across confidence and outcome phases*. In the outcome phase, both pgACC and VS represent payoff magnitude (see Figure 4). Payoff magnitude can in turn be decomposed into two components, first whether or not any payoff is delivered which defines a binary variable and second, how much reward was allocated with is dependent on the confidence interval set in the previous stage and defines the second component of payoff magnitude. We show time courses for three areas that survived cluster-correction during the confidence phase (IPLd, pgACC, ventral striatum). Each row shows activation for the same brain area, either as average across conditions* ***(A)****, or for social condition only* ***(B)*** *or for non-social condition* ***(C)****. The activation pattern was similar for social and non-social conditions. (n = 24; whole-brain effects family-wise error cluster corrected with z > 3.1 and p < 0.05).*

4.7.2. Unique social substrate: activation in posterior TPJ correlates with confidence judgment about social predictors

Next, we tested whether there were neural substrates that coded confidence judgments specifically when they were made about the performance of social predictors. Activation in right posterior temporal junction (pTPJ) survived cluster correction for social predictors, but not for non-social predictors (Figure 8A). The cluster covered the right pTPJ and extended dorsally into IPLd. Both pTPJ and IPLd showed a similar profile of confidence coding: activation in both areas decreased with increasing confidence about the social predictor (compare with Figure 6,7).

We selected an unbiased ROI to compare conditions in right pTPJ (Mars et al., 2013). Using small volume correction in this ROI, we found that activity levels differed between conditions in a ventral part of the right pTPJ (MNI x/y/z-peak coordinates: [46 -54 22]) (Figure 8A; green ROI-corrected cluster). Further analysis of activity in the ROI confirmed significant effects of confidence judgements in the social condition, an absence of such signals in the non-social condition, and a significant difference between both conditions (Figure 8B; paired t-test: $t(23) = 3.1$, $p=0.005$, $d=0.63$, 95% confidence interval=[0.02 0.08]; social: $t(23) = 3.9$, $p<0.0001$, $d=0.8$, 95% confidence interval =[0.024 0.08]; non-social: $t(23) =0.1$, $p=0.9$, $d=0.02$, 95% confidence interval=[-0.03 0.03]; see ROI-GLM2 in Methods for details on analysis).

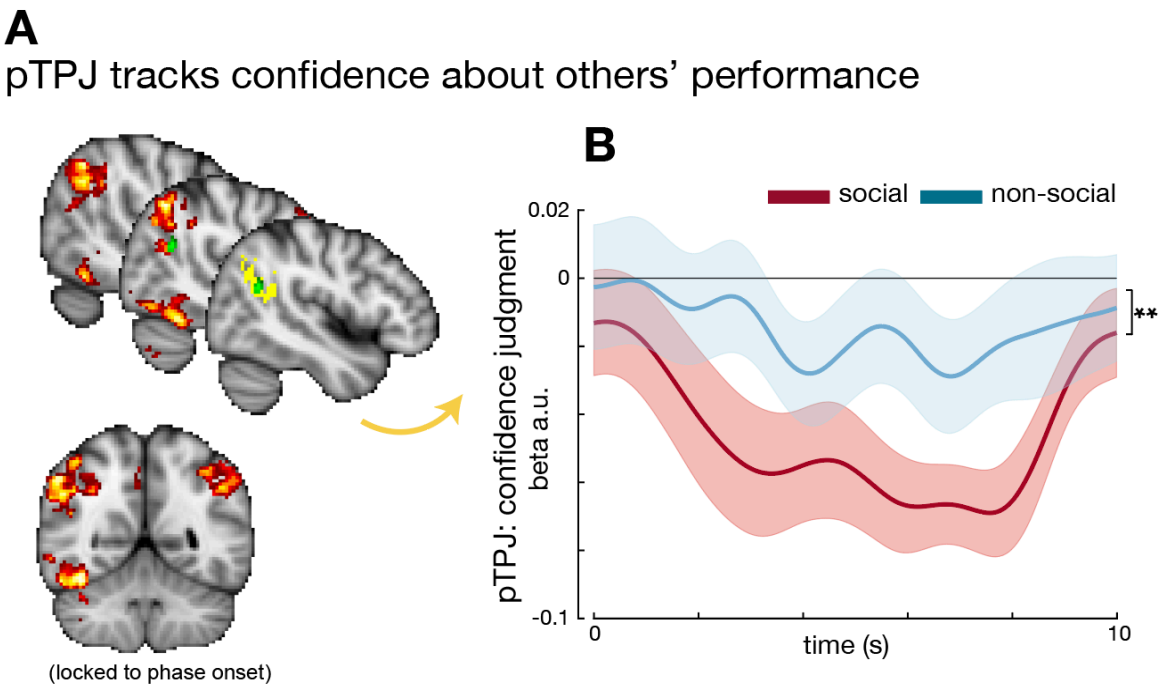


Figure 8. Unique social substrate: posterior TPJ represent confidence judgments about the performance of social predictors. (A) *Trial wise confidence judgment correlates with activity in right posterior TPJ when judging the performance of social advisors (red cluster shows cluster-corrected activation for social condition only). The difference in confidence judgement signals between social and non-social condition were strongest in a ventral part of right pTPJ (green cluster). This difference was observed when using small volume correction for a ROI based on a 90% probability pTPJ mask (yellow mask). Activation in this analysis was time-locked to the onset of the confidence phase.* ***(B)*** *Differences between conditions are shown when extracting a time course from an independent ROI based (yellow mask in A). (n = 24; whole-brain effects family-wise error cluster corrected with z > 3.1 and p < 0.05; pTPJ ROI-corrected with z >2.3 and p<0.05).*

In the behavioural analysis, we showed that participants set their confidence judgments as a function of multiple influences: their beliefs in the accuracy of the selected and forgone predictors and the uncertainty that they had in those beliefs, observations made on previous encounters, and previous confidence judgments. Next, we tested whether right pTPJ's activity could be shown to reflect each of the factors that influenced confidence setting (Figure 9). We used a time course analysis and included the same variables as in the behavioural GLM (see ROI-GLM3 in Methods for details). Time courses were extracted from the same ROI in

right pTPJ (Figure 8A, yellow mask). We found that right pTPJ activity covaried with beliefs in the social predictor's accuracy: activation decreased when participants had selected a social predictor that was believed to be accurate ($t(23)=-3$, $p=0.007$, $d=-0.6$,95% confidence interval=[-0.09 -0.02], but not when they had selected a non-social predictor they believed to be accurate ($t(23)=-1$, $p=0.3$, $d=-0.2$,95% confidence interval=[-0.054 0.019]; paired t-test for social vs non-social condition: $t(23)=-1.2$, $p=0.2$, d=,95% confidence interval =[-0.09 0.02]; Figure 9, top row). Note that the direction of this effect is consistent with the effect of confidence judgement that we had initially found; more negative fMRI signals indicate more confidence and the belief that a predictor is more accurate. Right pTPJ activation also covaried with the confidence judgment made on the previous trial: activity decreased when participants were more confident (i.e. selected a smaller interval) on the last encounter of the same predictor. Although this signal did not differ between conditions (paired t-test: $t(23) = -1.6$, $p=0.13$, $d=-0.33$,95% confidence interval =[-0.1 0.01], it was only present for social (social: $t(23) =-4$, $p<0.0001$, $d=-0.8$, 95% confidence interval=[-0.1 -0.04]) and not non-social predictors (non-social: $t(23)= -1.7$, $p=0.1$, $d=-0.35$, 95% confidence interval =[-0.07 0.006]; Figure 9, last row).

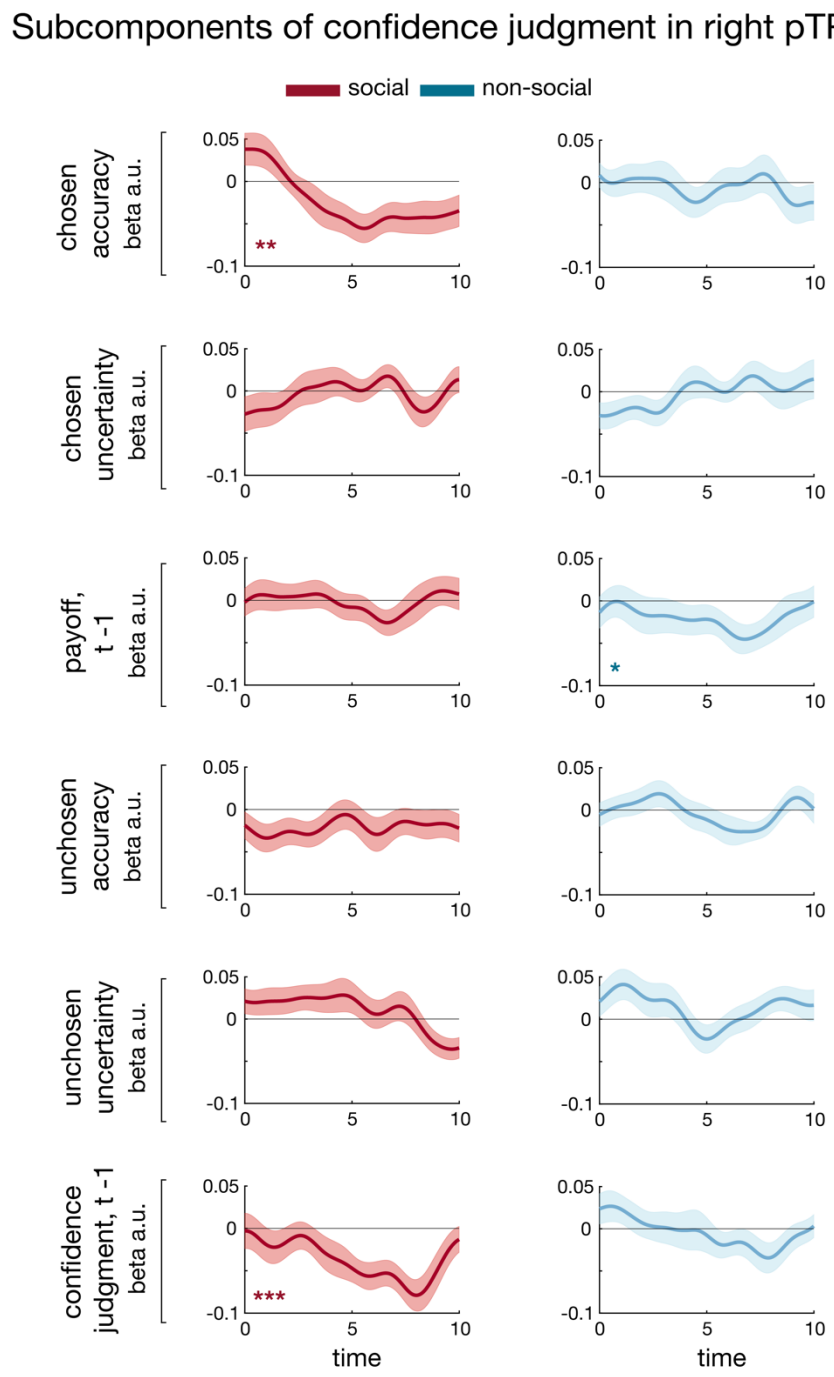


Figure 9. Subcomponents of confidence judgment in right pTPJ. *We tested whether the signal covarying with confidence judgment in right pTPJ could be decomposed into effects related to each of the factors that influenced confidence judgement. Signal in right pTPJ decreased when participants believed the selected predictor was accurate (regressor: chosen accuracy) and when participants were confident about the same predictor during previous encounters (regressor: confidence judgment t-1, t is for timepoint). These activation patterns were only present for social predictors. By contrast in the non-social condition right pTPJ activity decreased as the payoff received in the last encounter with the predictor increased (regressor: payoff t-1). (n = 24; error bars are SEM across participants).*

4.7.3. Confidence judgments about others' performance are based on recent and distant encounters

We have shown that activity in right pTPJ covaried with confidence judgments about social predictors made on the previous encounter. Further, we showed that participants were more confident (selected a smaller interval) when judging social predictors. These findings might relate to another set of results we have shown before, namely that previous confidence judgements influence the current confidence judgement over and above the effects mediated by accuracy and uncertainty in the chosen predictor. This suggests that participants tend to stick with their previous confidence in a predictor regardless of what their observations might have led them to believe about the predictor. However, the time period over which previous confidence is aggregated and carried over to the current trial is unclear. We tested whether participants integrated their prior judgments with different rates when judging social or non-social predictors. We first fitted a linear regression model to participants' behavioural confidence judgments as a function of judgments made at the last four encounters. Participants confidence judgments were influenced by previous judgments in a recency-weighted way (averaged across groups, repeated-measures ANOVA across four timepoints: $F(3,69) = 24.2$, $p<0.001$, $\eta^2= 0.36$; Figure 10A), however this was similar for both social and non-social conditions (difference between groups across four time points: repeated-measures ANOVA: 4 (timepoints) x 2 (group: social vs non-social) interaction: $F(3,69)= 0.79$, $p= 0.5$, $\eta^2=0.009$; Figure 8a).

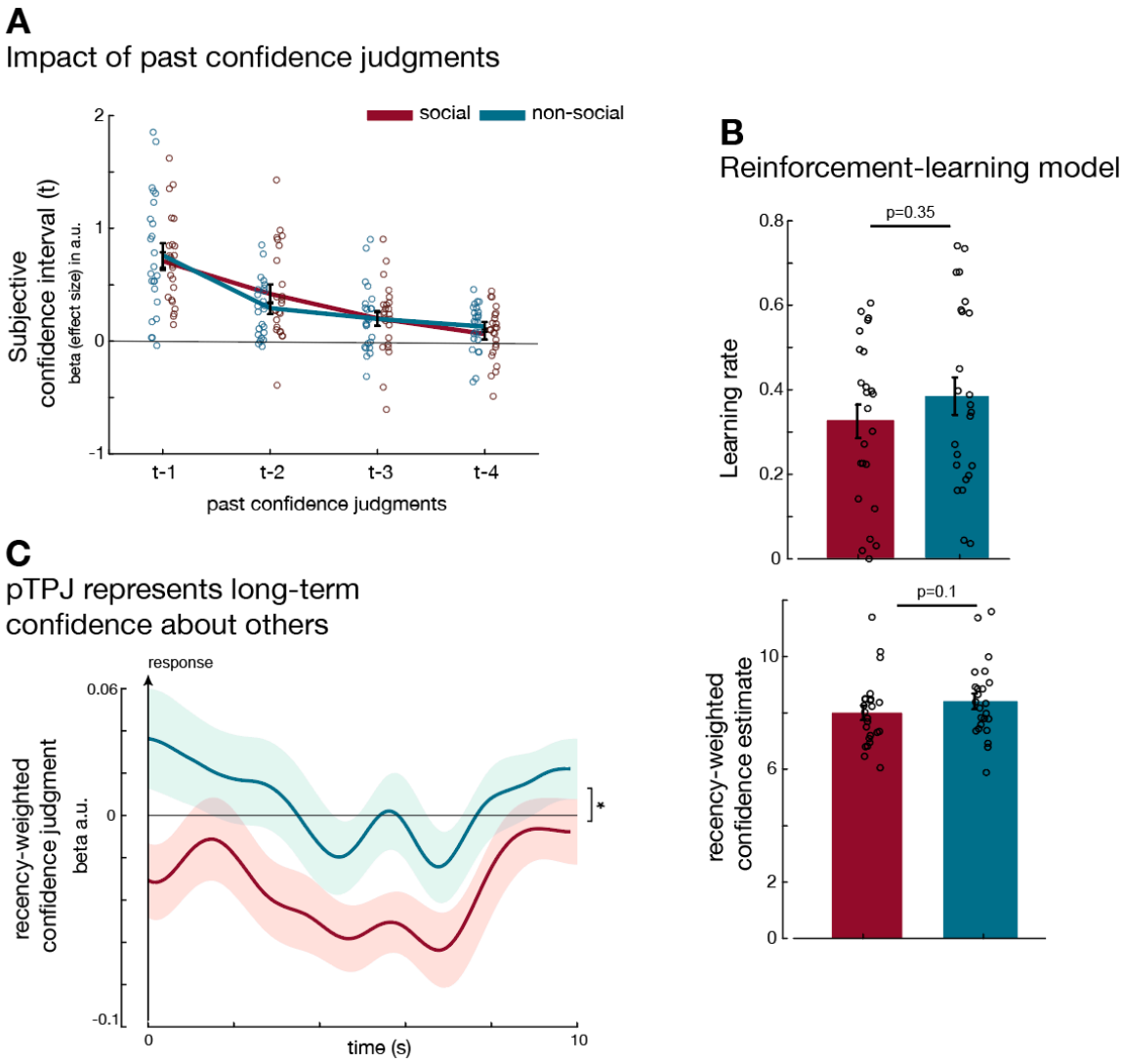


Figure 10. Confidence judgments are based on recent and distant judgments. (A) *Confidence judgments are formed in a recency-weighted way for social and non-social predictors (t represents time point).* ***(B)*** *A reinforcement learning (RL) model estimated the rate at which past judgments influence current judgement (expressed as leaning rate) and a recency-weighted confidence estimate (lower panel). Either parameter was not different between groups.* ***(C)*** *A time course extracted from right pTPJ showed that recency-weighted confidence judgments were represented for social, but not non-social predictors. (n = 24; error bars are SEM across participants).*

We used a reinforcement learning (RL) model to determine the rate at which past confidence judgments informed the current judgment. The learning rate parameter determines the rate at which new information replaces old information. Although the learning rate was lower for social compared to non-social predictors,

this difference was not significant (paired t-test: t(23)=-0.95, p=0.35, d=-0.19. 95% confidence interval= [-0.18 0.07]; Figure 10B, upper panel). As in previous analyses that indicated greater confidence about social predictors, the recency-weighted confidence judgment estimated by the RL model was smaller for social predictors; however the difference was not significant (t(23)=-1.6, p=0.12, d=-0.33, 95% confidence interval =[-0.94 0.12]; Figure 10B, lower panel).

Although we did not find differences in the RL model-derived parameters, we previously showed that activity in right pTPJ covaried with the confidence judgment on the current trial (Figure 8) and the one made on the previous encounter (Figure 9) for social predictors. Next, we tested whether activity in the same region also reflected a recency-weighted confidence judgment – i.e. not only those made in the last encounter with the predictor but also in more distant encounters. We used a time course analysis in which we included the RL-derived recency-weighted confidence judgment and controlled for other variables (ROI GLM4). When extracting a time course from right pTPJ, we found that activity covaried with the recency-weighted confidence estimate only for social predictors (t(23)=-3.5,p=0.002, d=,95% confidence interval=[-0.1 –0.3]) and not for non-social predictors (t(23)=0.17,p=0.86,d=,95% confidence interval=[-0.025 0.03]). This difference was significant (paired t-test: t(23)= -2.8, p=0.01, d=-0.57,95% confidence interval=[-0.1 -0.016]) (Figure 10C).

4.8. Discussion

We monitor other people and track their abilities when interacting with them (Boorman et al., 2013a; Wittmann et al., 2016b). Here, we investigated which aspects of others' abilities inform participants' confidence judgments. We designed a paradigm during which participants had to make predictions about where targets would appear on every trial. The task design meant that they could not learn directly about the locations, but instead they had to rely on information provided by either a social or non-social predictor. This allowed us to compare two task versions that differed only in their social and non-social framing and not in terms of any statistics of the task itself.

We differentiated between two kinds of decisions: a binary decision during which participants selected between predictors (decision phase) and a confidence decision during which participants considered only a single predictor and indicated how well they thought it would predict the target (confidence phase). We found that both types of decisions were influenced by Bayesian-derived belief estimates: beliefs about how accurately a predictor predicted the target (accuracy) and uncertainty about those beliefs (uncertainty). Many studies in social neuroscience focus on the latter type of decision, during which participants evaluate another social agent, for example their performance or their intentions in a strategic game (Hill et al., 2017; Wittmann et al., 2016b). Behavioural differences between conditions became apparent during this second kind of decision: participants were less confident when judging non-social compared to social predictors and were influenced more strongly by their subjective uncertainty when evaluating their

confidence in the predictors when they were of a non-social nature. These differences could be attributed to different performance expectations for social versus non-social predictors. Participants might expect other people, as opposed to non-social stimuli, to exhibit more consistent patterns of prediction. This might have led them to converge quickly towards one confidence judgment. There was, however, no credible evidence for such a hypothesis; there was no significant difference in the learning rate between conditions after fitting an RL model on the confidence judgements. The initial aim of this analysis was to test the autocorrelation of confidence judgments across time: whether participants changed their confidence judgments more or less strongly across time in the social compared to non-social condition given their previous judgments. However, an alternative approach would be to estimate confidence judgments as a function of a weighted average of angular errors (i.e. the difference between target location and predictor estimate) across time. Similarly, the prediction error within the RL model should then represent the difference between the confidence judgment and angular error. This analysis might be more appropriate given the actual behaviour in the task as participants should not base their confidence judgments solely on their previous judgments, but instead update their judgments according to new observations of the angular error.

Additional analyses are, therefore, required to understand the difference in confidence judgments between conditions, for example by separating trials into different phases. In previous chapters, we have shown that subjective uncertainty beliefs are modulated by phases of exploration and exploitation. Correspondingly,

these phases should alter confidence judgments and possibly show whether differences are specific to some phases within the experiment.

Participants performed two statistically identical tasks which only differed in their social or non-social framing. This allowed us to identify both common and distinct neural mechanisms underlying the processing of social and non-social information within the same participants. This allowed us to assess whether there are areas that preferentially encode social information (social brain hypothesis), or whether there are similar networks involved which receive input from socially specific areas (common currency hypothesis) (Ruff and Fehr, 2014). We found brain areas that functioned different in both contexts and also areas that operated in ways when participants were dealing with both social and non-social predictors. Activation in pgACC and ventral striatum covaried with increasingly confident judgments. Note that confidence judgments reflect beliefs about the strength or usefulness of the predictor but in addition they determine how likely it is that the participant will succeed in the trial. This is similar to studies that showed pgACC's role in tracking expected performance in social (Boorman et al., 2013a; Will et al., 2017; Wittmann et al., 2016b) and non-social settings (Bang and Fleming, 2018). In contrast to areas that compute the value difference between options (Chapter 2,3), pgACC might track the value of a single option, increasing in activity with higher predictive reward success (Lockwood and Wittmann, 2018). This interpretation is consistent with pgACC's role in representing payoff magnitude during the outcome phase of the current experiment.

Finally, we found that activity in right pTPJ represented confidence judgments for social, but not for non-social predictors. Right TPJ has long been credited with a

role in mentalizing by making inferences about the knowledge, beliefs, or desires of another person (Carter et al., 2012; Hill et al., 2017; Mars et al., 2013; Saxe, 2006; Schurz et al., 2014), for example when inferring the performance of another person (Wittmann et al., 2016b) or predicting the untrustworthiness of another person (Behrens et al., 2008). Consistent with these findings, we found that pTPJ activity covaried with participants' belief' about how accurate the social predictor performed. Additionally, activity in right TPJ also represented long-term beliefs about others: activity integrated past confidence judgments in a recency-weighted way. Social versus non-social differences were strongest in a ventral part of pTPJ in an area of activity that crossed the border between pTPJ and dorsal IPL (IPLD). Although beliefs about the accuracy of social predictors drove activity more strongly in pTPJ, both social and non-social predictors were associated with the same direction of activity change in IPLd: both areas decreased in activation with increasing confidence judgments about the predictor's performance. The adjacent intraparietal sulcus is often implicated in evidence accumulation for the option that will be chosen and the tracking of confidence in the choice (Gold and Shadlen, 2007; Shadlen and Kiani, 2013). Similarly, right TPJ might accumulate evidence about the expected performance of the other person and therefore might explain why the signal reaches its maximum just prior to the judgment (Kiani et al., 2014).

4.9. Summary

In this Chapter, I have shown that how we deal with useful information in the environment, that allows us to make useful predictions about what is going to happen next, differs depending on whether that information is derived from a social

or non-social source. In general, participants were more confident when using predictive information derived from social sources and simultaneously, they were less influenced by their uncertainty about the social information's accuracy. We showed common brain areas representing confidence judgments about expected performance for social and non-social predictors, however only right pTPJ coded immediate and long-term judgments about social predictors.

4.10. Supplementary Information

Contrast	Region	Peak Coordinates x/y/z (in mm MNI Space)	Z Value
Confidence judgment (average across social and non-social)	Dorsal Inferior Parietal lobe (right)	-42 -62 44	-4.4
	Perigenual anterior cingulate cortex	-4 42 6	4.4
	Ventral striatum (right)	10 12 0	6.1
	Fusiform face area	54 12 -22	4.4
Confidence judgment (social > non-social)	Posterior temporal junction (rpTPJ)	46 -54 22	3.4

Family-wise error cluster corrected, $z > 3.1$, $p < 0.05$ for average analysis

Family-wise error cluster corrected, $z > 2.3$, $p < 0.05$ for social vs non-social analysis

Table 3. *Cluster-corrected coordinates of fMRI-GLM1.*

5. CHAPTER: Learning about others through direct and vicarious experiences

5.1. Introduction

In order to thrive in a social environment, we need to learn about the people we encounter. Learning about their attitudes and abilities allows for accurate predictions of their behaviour and in turn flexible adaptation of one's own choices (Hampton et al., 2008; Hill et al., 2017; Ruff and Fehr, 2014). For example, knowing who has the tendency to cooperate and who does not, is a prerequisite for decisions about joint actions such as building shelter or gathering food supply. We can, however, learn about other people's cooperative tendencies in more than one way. We can, of course, learn by direct experience of those other people. We might estimate how likely they are to cooperate with us as a consequence of considering how frequently they have done so in the past. However, in addition, we can learn vicariously by observing the experiences of others when they interact with a given person. The first aim of the current experiment was to examine how we come to estimate the likelihood that someone will cooperate with us as a result of past, direct experience or as a result of past, vicarious experience.

The representations we form about others should be reflective of information observed during multiple encounters and during varying contexts (Wittmann et al., 2016b). The tendency to cooperate can fluctuate across contexts; for example, one

might learn that when time and resources are abundant, a specific other person might be happy to assist, but when time and resources are scarce, this person is no longer willing to help. If a representation of the other person's willingness to help was only based on one of these encounters, it would make incorrect predictions about the other's behaviour when context changes. The second aim of the current experiment was to examine whether our representations of others are contextually rich; do they allow us not just to make predictions about another person's behaviour, but do they allow for different predictions about others in different contexts.

While learning vicariously is often an important way in which to learn and it is important to consider that the persons we learn about behave differently in different contexts, another important contextual factor in social situations is towards whom a certain behaviour is directed; whether someone is interacting with us directly or whether we observe them interacting with others. Their behaviour might differ depending on who they are interacting with; for one reason or another, perhaps because of a prior friendship or some other link, they may behave more helpfully towards us as opposed to someone else or, *vice versa*, they may do the opposite. For example, people's helping behaviour is often influenced by whether it is directed towards a member of our own or a different group (De Dreu and Kret, 2016). Hence, considering information regarding the environmental context (e.g. resources held by each person) and social context (e.g. social partner and the history of previous interactions) are crucial determinants for accurate predictions of other people's behaviour. A third aim of the current study was to consider how the direction or target of a social behaviour might be taken into account when we learn

about other social agents; we, therefore, also looked at whether people can learn that another person might behave differently towards us or towards another person.

We designed a multi-player strategic game which followed similar principles to previous dilemma games (Hampton et al., 2008; Hill et al., 2017) during which participants can either cooperate with or defect against the other player to maximize points. Participants made two types of decisions: first, they decided how they themselves would behave towards another player. Second, they also predicted the other's behaviour towards themselves. These predictions were made either when directly interacting with another player (direct trials) or when observing the interaction between two other players (observational trials). To investigate how participants integrated information in the representations they formed of others, we manipulated contextual factors such as the endowment of each player on a given trial and social preference amongst players. We show that participants make their predictions and choices adaptively according to varying context and social partner. Further, we provide evidence that representations about others are formed from information observed during direct and vicarious experience, even if the information observed in the different contexts is contradictory. In this Chapter I will present the results of a behavioural study. We are planning to record neural data with this experiment to investigate how representations of others are formed from direct and vicarious experiences and how information is merged across these two types of learning.

5.2. Methods

5.2.1. Participants

Twenty-two participants took part in the experiment. We excluded participants on the basis of information that were collected during a debriefing questionnaire. The debriefing questionnaire assessed multiple aspects of participants' beliefs about other players. For example, they rated each 'other player' on objective measures that were manipulated in the experiment, such as the amount of rejection towards themselves and towards others. Some players had a very low (or very high) rejection rate towards both 'Self' and 'Other' and were clearly differentiable compared to players that had a mixed rejection rate towards 'Self' and 'Other'. We excluded one participant, who rated the rejection rate of these players as behaving the opposite way around: a loyal character was estimated as being opportunistic and vice versa. This participant was excluded on the basis of their subjective report. We asked in the same questionnaire whether participants had any suspicions about the experimental set-up, thereby probing their beliefs about the deception procedure, i.e. whether other players were in fact playing the same game (Question: "Was there anything you found suspicious during the experiment? If yes, please tell us what you think.") We excluded another participant who did not believe they were playing with 'real' other players, i.e. did not believe the deception procedure (two participants excluded in total; final sample: 20 participants, 14 female). The study was approved by the Central Research Ethics Committee (R64079/RE001) at the University of Oxford. All participants gave informed consent.

5.2.2. Experimental procedure

On the day of the experiment, participants first received an explanation of the experimental task and performed a practice session. Due to the complexity of the task, the experimenter was present during the practice session to simultaneously probe the participant's understanding of the task. The instruction part lasted between 30 and 60 minutes, the experimental task took one hour such that the entire visit lasted approximately 2 ½ hours. Participants received £7.50 per hour with the opportunity of earning a bonus payment during the experimental session (range: £2.50-£7.50, mean: £4.77).

The experimental procedure was precisely scripted and involved two experimenters, one of which was a confederate experimenter. The confederate experimenter was instructed to be one of multiple experimenters. This allowed us to make it seem like both experimenter were coordinating incoming participants and the start of the experiment, hence making the participant believe that other participants were simultaneously taking part in the experiment. Other features were implemented to reinforce the belief that other participants were taking part in the experiment. For example, online time-tables to register prior to the experiment contained the initials of other participants already filled-in. The in-person and phone communication between experimenters and participants also emphasized the need to coordinate the testing of participants and also gave participants a sense of working with other individuals. Artificial delays were inserted in practice sessions to give the impression of another player taking time to respond. At the end of the experiment, participants filled in a debriefing questionnaire that assessed participants' comprehension of the task and the verisimilitude of their

impressions regarding the other players with whom they interacted during the experiment. This allowed us to probe participants' beliefs about the session without directly discussing the possibility of deception. Participants were debriefed via e-mail at the end of data collection.

Note that, before attending the experimental procedure, participants filled in online questionnaires via the web platform Qualtrics (www.qualtrics.com). We acquired data from five questionnaires: The Questionnaire of Cognitive and Affective Empathy, Rosenberg Self-Esteem Scale, Social Reward Questionnaire, Self-report Psychopathology and the Autism Quotient. Questionnaires were selected to assess the relation between behavioural measurements and social and cognitive factors. These data will be analysed in future analyses.

5.2.3. Experimental design

Participants took part in a multi-player strategic cooperation game. On each trial, participants were endowed with an amount of resources which determined their possible payoff depending on the behavioural strategy they chose. The aim was to maximize points by strategically cooperating with other players and by predicting others' actions accurately. Points could be maximized in two different conditions: participants either interacted with other players directly, referred to as Self-Other trials (SO trials), or they indirectly observed other players interacting with each other, referred to as Other-Other trials (OO trials). SO and OO trials were pseudorandomly interleaved. Across the experiment, participants encountered four other players that appeared in SO and OO trials with equal frequency. There were 162 trials in total, of which 108 were SO trials and 54 were OO trials. The number of SO trials is double that of OO trials because participants interacted with one

player at a time during SO trials, while they observed two players simultaneously during OO trials. This meant that overall, the number of others' actions observed directly in SO trials (108) was the same as the number of others' actions observed indirectly in OO trials (54 times actions of both others were observed). Although this was not explicitly cued, the experiment was divided into three blocks. Observations made about others within each block were reflective of their overall behaviour throughout the task. The same resource differences were assigned to players across blocks. Further, SO and OO trials occurred with the same frequency within blocks (54 trials within each block: 36 SO and 18 OO trials). These task features ensured that participants did not form representations of others on the basis of information that was not representative of the general behaviour of each player. The order of blocks was counterbalanced across participants.

5.2.3.1. Trial timeline

At the beginning of a trial, participants saw the amount of resources that were allocated to each player (Figure 1). An amount was indicted on each player's resource bars (indicated in yellow). This ranged between one to nine per player, while its sum across players resulted always in ten resource bars. Hence, if one player had many resource bars, the other player had few resource bars. All resource differences occurred with the same frequency for all player combinations.

Participants could also distinguish between SO and OO trials depending on the player identities presented at the beginning of the trial. Other players were represented by female or male facial stimuli, while the participant was represented by their initials (MK in Figure 1). Facial stimuli were taken from a database, not

matched to the participants' gender, but instead counterbalanced across participants.

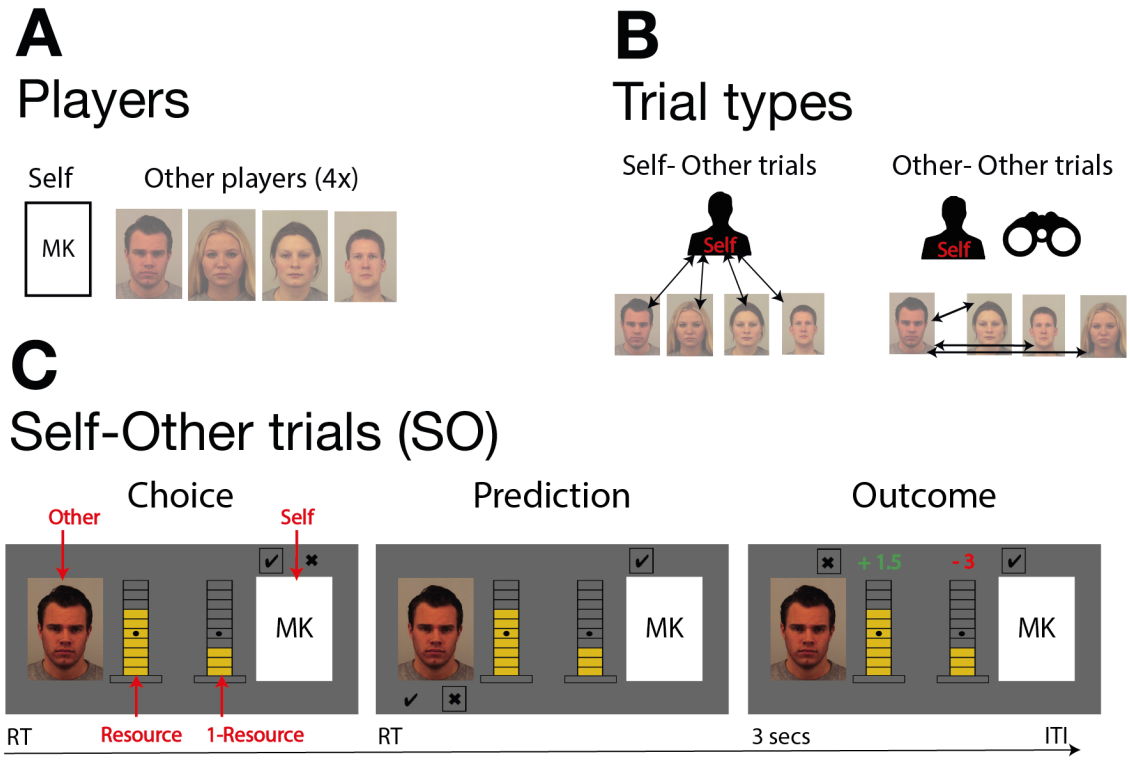


Figure 1. Trial timeline. **(A)** Participants engaged in an interactive strategic game with four other players. Participants were indicated by their initials (here: MK), while other players were indicated by a facial stimuli that was randomly drawn from a set of six possible faces. **(B)** Participants either directly interacted with another player (left panel: Self-Other trials, SO trials) or indirectly observed two other players interacting with each other (right panel: Other-Other trials, OO trials). **(C,D)** At every trial, each player received an amount of resource on a resource bar (yellow bars) which summed up to ten across the two players on a given trial: hence, lower resource bars for one player meant higher resource bars for the other player. **(C)** SO-trial: participants first decided whether to cooperate (check-symbol on the right above the participant's initial indicates an 'Accept' choice) or not cooperate (cross-symbols indicated 'reject' on the left above the participant's initials). Next, a prediction was made about how the other player was likely to behave towards the participant. Here, check-symbols meant that participants predicted the other player to accept the offer and cross-symbols meant that they predicted the other player to reject the offer. In this example, the participant predicts to be rejected by the other player. During the outcome phase, players were allocated their payoffs while participants could additionally verify whether their predictions about the other player's

behaviour were correct. **(D)** OO-trial: participants made two consecutive predictions about the behaviour of each player towards the other. Here, the player who is making a prediction is referred to as 'relevant other', while the other player is referred to as 'irrelevant other'. Again, checks and crosses indicate the predictions as in the SO trials. During the outcome phase, players received their payoff, while the participant could update their beliefs about the other's cooperative behaviour.

SO trials included a choice, prediction and outcome phase (Figure 1C). During the choice phase, participants could choose to cooperate by selecting "accept" (A) or choose not to cooperate by selecting "reject" (R). The location of (A) and (R) was counterbalanced across participants. During the prediction phase, participants predicted how the other player would behave towards themselves, similarly by either predicting (A) or (R). During the outcome phase, participants observed their mutual responses towards each other, and the payoff gained or lost by each player. Participants collected points through these cooperative decisions, but also by making correct predictions. Points displayed only represent points resulting from cooperative decisions and did not include possible bonus points earned. The aim was to maximize both types of points. To avoid any bias in the sequence of decisions, the order of prediction and choice during SO trials was randomised across trials. During OO trials, participants observed choices of two other players towards each other (Figure 1D). Here, the three phases consisted of two prediction phases and an outcome phase. The participant predicted how each player would behave towards the other player given the players' resource difference.

The duration of each decision phase (choice and prediction phase) was determined by the participant's reaction time, while the outcome was shown for a

fixed amount of 3 seconds. We included time jitters between phases that were supposed to represent other players' reaction times.

5.2.3.2. Payoff structure

We designed a payoff structure that resembles that used in previous prisoner dilemma games during which participants can either cooperate with or defect against the other player (Hampton et al., 2008; Hill et al., 2017). Whether a player wins or loses points on a given trial was determined by the choice combination between players, while its payoff magnitude (win or loss) was dependent on the amount of resource bars each player had. To maximize points, players had to consider both the possible decisions that might be made by the other player as well as possible gains or losses associated with different choice combinations.

There were four possible choice combinations (Figure 2A), of which a mutual cooperative decision (Accept | Accept) led to the highest payoff for both players. The subsequent highest outcome was to reject another player who accepted (Reject | Accept). A payoff loss could only occur when accepting another player who rejected (Accept | Reject). A mutual rejection resulted in a zero payoff for both players (Reject | Reject). In general, if the other player decided to cooperate the participant received points, but when the other player decided to reject, the participant in most cases lost points.

In this game, mutual cooperation (Accept | Accept) leads to higher points compared to any other strategy with any possible resource differences between players. In the example illustrated in Figure 2A-ii, mutual cooperation results in 7 points for Self, while rejecting the other player, when they are cooperating, only

results in 3.5 points. When rejecting another player who accepted the interaction, the win and loss for each player respectively was determined by the resources owned by the player who was rejected. In more detail, the person who rejected receives 50% of the resources owned by the person who was rejected, while the person who was rejected lost points equal to their own resources. In contrast, during mutual cooperation both players won 100% of their opposing players' resources and therefore this type of interaction was always the most rewarding.

Winning points, either during mutual cooperation or unilateral defection is even more beneficial when one has been endowed with low resources at the start of the trial. This design feature reflects a natural imbalance inherent in many cooperative situations: often, one cooperation partner is in a privileged position compared to others and therefore partners with lesser resources can gain more from joining forces. We illustrate the payoff structure for each possible choice combination and possible resource differences between players in Figure 2A-iii. The decreasing slopes from left to right demonstrate two features: it indicates that the magnitude of potential wins decreases with increasing resources (blue and orange lines), while the risk of losing many points increases (yellow line). Hence, the balance between risk and benefit for each strategy changes when resource differences between players are varied. This means, although mutual cooperation always leads to the highest win, being defected against by another player when one has high resources can cause a substantial payoff loss. In such a case, it might be more tempting to defect against the other player. Although the potential payoff of taking this action is less, so is the potential cost. In summary, a payoff win or loss depends on the choice combination between players, while its magnitude

depends on the resources that the player holds and the recourses that the other player holds: if a player has a lot of resources at the start of a trial, there is a risk of a substantial loss and no possibility of a large win; *vice versa* when a player has low resources, then there is the potential to earn a lot and only to lose a little (Figure 2B-iv).

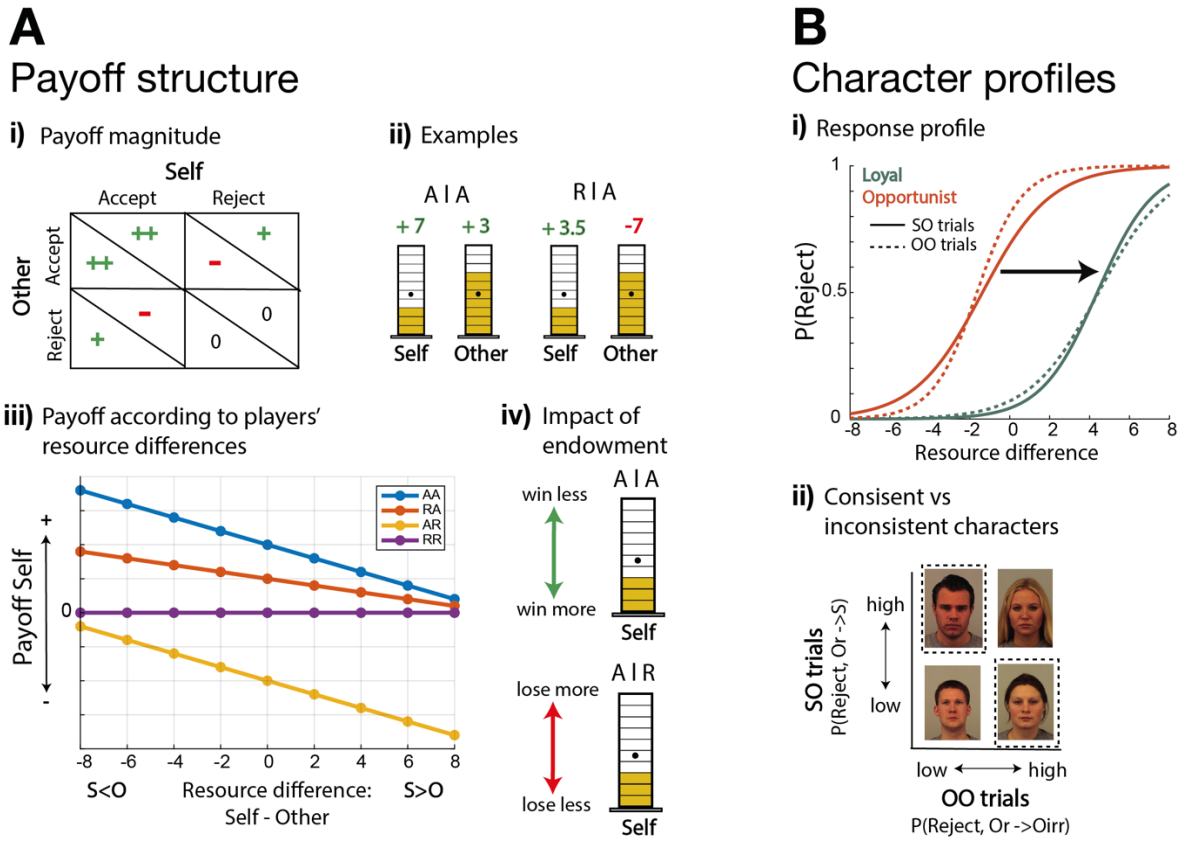


Figure 11. Experimental Design: payoff structure and character profiles. **(A-i)** Each player's payoff is determined by the choice combination of both players. Mutual cooperation yields the highest payoff for each player, while being rejected by another player can result in a payoff loss for the participant. **(A-ii)** Example of two choice combinations with the same resource difference: mutual cooperation results in a higher payoff compared to rejecting the other person (compare payoff for Self [S] on the left and right). The payoff for mutual cooperation was for each player to receive the resources of the other player. The payoff for rejecting the other player who had offered to cooperate was just half of the other player's resources. For the player who offered to cooperate, however, being defected against resulted in a payoff loss equalling their own resources (see payoff '-7' for Other). **(A-iii,iv)** A win or loss is determined by the choice combination between players, while its payoff magnitude is determined by the resource difference between players. The balance between the potential cost and benefit of each strategy changes as a function of whether one has high or low resources. When one has high resources (right-hand side), there is a high risk of losing a lot

while there is only a small amount to potentially gain. *Vice versa* there is potentially more to be gained when one has low resources and a potential of only a small loss (left-hand side). **(B-i)** As a trait-feature of the other players in the game, they differed in their tendency to reject cooperation, partly as a function of who that cooperation is directed towards. This feature was superimposed on the general tendency of all players to increase rejection as the resource difference became disadvantageous for them (positive x-values indicate that the other player had more resources than oneself). Players differed in their rejection rate in the following way: a loyal player had a low rejection rate and cooperated even if the possible loss was larger than the potential gain (shift to the right). The rejection rate for an opportunistic player was high such that they only cooperated when the benefit of cooperation exceeded the possible costs it would incur (shift to the left). Each line denotes the mean across two players of the same character profile towards the same group (Self or Other). For example, the solid orange line represents the average rejection rate across both opportunistic players directed towards Self in SO trials. The dotted orange line represents the average rejection rate across both opportunistic players directed towards the relevant Other in OO trials. The analogue version holds for the green line, here representing an average across loyal characters for SO trials (solid line) and OO trials (dotted line). The aim of this panel is first to show the difference between opportunistic and loyal characters and second, to show the similarities of the same characters across SO and OO trials – thereby making it possible to compare choice behaviour across SO and OO trials in response to different character profiles. **(B-ii)** Two of the other players had similar rejection rates towards Self in SO trials (y-axis) and towards Other in OO trials (x-axis) by being loyal or opportunistic towards Self and Other (consistent character profiles are on the bottom left and top right). The other two players had a different rejection rate towards Self and Other by being loyal towards Self and opportunistic towards Others *or vice versa* (inconsistent character profiles, indicated with a dotted frame on the top left and bottom right).

5.2.3.3. Character profiles

The rate at which a player cooperated or defected was dependent on their character profile (Figure 2B). Players could either be loyal or opportunistic and which character profile they obtained was experimentally pre-determined. Their character profile was determined by a Softmax function that integrated the resource difference between players and used this as the decision variable to determine the choice that would be taken (Figure 2B-i). A loyal player had a low rejection rate and cooperated, even if the potential loss outweighed the potential gain, while an opportunistic player had a relatively high rejection rate and only cooperated when the benefit of cooperation exceeded the possible costs that might be incurred. The difference between character profiles was manipulated by the

offset of the Softmax function, while their slope was kept as identical as possible across characters and SO and OO trials. This allowed us to compare predictions made about others' behaviour between SO and OO trials.

We were interested whether participants learned from both direct and observational experience and the degree to which they integrated the information obtained under these two conditions. To study whether participants form social representations separately from direct and observational experience, or whether they merge the information gained from both sources, we designed characters who had either the same or different rejection rate across SO and OO trials (Figure 2B-ii). Two characters behaved consistently across SO and OO trials, by either being loyal or opportunistic towards both Self and Other, while inconsistent characters behaved loyally towards Self and opportunistically towards Others, or *vice versa* (Figure 2B-ii: inconsistent characters are indicated with a dotted frame).

In sum, other players had different character profiles which determined two aspects of behaviour: first, they determined how likely they cooperated as a function of resource difference and second, they determined the consistency of choice pattern. Consistent characters behaved in the same way when interacting with the participant and other players, but inconsistent characters behaved differently towards the participant and the other players. Therefore, participants had an incentive to tell characters apart, learn about them, and to make choices adaptively.

5.3. Behavioural results

5.3.1. Predictions about others' behaviour vary according to contextual and social factors

We were interested in how participants estimated the cooperative attitudes of another person when directly interacting with them (Self-Other trials, SO trials) and when indirectly observing two other players interact with each other (Other-Other trials, OO trials). With our paradigm, we measure how participants *predict* another player to act in trials in which they interact with the other player directly (SO trials) and trials in which they observe them acting towards another person (OO trials). In addition, in SO trials, participants also *chose* how to act towards the other player. First, to test estimation during direct and indirect experiences, we analysed whether participants adapted their prediction rate according to two experimental factors: first, whether they anticipated other players' behavioural changes according to the resource difference between players. Second, whether they anticipated that other players had different character profiles which determined their rejection rate. For both, SO and OO trials, participants adapted their predictions according to the resource difference between players at a given trial (Figure 3A; main effect of resource difference, SO trials: Greenhouse-Geisser correction due to Sphericity violation, $F(2.6,152)=73$, $p<0.0001$; OO trials: Greenhouse-Geiser correction due Sphericity violation, $F(4.2,80.7)=76.3$, $p<0.001$). When resource differences were positive (SO trials: resources are higher for Other compared to Self; OO trials: resources are higher for the relevant Other compared to the irrelevant Other), participants predicted rejection at a higher rate than when resource differences were negative (SO trials: resources are higher for Self

compared to Other; OO trials: resources are lower for the relevant Other compared to the irrelevant Other). Additionally, participants also integrated information about the character profile of the currently relevant player. During both SO and OO trials, participants expected higher rejection rate from opportunistic compared to loyal characters (Figure 2A; main effect of character profile; SO trials: $F(1,19)=86, p<0.0001$; OO trials: $F(1,19)=18, p<0.0001$). Finally, participants integrated both features dynamically and learnt that different character profiles could behave differently according to the resource difference (Figure 3A, interaction character profile x resource difference, SO trials: Greenhouse-Geisser correction due to sphericity violation, $F(8,152)=3.3$, $p=0.01$; OO trials: $F(8,152)=3.7$, $P=0.001$). These results show that participants understood the payoff structure but in addition, also tracked the behaviour of others by dynamically combining information about the context and social partner.

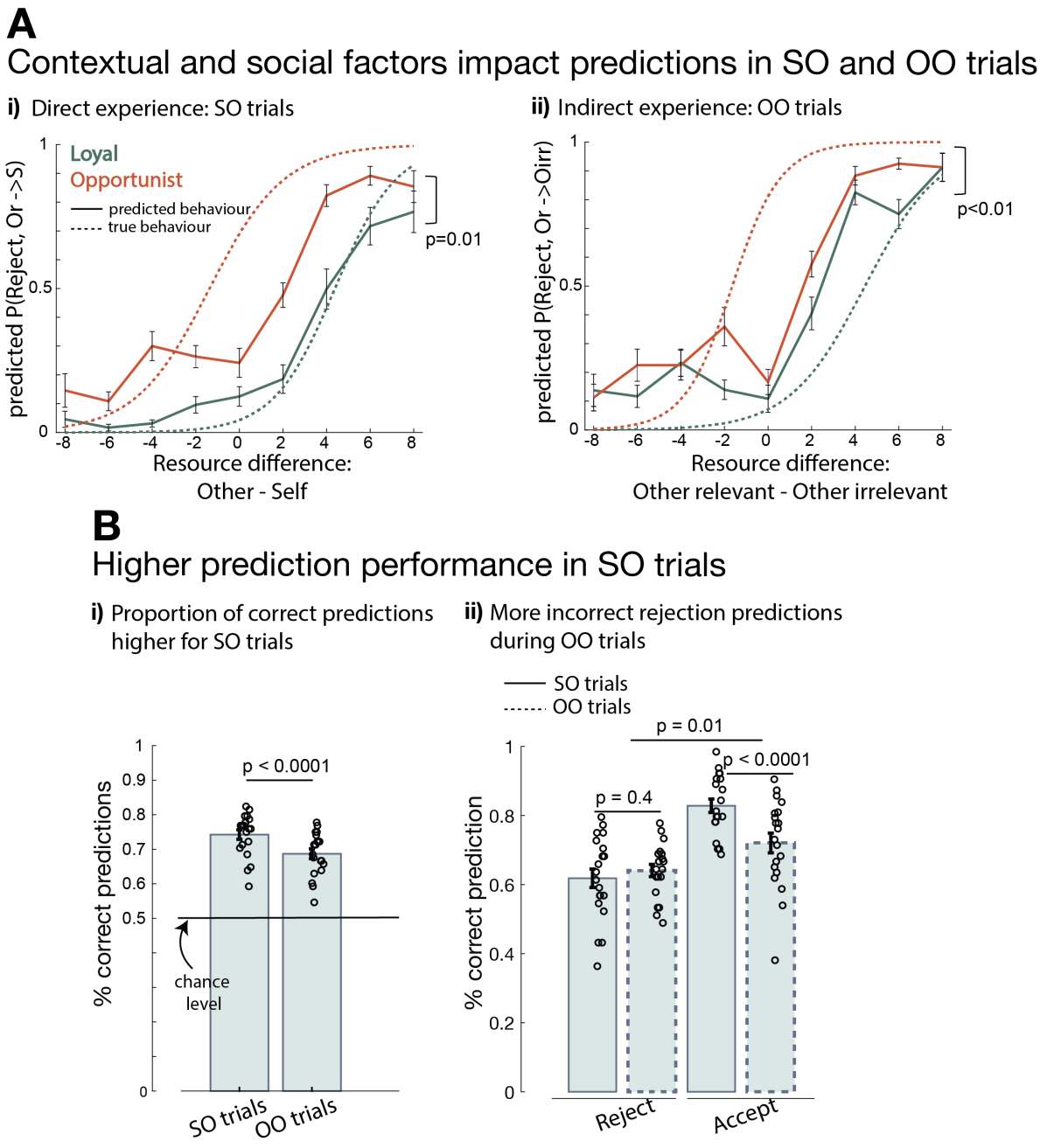


Figure 12. Prediction performance during SO and OO trials. **(A)** Participants adapt their predictions according to the character profiles, resource difference and its interaction of both contextual and social factors during SO trials (A-i) and OO trials (A-ii). The dashed lines indicate the ground truth and the solid lines indicate participants' prediction binned by resource difference. The orange and green lines aggregate over both consistent and inconsistent types of players. **(B-i)** Proportion of correct predictions is higher for SO compared to OO trials. **(B-ii)** Participants were better in predicting others' cooperative response towards themselves (SO trials) but there was no difference in predicting (A) or (R) responses towards others (OO trials); this difference between SO and OO trials was significant (see main text for statistical details). Further, the difference in total proportion of correct trials between SO and OO trials is additionally due to participants incorrectly predicting the other person to reject during OO trials,

while in fact they accepted (compare % correct predictions for accept trials between SO and OO trials).

It has been shown that humans learn better when information is related to oneself (Lockwood et al., 2018). We tested whether the accuracy of predicting others' behaviour differed between SO and OO trials, that is when being directly affected by the other's behaviour or when indirectly observing two other players interacting. There was a higher proportion of correct predictions when interacting directly with the other person compared to observing other players interacting with each other (paired t-test, SO vs OO trials: $t(19)=3.9$, $p<0.0001$, $d=0.87$, 95% Confidence interval, from now referred to as CI=[0.03 0.08]). Next, we tested whether the difference in prediction accuracy resulted from the frequency of a specific type of response, i.e. cooperation (Accept, A) vs defection (Reject, R). We calculated the percentage of correct predictions for (A) and (R) for each SO and OO trials (Figure 3B-ii). Participants were better in predicting others cooperative behaviour (A) towards themselves during SO trials (% correct prediction for accept > reject, paired t-test: $t(19) = 5.7$, $p<0.0001$, $d=1.3$, 95% CI= [0.13 0.29]), but there was no significant difference between (A) or (R) predictions during OO trials (% correct prediction for accept > reject, paired t-test: $t(19) = 1.9$, $p=0.07$, $d=0.44$, 95% CI= [-0.007 0.17]). This difference between response strategies was significant when comparing SO and OO trials (paired t-test of their respective differences, $t(19) = -2.8$, $p=0.01$, $d=-0.65$, 95% CI= [-0.23 -0.03]). Further, the proportion of correct predictions differed between SO and OO trials when comparing cooperative responses: in other words, participants incorrectly predicted the other person to reject during OO trials, while in fact they accepted ((A) for SO vs OO trials, paired t-test: $t(19) = 4$, $p<0.0001$, $d=$, 95% CI=[0.05 0.16]). This was less frequently the

case during SO trials, possibly indicating a 'rejection bias' during OO trials, but not during SO trials. In sum, participants made more mistakes during OO trials which were caused mostly by incorrectly predicting a defection behaviour from the other person.

5.3.2. Impact of irrelevant interaction on current prediction

We have shown that participants adapted their predictions within SO and OO trials flexibly according to the context of resource differences and the identity of the current social partner. One can construct two accounts to explain the prediction data. First, participants' predictions might be solely informed by the relevant interaction. For example, only information from previous SO interactions might have been used to make predictions about future SO interactions, but not about OO interactions. Similarly, only information observed during OO interactions might be used to make predictions about how others would behave towards each other in OO trials in the future. An alternative account hypothesizes that participants draw on information learned about a particular agent during SO interactions to inform predictions about the same agent during OO interactions and, *vice versa*, they draw on OO experiences to inform predictions about the same agent during SO interactions. To test this hypothesis, we compared prediction data between consistent and inconsistent characters (Figure 4). Specifically, we compared the mean rejection rate for consistent and inconsistent loyal characters within each SO and OO trials. If participants used information from the relevant interaction only (SO interactions only inform SO predictions and OO interactions only inform OO predictions), then characters with the same profile should be predicted to have the

same mean rejection rate, independent of whether they behaved consistently or inconsistently across SO and OO trials. However, if participants merged information from irrelevant interactions into the current prediction, then they should predict a mean rejection rate for inconsistent characters that differ from the mean predicted rejection rate for a consistent character even if both characters' interactions with the self were similarly loyal or disloyal. Similarly, predictions for characters in OO interactions should also be influenced by their behaviours in SO interactions and not just previous OO interactions. We found that within both interactions, information from the other interaction type (SO or OO) was used to inform current predictions: inconsistent loyal characters were predicted to have a higher rejection rate compared to the consistent loyal profile. This increase in rejection rate can be explained by the observations made during the other interaction type when behaviour associated with an opportunistic character profile would have been observed (SO trials: paired t-test LL vs LO (the first letter refers to SO interaction, the last letter refers to OO interaction), $t(19) = -3.2$, $p=0.004$, $d=-0.7$, 95% CI= [-0.13 -0.028]; OO trials: paired t-test LL vs OL, $t(19) = -5.4$, $p<0.0001$, $d=-3.8$, 95% CI=[-0.2 -0.09]). Similarly, participants integrated information from irrelevant interactions to inform their current predictions about opportunistic characters. Mean rejection rates were lower for inconsistent compared to consistent opportunistic character presumably because of the observations of loyal behaviour that would have been made when the character was interacting with other individuals (SO trials: paired t-test OO vs OL, $t(19) = 2.8$, $p=0.01$, $d=0.63$, 95% CI= [0.03 0.02]; OO trials: paired t-test OO vs LO, $t(19) =3$, $p=0.006$, $d=0.69$, 95% CI=[0.03 0.17]). These results show that participants merge

information derived across both SO and OO trials to form representations of others, even if the information gathered in the two contexts is contradictory.

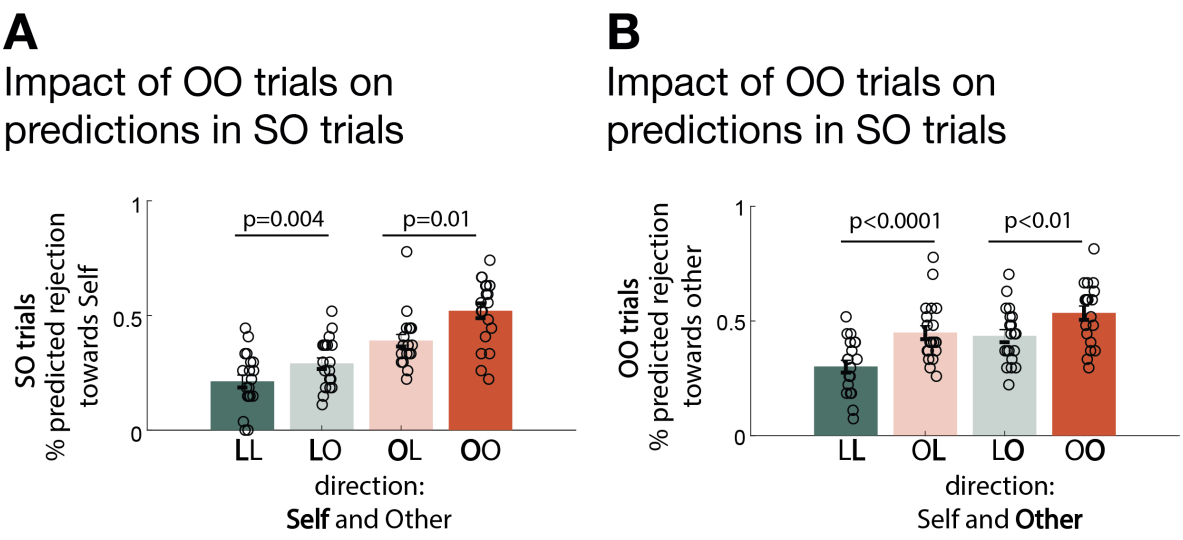


Figure 13. Impact of irrelevant interaction on current prediction. We compared the mean rejection rate across consistent and inconsistent characters who otherwise had a similar behavioural profile when interacting with the self (A) or the other (B). This allowed us to test whether participants merged information gathered in both SO and OO trials. Information from the other type of interaction was used to inform current predictions about the behaviour of others. Inconsistent loyal characters were predicted to have a higher mean rejection rate compared to consistently loyal character profiles: compare LL and LO in SO trials (A) and LL and OL in OO trials (B). Similarly, inconsistent opportunistic characters were predicted to be more loyal than they really were: compare OO and OL in SO trials (A) and OO and LO in OO trials (B).

5.3.3. Choices towards others are made prospectively

We have shown how participants integrate information to make *predictions* about others' behaviour. Next, we tested how participants adapted their own *choice* to cooperate with others during SO trials (Figure 5A). First, we tested whether participants integrated information about whom they were currently interacting with (character profile) and whether they changed their choice depending on the

resource difference between players. Participants differentiated between character profiles and adapted their choices correspondingly by rejecting opportunistic others more often compared to loyal characters (repeated measures ANOVA, main effect of character profile: $F(1,19) = 74.4$, $p<0.0001$). They also integrated information about the resource differences between players into their decision making when making choices (repeated measures ANOVA, main effect of resource difference, Greenhouse-Geisser due to Sphericity violation: $F(2.7,50.7) = 4.4$, $p=0.01$). Moreover, they behaved differently towards loyal and opportunistic characters depending on whether resource differences were low or high (repeated measures ANOVA, resource difference x character profile: $F(8,152)=3.7$, $p=0.001$). Thus, participants integrated information about both contextual and individual character factors and adapted their choices flexibly.

Notably, the effect of resource difference exerted a different impact on participants' choice patterns as opposed to their prediction patterns. Participants predicted others' behaviour according to a SoftMax function and predicted others to reject more often when they had more resources compared to the other player. In these cases, defecting against the other player might be beneficial as there was a risk for a substantial loss when being defected against by another player. However, the choice pattern did not follow the same principle and instead it resembled a U-function with higher rejection rates for either extreme of resource difference between Self and the Other. Congruent with the payoff structure and with the participants own prediction patterns, participants rejected others when there was a risk of a substantial loss and only a potential for a small gain (high resources, left-hand side). However, unlike the previous prediction pattern,

participants also rejected others when there was a potential for a substantial gain and only a small risk of loss as not many points could be lost when being defected against (low resources, right-hand side). This shows that participants integrate information about others in a prospective manner by anticipating others' rejections towards themselves in such circumstances.

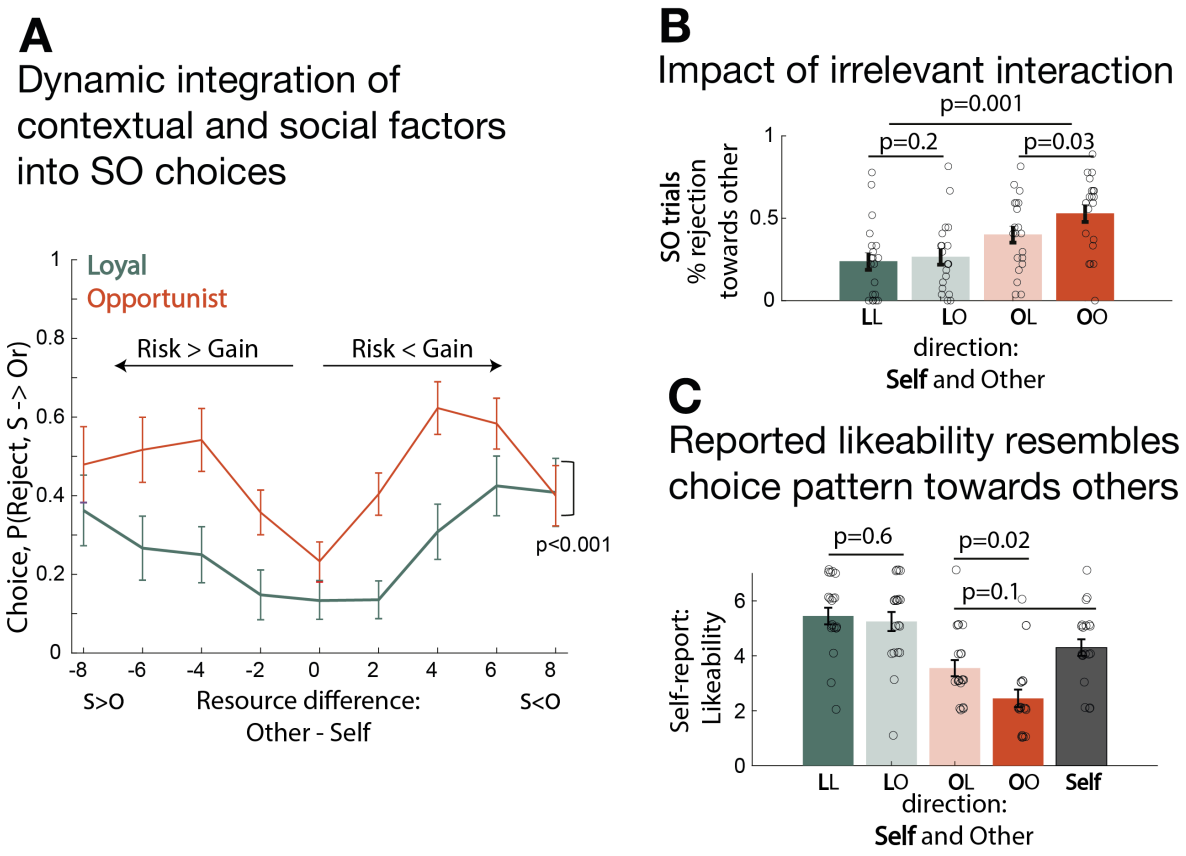


Figure 14. Choices are made prospectively. **(A)** Choices towards others integrated information about resource differences and character profiles. However, they were also made prospectively by anticipating other's behaviour during extreme resource differences. This was particular evident during scenarios where the risk was smaller than the potential gain (right-hand side); participants still decided to defect the other person due to anticipation of a rejection. **(B)** Participants were influenced by observational experiences when making a decision towards opportunistic players: participants rejected the inconsistent (OL, light orange) player less often compared to the consistently opportunistic character (OO, dark orange). This mergence across experiences was not observed for loyal characters: there was no difference in mean rejection rate between the consistent (LL, dark green) and inconsistent (LO, light green) players. **(C)** The reported likeability of each player resembles the observed choice pattern: there was no difference in likeability between loyal players, while likeability was affected by information observed during direct and observational experiences: the inconsistent player (OL, light orange) is rated higher in likeability compared to the consistently opportunistic player (OO, dark orange). Participants judged themselves similar to one other player (OL, light orange).

Next, we tested whether participants integrated information from other interactions with other agents when deciding to cooperate with or defect against the other player. We used the same analysis principle as before and compared the mean rejection rate across consistent and inconsistent characters with otherwise similar character profiles. We observed a mergence across interactions only for opportunistic characters: participants rejected opportunistic characters less often when they were loyal towards others [paired t-test, OO vs OL: t(19)=-2.3, p=0.03, d=-0.5, 95% CI=[-0.24 -0.01]). However, loyal characters were rejected at similar rates regardless of how they behaved differently towards others [paired t-test, LL vs LO: t(19)=-1.2, p=0.24, d=-0.27, 95%CI=[-0.08 0.02]). The difference in choice behaviour towards loyal and opportunistic behaviour was significant (paired t-test: absolute of (LO -LL) vs absolute of (OL- OO): t(19) = -3.9, p=0.001, d=-0.9, 95% CI=[-0.02 -0.06]). This last observation made during actual decision making differs from the predictions that were made in the same trials. This might suggest that there is a difference in how participants predict others' behaviour towards themselves and their own choices towards others.

We also examined how much participants reported that they liked the other players in the game. The mean rejection rate towards others resembled the likability rating for other players that participants reported: participants were not influenced by information from other interaction types when rating the likeability of others who were loyal towards themselves, independent of whether they behaved differently towards others. However, when judging opportunistic characters, participants preferred to interact with the inconsistent character that was estimated

to be more loyal due to its behaviour during OO interactions that were less directly relevant to SO interactions.

Finally, participants rated themselves on how much they thought others liked to interact with them. Their rating they ascribed themselves differed from that they ascribed to all other characters (paired t-tests; Self vs LL: $t(19) = -4.5$, $p<0.0001$, $d=-1$, 95%CI = [-1.7 -0.6]; Self vs LO: $t(19)= -2.2$, $p=0.038$, $d=-0.5$, 95% CI=[-1.8 -0.06]; Self vs OO: $t(19)=4.1$, $p<0.0001$, $d=0.9$, 95% CI=[0.9 2.8]), except the one that behaved opportunistically towards themselves and loyally towards others (paired t-test, Self vs OL: $t(19) = 1.6$, $p=0.13$, $d=0.36$, 95% CI=[-0.24 1.7]).

5.4. Summary

We have shown that participants integrate both contextual and social identity factors, and their dynamic interactions when forming representations of others. Participants used both sources of information to inform their own choices about whether to cooperate with or defect against others and when they made predictions about the behaviour of others towards themselves and others. When deciding whether to cooperate with or defect against another player, participants used information in a prospective manner by anticipating other's behaviour. We also showed that in general, predictions were more accurate when interacting directly with another person, while during observational interactions most errors occurred by incorrectly predicting others to defect more frequently than was actually the case.

5.5. Discussion

In this study, we designed a multi-player game to examine what types of information are used to form representations of others. We showed that participants integrated information about the context and social partner and adapted their choices towards others and predictions about them accordingly. Choices and predictions were adapted in tandem with changes observed in the other's behaviour in previous interactions that were due to environmental context such as resource endowment and social factors such as the general tendency to be cooperative or to defect. This shows that participants integrate multiple information sources and their dynamic interactions when forming representations of others.

As was shown in previous studies, participants do not only learn through direct experience, for example by using previous outcomes to inform their future actions, but also by observing other's actions with one another and their outcomes. Previous studies have already shown that people learn about other social agents just as they learn about stimulus-outcome associations (Behrens et al., 2008; Boorman et al., 2013a; Burke et al., 2010; Wittmann et al., 2016b) and that they take character features of the other person into account when learning about them (Garvert et al., 2015; Izuma and Adolphs, 2013). However, here we show that such learning occurs even when participants have opportunities for vicarious as opposed to direct experiences of seeing how cooperatively another agent will behave. We show that participants use information from both direct and observational experiences to predict others' behaviour during subsequent interactions both with themselves and with others. In other words, they extrapolate

from the behaviour another person exhibits towards others to predict how they will behave towards themselves and, in a complementary manner, they extrapolate from how another person behaves towards themselves to predict how they will behave towards others. Sometimes merging information across experiences is beneficial when information obtained in different contexts is consistent and leads to increased learning (Burke et al., 2010). However, it can also lead to suboptimal outcomes when information derived from different experiences is contradictory: here, participants used information from direct experiences to inform their predictions during observational trials, and *vice versa*, even when the information was contradictory across trials. The proportion of inaccurate predictions was higher during observational compared to self-experienced trials, which might suggest that the effect of mergence was asymmetrical across direct and indirect experiences. Although this would be consistent with the notion that information related to oneself is learnt better compared to learning about others (Lockwood et al., 2018), this suggestion still has to be verified in future analyses. Other studies have shown that mergence across social representations can occur between oneself and another player (Wittmann et al., 2016b), while we show here that representations of others are merged across different types of experiences. During future analyses, I will explore the consistency of representations of various social agents to see if they reflect the consistency of their behaviour and the consistency of the participants' behaviour towards them. The similarity between representations of self and other might also be predictive of when participants will predict others' behaviour effectively.

We found that predictions about others differed from choices made towards others and this was particularly true when there was extreme resource inequalities. Decisions to defect were motivated by the payoff schedules, but they were also made prospectively in anticipation of the other players' behaviour. As in their predictions about others, participants defected against the other person when the risk of a loss was larger than the potential gain. However, they also defected against the other person when their potential gain was greater than their risk (in cases where they had less resources than the other person), thereby mirroring the expected behaviour of the other player.

6. CHAPTER: General discussion

6.1. Behavioural flexibility and the exploration/exploitation trade-off

In this thesis, I have presented multiple experiments that aimed to contribute to current knowledge concerning the behavioural and neural mechanisms of behavioural flexibility. Many decision-making paradigms that are used to study behavioural flexibility present a scenario in which one has to weigh the value of the current option – an option that has been the focus of behaviour for some time – with the hypothetical value of counterfactual options – the options that have not recently been chosen. Regardless of the type of decision problem -- a binary neuroeconomic type of decision, a foraging problem or an approach-avoid decision, the explore-exploit dilemma defines the decision problem during which one has to make a decision about whether it is worth forgoing an option that has been taken and which has a relatively well known value to instead select an alternative option with uncertain value and which might turn out to be more profitable in the long-term. This decision process underlies prospective decision-making because it integrates changes in the external (e.g. novelty of options) and internal (e.g. subjective uncertainty about options) environment to match future goals (e.g. maximization of reward). The exploration/exploitation has been one of the key decision-making processes that I investigated in the thesis (Chapter 3,4).

As in previous work (Wilson et al., 2014c), I have made use of time horizons to examine exploratory behaviour. The time horizon is the number of future choice

opportunities that remain with the same set of options. Exploratory choices increased as future opportunities increased when time horizons were longer. When time horizons are long then the knowledge that is gained by exploring alternative options can improve future reward maximization. It was possible to certain that the result really was driven by the number future opportunities because when I manipulated time horizon I took care to that the previous learning were matched across conditions. Hence, these results show that exploratory choices were aimed at reducing uncertainty about unknown options, and thereby capturing processes of directed exploratory behaviour.

This result and conclusion differ from those in a recent account that suggests some exploratory choices in fact result from noise during the learning process (Findling et al., 2019). While I also found that some exploratory decisions appeared to be noise-driven, it was possible to establish that some were guided towards the choices about which participants were most uncertain. The uncertainty-guided exploratory decisions were associated with an uncertainty signal in both vmPFC and dACC but the vmPFC signal was not present during random exploration.

In Chaptesr 3 and 4 I showed that, unlike in other previous studies, both behavioural modes of exploration and exploitation are most effectively described in terms of both their relative value (Daw et al., 2006) and their relative uncertainty. To show this, I used a Bayesian model that allowed extraction of subjective beliefs about the accuracy of an information source (‘accuracy’) and the uncertainty in that belief (‘uncertainty’), both of which impacted on the frequency of exploratory versus exploitative choices across time.

In Chapter 4, I extended the results from binary selections between information sources to examining explicit judgements about one's confidence about how successful the chosen information source was likely to be. Estimates of uncertainty are often derived using careful Bayesian learning analyses but the degree to which they correspond to participants' self-reported and subjective sense of uncertainty has been unclear. I showed that model-based estimates of accuracy and uncertainty not only determined the balance between of exploratory and exploitative behavioural modes of behaviour, but they also determine the subsequent subjective confidence in the choice, regardless of whether it was exploitative or exploratory once it was made. Participants' reported subjective confidences increased with the estimated accuracy in the chosen predictor and decreased with as the estimate of uncertainty associated with that belief increased. This is consistent with previous observations that confidence can be defined as the evaluation of accuracy in one's choice (Lak et al., 2020; Stolyarova et al., 2019). Both retrospective evaluation (Fleming et al., 2018) and the evaluation of counterfactual choices (Boorman et al., 2011; Fouragnan et al., 2019) are crucial determinants of adaptive behaviour.

6.2. Behavioural flexibility during social interactions

We rarely make decisions in social isolation. Instead, we are continuously influenced by other people, which can alter our preferences (Garvert et al., 2015; Izuma and Adolphs, 2013), learning (Behrens et al., 2008; Boorman et al., 2013a; Wittmann et al., 2016b) and decision-making (Bang et al., 2017; Kappes et al., 2020; Will et al., 2017). I first aimed to test whether there are differences in the

exploration and exploitation rates of social compared to non-social information sources. I did not find behavioural differences between conditions in the impact model-based beliefs of uncertainty and accuracy had on binary selections between information sources. However, when comparing participants' explicit, subjective confidence judgments , then differences between social and non-social conditions emerged. Participants were less confident when judging non-social compared to social information sources and the belief uncertainty estimates derived from the Bayesian learning model had a bigger impact in determining the explicit uncertainty judgements when they were of a non-social nature. These differences could be attributed to different performance expectations for social versus non-social predictors. Participants might expect other people, as opposed to non-social stimuli, to exhibit more consistent patterns of performance.

In Chapter 5, I have shown that we form estimates about others based on multiple factors that are dynamically integrated and used to adapt one's choices and predictions about others. As in previous studies, estimates were based on direct experience but also on vicarious observation (Lockwood et al., 2015). Learning was improved when directly interacting with another person as opposed to passively observing interactions between other agents (Lockwood et al., 2018). When other agents consistently engaged in different patterns of behaviour towards the participants and towards other individuals then participants should base their estimates of how the other agent is going to behave towards them on the next occasion solely on the basis of their past, direct interactions. Only these are relevant for predicting what the other agent will do to the participant on the next occasion. However, interestingly, participants were not able to completely separate

their direct experiences with the other agent and the other observations that they had made of the other agent; even what should have been irrelevant interactions that they had had with others exerted an influence on the participants' estimates of what they would do to the participants themselves next. Other studies have shown that mergence across social representations can occur between oneself and another player (Wittmann et al., 2016b) or that one can assume the predispositions of another person (Garvert et al., 2015). The current results complement these previous observations and show that just as self and other representations may merge with one another, so representations of different others merge across different types of experiences.

6.3. Neural correlates of exploration and the unique role of vmPFC

I have identified a set of brain regions that was involved in balancing exploration and exploitation of information sources. I replicated previous results that showed a role for dorsal anterior cingulate cortex and anterior frontal cortex including area 46 (Badre et al., 2012; Kolling et al., 2018; Zajkowski et al., 2017) and frontal pole (Badre et al., 2012; Boorman et al., 2009; Daw et al., 2006; Fleming et al., 2018) during flexible decision-making. These brain areas have been predominantly implicated during behavioural change and adaption. For example, dACC has been associated with representing the value of the counterfactual choices (Boorman et al., 2011; Fouragnan et al., 2019), driving foraging decisions (Hayden et al., 2011a; Kolling et al., 2012; Wittmann et al., 2016a) or driving behavioural change in other contexts (Karlsson et al., 2012; Tervo et al., 2014). Area 46 and frontal pole have been associated with exploratory decisions that were

either defined by the relative value between options (Daw et al., 2006) or by the relative uncertainty between options (Badre et al., 2012; Zajkowski et al., 2017). However, I showed that the difference between exploration and exploitation is best explained in relation to both the relative values and uncertainties of choices and this was apparent both in the behavioural and neural data: vmPFC did not only represent one variable relevant for guiding the current decision, but instead it represented both belief estimates and uncertainties in belief estimates in parallel – that is, we found neural correlates of how accurate participants believed the information source to be ('accuracy') and how uncertain they were in this accuracy estimate ('uncertainty') in vmPFC. This is similar to previous studies that showed multiple representations of value and confidence in vmPFC (De Martino et al., 2012; Lebreton et al., 2015; Lopez-Persem et al., 2020). Previous accounts suggested that confidence is an automatic process or by-product of value representation (Barron et al., 2015; Lebreton et al., 2015), here however I show how it exerts an influence on behaviour. Second, I showed that which decision variables are represented in vmPFC depends on the current behavioural mode – exploration, exploitation, or the transition between both. This implies that a single brain area can represent multiple choice-relevant decision variables that vary according to changes between exploratory and exploitative behavioural modes. An analogous interpretation has recently put forward in the context of two subcortical areas: amygdala and ventral striatum (Costa et al., 2019). Both amygdala and ventral striatum have previously been linked to reward-guided exploitative behaviour and Costa and colleagues used electrophysiological data from macaques to confirm that neural activity in both areas reflected the reward value of

choices during exploitation. In addition, however, they show that activity in both areas was present during novelty-driven exploration and then activity reflected the information bonus of an option. In sum, these results extend our understanding of the types of information that vmPFC represents during decision-making.

6.4. Common and distinct representations of confidence across social and non-social contexts

A different set of brain areas was involved when making an evaluation about a single choice compared to the prior binary selection between them. Activity in perigenual anterior cingulate cortex (pgACC) and ventral striatum correlated positively with participant confidence in the likely success of the predictor's performance. This finding was independent from the framing of the information source -- whether it was of a social or non-social kind -- suggesting a common role for pgACC in both types of processes. This suggestion is further supported by previous studies that have reported a role for pgACC during social (Will et al., 2017; Wittmann et al., 2016b) and non-social decision-making studies (Bang and Fleming, 2018).

There was one brain area that uniquely encoded social information: the right posterior temporoparietal junction (pTPJ). Its activity covaried with the reported confidence judgment about another agent, which could be further decomposed into, amongst other factors, a belief of how accurately the social predictor would perform. Additional analyses are needed to investigate the link between right pTPJ and areas that underlie common processes, for example by investigating the connectivity between them. Further, it would be interesting to use multivariate

analysis approach to examine whether activity in brain areas that are activated in both social and non-social situations, such as vmPFC and dACC are similar in terms of the representations they form.

In sum, I have used a combination of neuroimaging and computational methods to investigate learning and decision-making in a social and non-social context. I have identified computational mechanisms that impacted the trade-off between exploration and exploitation. Medial and lateral frontal areas were implicated during exploratory decisions, while vmPFC had a unique role in representing multiple decision variables depending on the behavioural mode of exploration, exploitation, or the period of transition between them. I showed that, at neural level, binary selections between information sources differ from the evaluation of a single source. Moreover, I have shown that there are neural level differences between situations in which participants learn about social as opposed to non-social predictors. The estimates we form about others are not only influenced by the social identity of the people we interact with but also by other contextual factors, such as the manner in which we learn about others and whether this is through direct experience or vicarious observation.

Bibliography

Akaishi, R., Kolling, N., Brown, J.W., and Rushworth, M. (2016). Neural Mechanisms of Credit Assignment in a Multicue Environment. J. Neurosci. *36*, 1096–1112.

Amiez, C., and Petrides, M. (2017). Functional rostro-caudal gradient in the human posterior lateral frontal cortex. Brain Struct Funct.

Amiez, C., Joseph, J.P., and Procyk, E. (2006). Reward Encoding in the Monkey Anterior Cingulate Cortex. Cerebral Cortex *16*, 1040–1055.

Amiez, C., Neveu, R., Warrot, D., Petrides, M., Knoblauch, K., and Procyk, E. (2013). The Location of Feedback-Related Activity in the Midcingulate Cortex Is Predicted by Local Morphology. Journal of Neuroscience *33*, 2217–2228.

Amodio, D.M., and Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. Nat Rev Neurosci *7*, 268–277.

Amunts, K., Schleicher, A., Rgel, U.B., Mohlberg, H., Uylings, H.B.M., and Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. 23.

Apps, M.A.J., Rushworth, M.F.S., and Chang, S.W.C. (2016). The Anterior Cingulate Gyrus and Social Cognition: Tracking the Motivation of Others. Neuron *90*, 692–707.

Badre, D., Doll, B.B., Long, N.M., and Frank, M.J. (2012). Rostrolateral Prefrontal Cortex and Individual Differences in Uncertainty-Driven Exploration. Neuron *73*, 595–607.

Bang, D., and Fleming, S.M. (2018). Distinct encoding of decision confidence in human medial prefrontal cortex. PNAS *115*, 6082–6087.

Bang, D., Aitchison, L., Moran, R., Herce Castanon, S., Rafiee, B., Mahmoodi, A., Lau, J.Y.F., Latham, P.E., Bahrami, B., and Summerfield, C. (2017). Confidence matching in group decision-making. Nat Hum Behav *1*, 0117.

Barron, H.C., Garvert, M.M., and Behrens, T.E.J. (2015). Reassessing VMPFC: full of

confidence? Nat Neurosci *18*, 1064–1066.

Basten, U., Biele, G., Heekeren, H.R., and Fiebach, C.J. (2010). How the brain integrates costs and benefits during decision making. PNAS *107*, 21767–21772.

Beckmann, M., Johansen-Berg, H., and Rushworth, M.F.S. (2009). Connectivity-Based Parcellation of Human Cingulate Cortex and Its Relation to Functional Specialization. Journal of Neuroscience *29*, 1175–1190.

Behrens, T.E.J., Woolrich, M.W., Walton, M.E., and Rushworth, M.F.S. (2007). Learning the value of information in an uncertain world. Nature Neuroscience *10*, 1214–1221.

Behrens, T.E.J., Hunt, L.T., Woolrich, M.W., and Rushworth, M.F.S. (2008). Associative learning of social value. Nature *456*, 245–249.

Berger-Tal, O., and Avgar, T. (2012). The Glass is Half-Full: Overestimating the Quality of a Novel Environment is Advantageous. PLoS ONE *7*, e34578.

Bernacchia, A., Seo, H., Lee, D., and Wang, X.-J. (2011). A reservoir of time constants for memory traces in cortical neurons. Nat Neurosci *14*, 366–372.

Bickart, K.C., Wright, C.I., Dautoff, R.J., Dickerson, B.C., and Barrett, L.F. (2011). Amygdala volume and social network size in humans. Nat Neurosci *14*, 163–164.

Bijsterbosch, J.D., Woolrich, M.W., Glasser, M.F., Robinson, E.C., Beckmann, C.F., Van Essen, D.C., Harrison, S.J., and Smith, S.M. (2018). The relationship between spatial configuration and functional connectivity of brain regions. ELife *7*, e32992.

Blanchard, T.C., Hayden, B.Y., and Bromberg-Martin, E.S. (2015). Orbitofrontal Cortex Uses Distinct Codes for Different Choice Attributes in Decisions Motivated by Curiosity. Neuron *85*, 602–614.

Bludau, S., Eickhoff, S.B., Mohlberg, H., Caspers, S., Laird, A.R., Fox, P.T., Schleicher, A., Zilles, K., and Amunts, K. (2014). Cytoarchitecture, probability maps and functions of

the human frontal pole. NeuroImage *93*, 260–275.

Boorman, E.D., Behrens, T.E.J., Woolrich, M.W., and Rushworth, M.F.S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. Neuron *62*, 733–743.

Boorman, E.D., Behrens, T.E., and Rushworth, M.F. (2011). Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. PLOS Biology *9*, e1001093.

Boorman, E.D., O'Doherty, J.P., Adolphs, R., and Rangel, A. (2013a). The Behavioral and Neural Mechanisms Underlying the Tracking of Expertise. Neuron *80*, 1558–1571.

Boorman, E.D., Rushworth, M.F., and Behrens, T.E. (2013b). Ventromedial Prefrontal and Anterior Cingulate Cortex Adopt Choice and Default Reference Frames during Sequential Multi-Alternative Choice. J. Neurosci. *33*, 2242–2253.

Botvinick, M.M., Carter, C.S., Braver, T.S., Barch, D.M., and Cohen, J.D. (2001). Conflict Monitoring and Cognitive Control. 29.

Botvinick, M.M., Cohen, J.D., and Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences *8*, 539–546.

Bromberg-Martin, E.S., and Hikosaka, O. (2009). Midbrain Dopamine Neurons Signal Preference for Advance Information about Upcoming Rewards. Neuron *63*, 119–126.

Bromberg-Martin, E.S., and Hikosaka, O. (2011). Lateral habenula neurons signal errors in the prediction of reward information. Nat Neurosci *14*, 1209–1216.

Bromberg-Martin, E.S., and Sharot, T. (2020). The Value of Beliefs. Neuron *106*, 561–565.

Burke, C.J., Tobler, P.N., Baddeley, M., and Schultz, W. (2010). Neural mechanisms of observational learning. Proceedings of the National Academy of Sciences *107*, 14431–14436.

Cai, X., and Padoa-Schioppa, C. (2012). Neuronal Encoding of Subjective Value in Dorsal and Ventral Anterior Cingulate Cortex. Journal of Neuroscience *32*, 3791–3808.

Campbell-Meiklejohn, D., Simonsen, A., Frith, C.D., and Daw, N.D. (2017). Independent Neural Computation of Value from Other People's Confidence. J. Neurosci. *37*, 673–684.

Carmichael, S.T., and Price, J.L. (1994). Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. J. Comp. Neurol. *346*, 366–402.

Carter, R.M., Bowling, D.L., Reeck, C., and Huettel, S.A. (2012). A Distinct Role of the Temporal-Parietal Junction in Predicting Socially Guided Decisions. Science *337*, 109–111.

Chang, S.W.C., Gariépy, J.-F., and Platt, M.L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. Nat Neurosci *16*, 243–250.

Charnov, E.L. (1976). Optimal foraging, the marginal value theorem. Theoretical Population Biology *9*, 129–136.

Charpentier, C.J., Bromberg-Martin, E.S., and Sharot, T. (2018). Valuation of knowledge and ignorance in mesolimbic reward circuitry. Proc Natl Acad Sci USA *115*, E7255–E7264.

Chau, B.K.H., Kolling, N., Hunt, L.T., Walton, M.E., and Rushworth, M.F.S. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. Nature Neuroscience *17*, 463.

Cohen, J.D., Dunbar, K., and McClelland, J.L. (1989). The Artificial Intelligence and Psychology; Project. 80.

Corrado, G., and Doya, K. (2007). Understanding Neural Coding through the Model-Based Analysis of Decision Making. Journal of Neuroscience *27*, 8178–8180.

Costa, V.D., Mitz, A.R., and Averbeck, B.B. (2019). Subcortical Substrates of Explore-

Exploit Decisions in Primates. Neuron *103*, 533-545.e5.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.

Dayan, P., and Sejnowski, T.J. (1996). Exploration bonuses and dual control. Mach Learn *25*, 5–22.

Dayan, P., Kakade, S., and Montague, P.R. (2000). Learning and selective attention. Nat Neurosci *3*, 1218–1223.

De Dreu, C.K.W., and Kret, M.E. (2016). Oxytocin Conditions Intergroup Relations Through Upregulated In-Group Empathy, Cooperation, Conformity, and Defense. Biological Psychiatry *79*, 165–173.

De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2012). Confidence in value-based choice. Nature Neuroscience *16*, 105.

De Martino, B., Fleming, S.M., Garrett, N., and Dolan, R.J. (2013). Confidence in value-based choice. Nat Neurosci *16*, 105–110.

De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., and Love, B.C. (2017). Social Information Is Integrated into Value and Confidence Judgments According to Its Reliability. J. Neurosci. *37*, 6066–6074.

Dunbar, R.I.M. (1998). The social brain hypothesis. 13.

Ebitz, R.B., and Platt, M.L. (2015). Neuronal Activity in Primate Dorsal Anterior Cingulate Cortex Signals Task Conflict and Predicts Adjustments in Pupil-Linked Arousal. Neuron *85*, 628–640.

Fellows, L.K. (2004). Different Underlying Impairments in Decision-making Following Ventromedial and Dorsolateral Frontal Lobe Damage in Humans. Cerebral Cortex *15*, 58–63.

Findling, C., Skvortsova, V., Dromnelle, R., Palminteri, S., and Wyart, V. (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. Nat Neurosci.

FitzGerald, T.H.B., Seymour, B., and Dolan, R.J. (2009). The Role of Human Orbitofrontal Cortex in Value Comparison for Incommensurable Objects. J. Neurosci. *29*, 8388–8395.

Fleming, S.M., Putten, E.J., and Daw, N.D. (2018). Neural mediators of changes of mind about perceptual decisions. Nature Neuroscience *21*, 617–624.

Fouragnan, E.F., Chau, B.K.H., Folloni, D., Kolling, N., Verhagen, L., Klein-Flügge, M., Tankelevitch, L., Papageorgiou, G.K., Aubry, J.-F., Sallet, J., et al. (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. Nature Neuroscience *22*, 797–808.

Garrett, N., González-Garzón, A.M., Foulkes, L., Levita, L., and Sharot, T. (2018). Updating Beliefs under Perceived Threat. J. Neurosci. *38*, 7901–7911.

Garvert, M.M., Moutoussis, M., Kurth-Nelson, Z., Behrens, T.E.J., and Dolan, R.J. (2015). Learning-Induced Plasticity in Medial Prefrontal Cortex Predicts Preference Malleability. Neuron *85*, 418–428.

Gold, J.I., and Shadlen, M.N. (2007). The Neural Basis of Decision Making. Annu. Rev. Neurosci. *30*, 535–574.

Hampton, A.N., Bossaerts, P., and O'Doherty, J.P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. Proceedings of the National Academy of Sciences *105*, 6741–6746.

Hare, T.A., Schultz, W., Camerer, C.F., O'Doherty, J.P., and Rangel, A. (2011). Transformation of stimulus value signals into motor commands during simple choice. Proceedings of the National Academy of Sciences *108*, 18120–18125.

Hayden, B.Y., Pearson, J.M., and Platt, M.L. (2011a). Neuronal basis of sequential foraging decisions in a patchy environment. Nature Neuroscience *14*, 933–939.

Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., and Platt, M.L. (2011b). Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. Journal of Neuroscience *31*, 4178–4187.

Hertz, U., Palminteri, S., Brunetti, S., Olesen, C., Frith, C.D., and Bahrami, B. (2017). Neural computations underpinning the strategic management of influence in advice giving. Nat Commun *8*, 2191.

Hill, C.A., Suzuki, S., Polania, R., Moisa, M., O'Doherty, J.P., and Ruff, C.C. (2017). A causal account of the brain network computations underlying strategic social behavior. Nature Neuroscience *20*, 1142–1149.

Huettel, S.A., Song, A.W., and McCarthy, G. (2004). Functional Magnetic Resonance Imaging, Second Edition. 8.

Hunt, L.T., and Hayden, B.Y. (2017). A distributed, hierarchical and recurrent framework for reward-based choice. Nat Rev Neurosci *18*, 172–182.

Hunt, L.T., Kolling, N., Soltani, A., Woolrich, M.W., Rushworth, M.F., and Behrens, T.E. (2012). Mechanisms underlying cortical activity during value-guided choice. Nat Neurosci *15*, 470-S3.

Hunt, L.T., Malalasekera, W.M.N., de Berker, A.O., Miranda, B., Farmer, S.F., Behrens, T.E.J., and Kennerley, S.W. (2018). Triple dissociation of attention and decision computations across prefrontal cortex. Nature Neuroscience *21*, 1471–1481.

Iigaya, K., Hauser, T.U., Kurth-Nelson, Z., O'Doherty, J.P., Dayan, P., and Dolan, R.J. (2020). The value of what's to come: Neural mechanisms coupling prediction error and the utility of anticipation. Sci. Adv. *6*, eaba3828.

Ito, S. (2003). Performance Monitoring by the Anterior Cingulate Cortex During Saccade Countermanding. Science *302*, 120–122.

Izuma, K., and Adolphs, R. (2013). Social Manipulation of Preference in the Human Brain. Neuron *78*, 563–573.

Jenkinson, M., and Smith, S. (2001). A global optimisation method for robust affine registration of brain images. Medical Image Analysis *5*, 143–156.

Johnson, K.O., Hsiao, S.S., and Yoshioka, T. (2002). Review: Neural Coding and the Basic Law of Psychophysics. Neuroscientist *8*, 111–121.

Kaanders, P., Nili, H., O'Reilly, J.X., and Hunt, L.T. (2020). Medial frontal cortex activity predicts information sampling in economic choice. 31.

Kable, J.W., and Glimcher, P.W. (2007). The neural correlates of subjective value during intertemporal choice. Nat Neurosci *10*, 1625–1633.

Kalman (1960). A New Approach to Linear Filtering and Prediction Problems. Journal of Fluids Engineering *82*, 35–45.

Kappes, A., Harvey, A.H., Lohrenz, T., Montague, P.R., and Sharot, T. (2020). Confirmation bias in the utilization of others' opinion strength. Nat Neurosci *23*, 130–137.

Karlsson, M.P., Tervo, D.G.R., and Karpova, A.Y. (2012). Network Resets in Medial Prefrontal Cortex Mark the Onset of Behavioral Uncertainty. Science *338*, 135–139.

Kennerley, S.W., Walton, M.E., Behrens, T.E.J., Buckley, M.J., and Rushworth, M.F.S. (2006). Optimal decision making and the anterior cingulate cortex. Nat Neurosci *9*, 940–947.

Keysers, C., Gazzola, V., and Wagenmakers, E.-J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. Nat Neurosci *23*, 788–799.

Khamassi, M., Quilodran, R., Enel, P., Dominey, P.F., and Procyk, E. (2015). Behavioral

Regulation and the Modulation of Information Coding in the Lateral Prefrontal and Cingulate Cortex. Cereb. Cortex *25*, 3197–3218.

Kiani, R., and Shadlen, M.N. (2009). Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. Science *324*, 759–764.

Kiani, R., Corthell, L., and Shadlen, M.N. (2014). Choice Certainty Is Informed by Both Evidence and Decision Time. Neuron *84*, 1329–1342.

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

Kobayashi, K., and Hsu, M. (2019). Common neural code for reward and information value. Proc Natl Acad Sci USA *116*, 13061–13066.

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

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

Kolling, N., Wittmann, M.K., Behrens, T.E.J., Boorman, E.D., Mars, R.B., and Rushworth, M.F.S. (2016a). Value, search, persistence and model updating in anterior cingulate cortex. Nature Neuroscience *19*, 1280.

Kolling, N., Behrens, T., Wittmann, M., and Rushworth, M. (2016b). Multiple signals in anterior cingulate cortex. Current Opinion in Neurobiology *37*, 36–43.

Kolling, N., Scholl, J., Chekroud, A., Trier, H.A., and Rushworth, M.F.S. (2018). Prospection, Perseverance, and Insight in Sequential Behavior. Neuron *99*, 1069-1082.e7.

Kriegeskorte, N., and Douglas, P.K. (2018). Cognitive computational neuroscience. Nature

Neuroscience 1.

Lak, A., Costa, G.M., Romberg, E., Koulakov, A.A., Mainen, Z.F., and Kepecs, A. (2014). Orbitofrontal Cortex Is Required for Optimal Waiting Based on Decision Confidence. Neuron *84*, 190–201.

Lak, A., Okun, M., Moss, M.M., Gurnani, H., Farrell, K., Wells, M.J., Reddy, C.B., Kepecs, A., Harris, K.D., and Carandini, M. (2020). Dopaminergic and Prefrontal Basis of Learning from Sensory Confidence and Reward Value. Neuron *105*, 700-711.e6.

Le Heron, C., Kolling, N., Plant, O., Kienast, A., Janska, R., Ang, Y.-S., Fallon, S., Husain, M., and Apps, M.A.J. (2020). Dopamine Modulates Dynamic Decision-Making during Foraging. J. Neurosci. *40*, 5273–5282.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., and Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. Neuron *64*, 431–439.

Lebreton, M., Abitbol, R., Daunizeau, J., and Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. Nature Neuroscience *18*, 1159.

Leong, Y.C., Radulescu, A., Daniel, R., DeWoskin, V., and Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron *93*, 451–463.

Levy, I., Snell, J., Nelson, A.J., Rustichini, A., and Glimcher, P.W. (2010). Neural Representation of Subjective Value Under Risk and Ambiguity. Journal of Neurophysiology *103*, 1036–1047.

Lewis, P.A., Rezaie, R., Brown, R., Roberts, N., and Dunbar, R.I.M. (2011). Ventromedial prefrontal volume predicts understanding of others and social network size. NeuroImage *57*, 1624–1629.

Lim, S.-L., O'Doherty, J.P., and Rangel, A. (2011). The Decision Value Computations in the vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention. J. Neurosci. *31*, 13214–13223.

Lockwood, P.L., and Wittmann, M.K. (2018). Ventral anterior cingulate cortex and social decision-making. Neuroscience & Biobehavioral Reviews *92*, 187–191.

Lockwood, Patricia.L., Apps, M.A.J., Roiser, J.P., and Viding, E. (2015). Encoding of Vicarious Reward Prediction in Anterior Cingulate Cortex and Relationship with Trait Empathy. J. Neurosci. *35*, 13720–13727.

Lockwood, P.L., Wittmann, M.K., Apps, M.A.J., Klein-Flügge, M.C., Crockett, M.J., Humphreys, G.W., and Rushworth, M.F.S. (2018). Neural mechanisms for learning self and other ownership. Nat Commun *9*, 4747.

Loewenstein (1994). The psychology of curiosity: A review and reinterpretation (Psychological Bulletin).

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *412*, 8.

Lopez-Persem, A., Domenech, P., and Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. ELife *5*.

Lopez-Persem, A., Verhagen, L., Amiez, C., Petrides, M., and Sallet, J. (2019). The Human Ventromedial Prefrontal Cortex: Sulcal Morphology and Its Influence on Functional Organization. J. Neurosci. *39*, 3627–3639.

Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., et al. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. Nat Neurosci *23*, 664–675.

Mackey, S., and Petrides, M. (2010). Quantitative demonstration of comparable

architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains: Comparable cortical areas in human and macaque. European Journal of Neuroscience *32*, 1940–1950.

Mackey, S., and Petrides, M. (2014). Architecture and morphology of the human ventromedial prefrontal cortex. Eur J Neurosci *40*, 2777–2796.

Mackintosh, N.J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. Psychological Review *82*, 276–298.

Mansouri, F.A., Koechlin, E., Rosa, M.G.P., and Buckley, M.J. (2017). Managing competing goals — a key role for the frontopolar cortex. Nat Rev Neurosci *18*, 645–657.

Mars, R.B., Sallet, J., Neubert, F.-X., and Rushworth, M.F.S. (2013). Connectivity profiles reveal the relationship between brain areas for social cognition in human and monkey temporoparietal cortex. PNAS *110*, 10806–10811.

Matsumoto, M., Matsumoto, K., Abe, H., and Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci *10*, 647–656.

Mcfadden, D. (2001). Economic Choices. THE AMERICAN ECONOMIC REVIEW *91*, 28.

McGinty, V.B., Rangel, A., and Newsome, W.T. (2016). Orbitofrontal Cortex Value Signals Depend on Fixation Location during Free Viewing. Neuron *90*, 1299–1311.

Meder, D., Kolling, N., Verhagen, L., Wittmann, M.K., Scholl, J., Madsen, K.H., Hulme, O.J., Behrens, T.E.J., and Rushworth, M.F.S. (2017). Simultaneous representation of a spectrum of dynamically changing value estimates during decision making. Nat Commun *8*, 1942.

Mukamel, R. (2005). Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. Science *309*, 951–954.

Nakamura, K., Roesch, M.R., and Olson, C.R. (2005). Neuronal Activity in Macaque SEF and ACC During Performance of Tasks Involving Conflict. Journal of Neurophysiology *93*, 884–908.

Neubert, F.-X., Mars, R.B., Thomas, A.G., Sallet, J., and Rushworth, M.F.S. (2014). Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex. Neuron *81*, 700–713.

Neubert, F.-X., Mars, R.B., Sallet, J., and Rushworth, M.F.S. (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. PNAS *112*, E2695–E2704.

von Neumann, J., and Morgenstern, O. (1944). Theory of Games and Economic Behavior.

Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., and Malach, R. (2007). Coupling between Neuronal Firing Rate, Gamma LFP, and BOLD fMRI Is Related to Interneuronal Correlations. Current Biology *17*, 1275–1285.

Noonan, M.P., Walton, M.E., Behrens, T.E.J., Sallet, J., Buckley, M.J., and Rushworth, M.F.S. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. Proceedings of the National Academy of Sciences *107*, 20547–20552.

Noonan, M.P., Kolling, N., Walton, M.E., and Rushworth, M.F.S. (2012). Re-evaluating the role of the orbitofrontal cortex in reward and reinforcement: Re-evaluating the OFC. European Journal of Neuroscience *35*, 997–1010.

Noonan, M.P., Sallet, J., Mars, R.B., Neubert, F.X., O'Reilly, J.X., Andersson, J.L., Mitchell, A.S., Bell, A.H., Miller, K.L., and Rushworth, M.F.S. (2014). A Neural Circuit Covarying with Social Hierarchy in Macaques. PLoS Biol *12*, e1001940.

Noonan, M.P., Chau, B.K.H., Rushworth, M.F.S., and Fellows, L.K. (2017). Contrasting

Effects of Medial and Lateral Orbitofrontal Cortex Lesions on Credit Assignment and Decision-Making in Humans. J. Neurosci. *37*, 7023–7035.

Öngür, D., Ferry, A.T., and Price, J.L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. J. Comp. Neurol. *460*, 425–449.

O'Reilly, J.X., Jbabdi, S., and Behrens, T.E.J. (2012). How can a Bayesian approach inform neuroscience? European Journal of Neuroscience *35*, 1169–1179.

O'Reilly, J.X., Schüffelgen, U., Cuell, S.F., Behrens, T.E.J., Mars, R.B., and Rushworth, M.F.S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. PNAS *110*, E3660–E3669.

Papageorgiou, G.K., Sallet, J., Wittmann, M.K., Chau, B.K.H., Schüffelgen, U., Buckley, M.J., and Rushworth, M.F.S. (2017). Inverted activity patterns in ventromedial prefrontal cortex during value-guided decision-making in a less-is-more task. Nature Communications *8*, 1886.

Passingham and Wise (2012). The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight (Oxford University: Oxford University Press).

Payzan-LeNestour, E., and Bossaerts, P. (2011). Risk, Unexpected Uncertainty, and Estimation Uncertainty: Bayesian Learning in Unstable Settings. PLoS Comput Biol *7*, e1001048.

Pearce, J.M., and Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychological Review *87*, 532–552.

Petrides, M., Tomaiuolo, F., Yeterian, E.H., and Pandya, D.N. (2012). The prefrontal cortex: Comparative architectonic organization in the human and the macaque monkey brains. Cortex *48*, 46–57.

Philiastides, M.G., Biele, G., and Heekeren, H.R. (2010). A mechanistic account of value computation in the human brain. PNAS *107*, 9430–9435.

Picard, N., and Strick, P.L. (1996). Motor Areas of the Medial Wall: A Review of Their Location and Functional Activation. Cereb Cortex *6*, 342–353.

Picard, N., and Strick, P.L. (2001). Imaging the premotor areas. Current Opinion in Neurobiology *11*, 663–672.

Plassmann, H., O'Doherty, J., Shiv, B., and Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. Proceedings of the National Academy of Sciences *105*, 1050–1054.

Quilodran, R., Rothé, M., and Procyk, E. (2008). Behavioral Shifts and Action Valuation in the Anterior Cingulate Cortex. Neuron *57*, 314–325.

Reader, S.M., Hager, Y., and Laland, K.N. (2011). The evolution of primate general and cultural intelligence. Phil. Trans. R. Soc. B *366*, 1017–1027.

Rescrola, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory.

Rich, E.L., and Wallis, J.D. (2016). Decoding subjective decisions from orbitofrontal cortex. Nat Neurosci *19*, 973–980.

Rilling, J.K., Gutman, D.A., Zeh, T.R., Pagnoni, G., Berns, G.S., and Kilts, C.D. (2002). A Neural Basis for Social Cooperation. Neuron *35*, 395–405.

Rudebeck, P.H. (2006). A Role for the Macaque Anterior Cingulate Gyrus in Social Valuation. Science *313*, 1310–1312.

Ruff, C.C., and Fehr, E. (2014). The neurobiology of rewards and values in social decision making. Nat Rev Neurosci *15*, 549–562.

Ruggeri, K., Alí, S., Berge, M.L., Bertoldo, G., Bjørndal, L.D., Cortijos-Bernabeu, A., Davison, C., Demić, E., Esteban-Serna, C., Friedemann, M., et al. (2020). Replicating patterns of prospect theory for decision under risk. Nat Hum Behav *4*, 622–633.

Rushworth, M.F.S., and Behrens, T.E.J. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. Nature Neuroscience *11*, 389–397.

Rushworth, M., Walton, M., Kennerley, S., and Bannerman, D. (2004). Action sets and decisions in the medial frontal cortex. Trends in Cognitive Sciences *8*, 410–417.

Rushworth, M.F.S., Hadland, K.A., Gaffan, D., and Passingham, R.E. (2003). The Effect of Cingulate Cortex Lesions on Task Switching and Working Memory. Journal of Cognitive Neuroscience *15*, 338–353.

Rushworth, M.F.S., Noonan, M.P., Boorman, E.D., Walton, M.E., and Behrens, T.E. (2011). Frontal Cortex and Reward-Guided Learning and Decision-Making. Neuron *70*, 1054–1069.

Sallet, J., Mars, R.B., Noonan, M.P., Andersson, J.L., O'Reilly, J.X., Jbabdi, S., Croxson, P.L., Jenkinson, M., Miller, K.L., and Rushworth, M.F.S. (2011). Social Network Size Affects Neural Circuits in Macaques. Science *334*, 697–700.

Sanders, J.I., Hangya, B., and Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. Neuron *90*, 499–506.

Saxe, R. (2006). Uniquely human social cognition. Current Opinion in Neurobiology *16*, 235–239.

Scholl, J., and Klein-Flügge, M. (2017). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. Behavioural Brain Research.

Schultz, W., Dayan, P., and Montague, P.R. (1997). A Neural Substrate of Prediction and

Reward. Science *275*, 8.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neuroscience & Biobehavioral Reviews *42*, 9–34.

Schurz, M., Tholen, M.G., Perner, J., Mars, R.B., and Sallet, J. (2017). Specifying the brain anatomy underlying temporo-parietal junction activations for theory of mind: A review using probabilistic atlases from different imaging modalities: Brain Anatomy Underlying Temporo-Parietal Junction. Hum. Brain Mapp. *38*, 4788–4805.

Seo, H., and Lee, D. (2007). Temporal Filtering of Reward Signals in the Dorsal Anterior Cingulate Cortex during a Mixed-Strategy Game. Journal of Neuroscience *27*, 8366–8377.

Shadlen, M.N., and Kiani, R. (2013). Decision Making as a Window on Cognition. Neuron *80*, 791–806.

Sharot, T., and Garrett, N. (2016). Forming Beliefs: Why Valence Matters. Trends in Cognitive Sciences *20*, 25–33.

Sharot, T., and Sunstein, C.R. (2020). How people decide what they want to know. Nat Hum Behav *4*, 14–19.

Sharot, T., Korn, C.W., and Dolan, R.J. (2011). How unrealistic optimism is maintained in the face of reality. Nat Neurosci *14*, 1475–1479.

Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. Neuron *79*, 217–240.

Sliwa, J., and Freiwald, W.A. (2017). A dedicated network for social interaction processing in the primate brain. Science *356*, 745–749.

Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al. (2004). Advances

in functional and structural MR image analysis and implementation as FSL. Neuroimage *23 Suppl 1*, S208-219.

Stephen and Krebs (1986). Foraging theory.

Stoll, F.M., Fontanier, V., and Procyk, E. (2016). Specific frontal neural dynamics contribute to decisions to check. Nature Communications *7*, 11990.

Stolyarova, A., Rakhshan, M., Hart, E.E., O'Dell, T.J., Peters, M.A.K., Lau, H., Soltani, A., and Izquierdo, A. (2019). Contributions of anterior cingulate cortex and basolateral amygdala to decision confidence and learning under uncertainty. Nat Commun *10*, 4704.

Sutton and Barto (1998). Reinforcement learning: An introduction (Cambridge, Massachusetts: MIT Press).

Tervo, D.G.R., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., and Karpova, A.Y. (2014). Behavioral Variability through Stochastic Choice and Its Gating by Anterior Cingulate Cortex. Cell *159*, 21–32.

Trudel, N., Scholl, J., Klein-Flügge, M.C., Fouragnan, E., Tankelevitch, L., Wittmann, M.K., and Rushworth, M.F.S. (2020). Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. Nat Hum Behav.

Vogt, B.A., and Paxinos, G. (2014). Cytoarchitecture of mouse and rat cingulate cortex with human homologies. Brain Struct Funct *219*, 185–192.

Walker, A.E. (1940). A cytoarchitectural study of the prefrontal area of the macaque monkey. J. Comp. Neurol. *73*, 59–86.

Watkins, C.J., and Dayan, P. (1992). Q-learning. *8*, 279–292.

Will, G.-J., Rutledge, R.B., Moutoussis, M., and Dolan, R.J. (2017). Neural and computational processes underlying dynamic changes in self-esteem. ELife *6*.

Wilson, R.C., Geana, A., White, J.M., Ludvig, E.A., and Cohen, J.D. (2014a). Humans Use

Directed and Random Exploration to Solve the Explore–Exploit Dilemma. J Exp Psychol Gen *143*, 2074–2081.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., and Niv, Y. (2014b). Orbitofrontal Cortex as a Cognitive Map of Task Space. Neuron *81*, 267–279.

Wilson, R.C., Geana, A., White, J.M., Ludvig, E.A., and Cohen, J.D. (2014c). Humans Use Directed and Random Exploration to Solve the Explore–Exploit Dilemma. J Exp Psychol Gen *143*, 2074–2081.

Wilson, R.C., Bonawitz, E., Costa, V., and Ebitz, B. (2020). Balancing exploration and exploitation with information and randomization (PsyArXiv).

Wittmann, M.K., Kolling, N., Akaishi, R., Chau, B.K.H., Brown, J.W., Nelissen, N., and Rushworth, M.F.S. (2016a). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. Nature Communications *7*, 12327.

Wittmann, M.K., Kolling, N., Faber, N.S., Scholl, J., Nelissen, N., and Rushworth, M.F.S. (2016b). Self-Other Mergence in the Frontal Cortex during Cooperation and Competition. Neuron *91*, 482–493.

Wunderlich, K., Dayan, P., and Dolan, R.J. (2012). Mapping value based planning and extensively trained choice in the human brain. Nature Neuroscience *15*, 786.

Yoshida, W., and Ishii, S. (2006). Resolution of Uncertainty in Prefrontal Cortex. Neuron *50*, 781–789.

Yu, A.J., and Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. Neuron *46*, 681–692.

Zajkowski, W.K., Kossut, M., and Wilson, R.C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration.

Zhu, L., Mathewson, K.E., and Hsu, M. (2012). Dissociable neural representations of

reinforcement and belief prediction errors underlie strategic learning. Proceedings of the National Academy of Sciences *109*, 1419–1424.

Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., and Meyer-Lindenberg, A. (2008). Know Your Place: Neural Processing of Social Hierarchy in Humans. Neuron *58*, 273–283.

rmis