bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

How exerting control over outcomes affects the neural coding of tasks and outcomes

David Wisniewski[1], Birte Forstmann[2], Marcel Brass[1]

1 Department of Experimental Psychology, Ghent University, Belgium

2 Integrative Model-Based Cognitive Neuroscience Research Unit, University of Amsterdam, The Netherlands

Corresponding author

David Wisniewski

Department of Experimental Psychology, Ghent University

Henri Dunantlaan 2

9000 Gent

Tel: +32 (0)9 2649428

Email: david.wisniewski@ugent.be

Conflict of interest: The authors declare not competing financial interests.

Acknowledgements

We would like to thank Anita Tusche, Carlo Reverberi, and Ruth Krebs for valuable discussions on this project. This research was supported by the Research Foundation Flanders (FWO), the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 665501, FWO grant FWO.OPR.2013.0136.01, an ERC StG grant and NWO Vidi grant.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

We make countless choices every day to achieve desirable outcomes. While we often have perfect control over the outcomes of our choices, sometimes control remains low. Here, we investigate the effect of high vs low control over choice outcomes on the neural coding of outcome valuation and the implementation of the means to achieve these outcomes. In a value-based decision-making task, reward outcomes were either contingent on trial-by-trial choices between two different tasks (high control), or were unrelated to these choices (low control). Using fMRI, multivariate pattern analysis, and model-based neuroscience methods, we identified reward representations in a large network including the striatum, dorso-medial prefrontal cortex (dmPFC) and parietal cortex. These representations were amplified when rewards were contingent on subjects' choices. The means to achieve these outcomes were assessed by identifying brain regions encoding tasks during a preparation / maintenance phase, and results highlighted the role of both the dmPFC and parietal cortex in this process. Importantly, outcome contingency did not affect neural coding of tasks. This suggests that controlling choice outcomes selectively affects the neural coding of these outcomes, but has no effect on the means to reach them. Overall, our findings highlight the role of the dmPFC and parietal cortex in processing of value-related and task-related information, linking motivational and control-related processes in the brain. These findings inform current debates on the interaction of motivational and cognitive control processes.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Making decisions is an integral part of our life. Most of these choices are value-based, i.e. they are made with expected outcomes in mind. Value-based choices are made in separate stages: we first evaluate all options, and then select the option with the highest subjective value (Domenech et al., 2018). After implementing the chosen behavior (Rubinstein et al., 2001), predicted and experienced outcomes are compared, and prediction errors are computed (Matsumoto et al., 2007; Daw et al., 2011; Collins et al., 2017). This dopamine-mediated learning signal (Schultz, 2016) indicates the need to update our internal models of action-outcome contingencies (O'Reilly et al., 2013), which then leads to an adaption of future behavior.

This process is modulated by various properties of choice outcomes, e.g. their magnitude (Doya, 2008). However, one crucial aspect has received little attention in the past: to which degree our choices directly control possible outcomes. Clearly, whether or not we believe our choices to directly *cause* their outcomes affects decision-making considerably. If we know that a specific behavior predictably leads to a desired outcome (e.g. hitting a light switch to light up a room), we will choose it more often (Mobbs et al., 2013). If we know that our behavior and desired outcomes are only weakly correlated (e.g. refreshing your Facebook timeline), or not correlated at all, we might not prioritize any specific behavior. Despite this fact, previous research largely focused on high vs low control over behavior (i.e. classical research on agency, Sperduti et al., 2011), but not on high vs low control over its outcomes.

In principle, varying degrees of control of choice outcomes can affect two key processes: outcome valuation and the implementation of chosen behavior. Some previous research in non-human primates demonstrated that control over choice outcomes indeed affects valuation processes in the brain. Choice-contingent rewards elicit different responses in the caudate (Izquierdo et al., 2004) and anterior cingulate cortex (Chudasama et al., 2013), as compared to non-contingent rewards (see also Elliott et al., 2004). Importantly, one might expect similar effects on neural representations of the chosen behavior as well.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

This is due to the simple fact that in order to receive any reward, choosing a behavior is not enough, we need to implement it as intended first. One might expect chosen behaviors to be shielded more strongly against interference if outcomes are contingent on them (Dreisbach and Wenke, 2011), as not performing the behavior as intended is potentially costly. For non-contingent outcomes the need for shielding is lower, as e.g. executing the wrong behavior has no effect on received outcomes (see Waskom et al., 2014 for a related argument, but Botvinick and Cohen, 2014). Previous work demonstrated that implementation of chosen actions, which includes their maintenance and execution, is supported by a brain network including the frontopolar (Soon et al., 2013), lateral prefrontal and parietal cortex (Zhang et al., 2013; Wisniewski et al., 2016; Loose et al., 2017). Some initial evidence suggests that rewarding correct performance indeed enhances neural task representations (Etzel et al., 2016), but this work did not address the issue of varying degrees of control over choice outcomes.

Here, we report an experiment investigating the effects of control over choice outcomes on value-based decision making. We used a value-based decision task to assess the effects of reward contingency (choice-contingent vs. non-contingent rewards) on valuation and, more importantly, on choice implementation. For this purpose, we used a combination of multivariate pattern analysis (MVPA, Haynes, 2015) and model-based neuroscience methods (Forstmann and Wagenmakers, 2015). We first hypothesized that reward contingency affects the neural coding of outcome values in humans, as it does in non-human primates (Izquierdo et al., 2004; Chudasama et al., 2013). We further assessed whether implementation of chosen behavior (i.e. coding of chosen tasks) is similarly affected by contingency. We hypothesized that the lateral prefrontal cortex, and especially the parietal cortex to play a key role in the implementation of chosen behavior. The parietal cortex represents chosen tasks and actions (Wisniewski et al., 2016; Domenech et al., 2018), subjective stimulus and action values (Sugrue, 2004; Kahnt et al., 2014), as well as associations between choice options and their outcomes (Wisniewski et al., 2015a). Using MVPA, we

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

tested whether task representations in these brain regions were enhanced when rewards were choice-contingent vs when they were not.

Materials and Methods

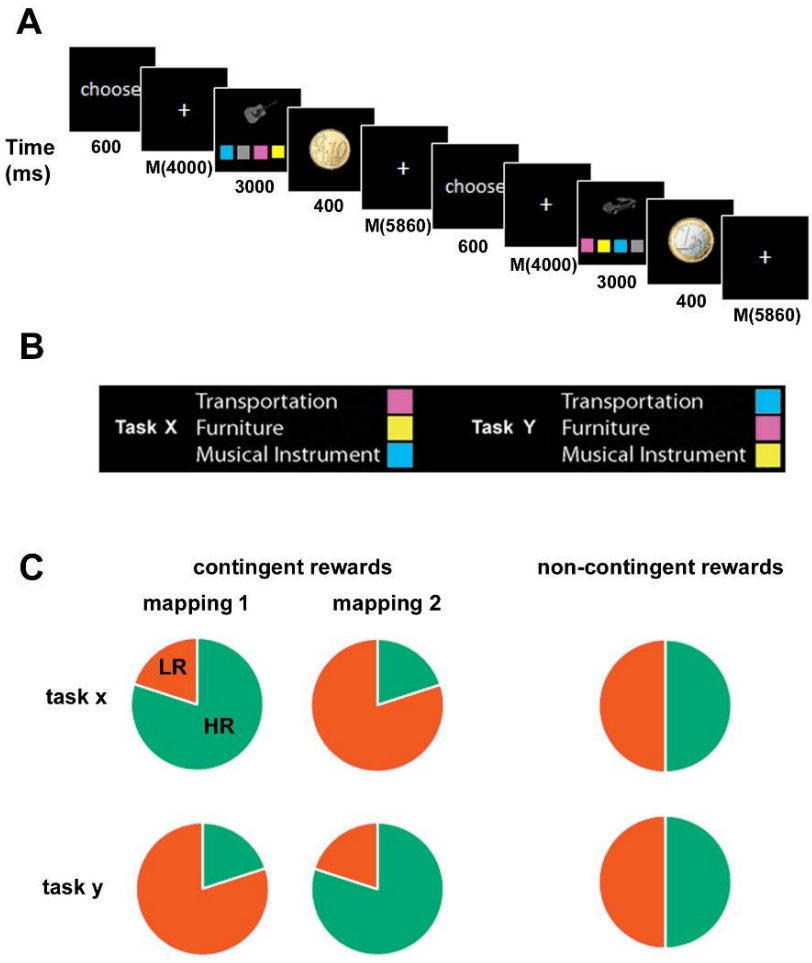


Figure 1. Experimental paradigm. **A.** Trial structure. Each trial started with the cue 'choose' presented on screen. After a variable delay, the task screen was presented for a fixed duration. Reward feedback was presented subsequently after each trial. All trials were separated by variable inter trial intervals. **B.** Tasks. Subjects were instructed to identify the visual object presented on screen, and press a corresponding colored button. The object-category to color mappings are depicted here. Note that the specific mappings were counterbalanced across subjects. Which task was implemented in each trial was chosen freely by the subjects. **C.** Reward contingencies. In contingent (RC) trials, one task always yielded a high reward with a higher probability (80%) than the other task (20%). Which specific task was currently the high-reward task depended on the current task-reward-mapping, which changed according to a probabilistic reversal learning procedure (see Materials and Methods for more details). In non-contingent (NCR) trials, the chance to receive a high and low reward were equal, irrespective of the chosen task.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Participants

A total of 42 subjects participated in this experiment (20 males, 21 females, 1 other). The average age was 22.6 years (min = 18, max = 33 years), 41 subjects were right-handed, one was left-handed. All subjects had normal or corrected-to-normal vision and volunteered to participate. Subjects gave written informed consent and received between 45€ and 55€ for their participation. The experiment was approved by the local ethics committee. Seven subjects showed excessive head movement in the MR scanner (>4mm) and were excluded. All reported analyses were thus performed on a sample of 35 subjects. Despite the fact that the multivariate analyses performed in this experiment (see below for details) show notoriously small effects (Bhandari et al., 2018), we believe to have sufficient statistical power with the given sample size.

Experimental Design

The experiment was programmed using PsychoPy (version 1.85.2, psychopy.org, RRID:SCR_006571, Peirce, 2007)). In each trial, subjects were free to choose between two different tasks, and could either earn a high or a low reward for correct performance. The paradigm is described in more detail below.

Trial structure

Each trial started with the presentation of a fixation cross centrally on-screen for 300ms (Figure 1 A). This was followed by the presentation of a choice cue, the word 'CHOOSE', for 600ms. This cue instructed subjects to freely choose one of the two tasks to perform in this trial. After a variable delay period (2000-6000ms, mean delay duration = 4000ms), the task screen was presented for a total of 3000ms. In this experiment, we used the same tasks as (Wisniewski et al., 2015b), in order to better compare current results to this previous experiment on value-based decision-making. The task screen consisted of a visual object presented centrally on screen (Figure 1 B). This object was picked pseudo-randomly out of a pool of 9 different objects in 3 categories: musical instruments, furniture, means of transportation. Below, 4 colored squares were presented (magenta, yellow, cyan, gray), with the square positions being mapped onto 4 buttons, operated using the left and right index and middle fingers. Subjects were given the option

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to choose which of two stimulus-response-mappings to apply to the presented object. For instance, in task 'X', means of transportation were associated with the magenta, furniture with the yellow, and musical instruments with the cyan button. In task 'Y', means of transportation were associated with the cyan, furniture with the magenta, and musical instruments with the yellow button. Thus, depending on the chosen task and the presented object, one of the colored buttons was correct for each task, and subjects were instructed to react as quickly and accurately as possible. Here, we use the term task to describe a specific link between stimuli and responses, and we do not claim that the cognitive processes required to perform both tasks differed substantially. We inferred subjects' choices from their responses. Note, that the grey button was never task-relevant and was merely included to balance left and right hand responses. Furthermore, the mapping of the colored buttons on screen was pseudo-randomized in each trial, preventing subjects from preparing a specific motor response before the onset of the task screen. The specific stimulus-response-mappings called *task X* and *task Y* were counter-balanced across subjects. Subsequently to the task-screen presentation, subjects were given trial-by-trial reward feedback, by presenting either an image of a 1€ coin (high reward), a 10€cent coin (low reward), or a red circle (no reward). The feedback was presented for 400ms. After a variable inter-trial-interval (4000-14000ms, geometrically distributed, mean duration = 5860ms), the next trial began.

Reward conditions

Subjects were rewarded for correct performance on every trial. There were a total of two different reward conditions: contingent rewards (CR) and non-contingent rewards (NCR). In the NCR condition, the chosen reward in each trial was determined randomly. Irrespective of the chosen task, subjects had a 50% chance of receiving a high and a 50% chance of receiving a low reward (Figure 1 C). Subjects were instructed to choose tasks randomly in this condition, by imagining flipping a coin in their head in each trial (Zhang et al., 2013). In the CR condition, subjects performed a probabilistic reward reversal-learning task, similar to (Hampton and O'Doherty, 2007). In each trial, one task led to a high reward with an 80% and a low reward

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

with a 20% probability (high-reward task, HR). These probabilities were reversed for the other task (low-reward task, LR), e.g., in a specific trial, *task X* might be the HR task, while *task Y* might be the LR task. Subjects were unaware which of the two tasks was the HR task, and needed to learn this from the reward-feedback provided after each trial. Once they chose the HR task on 3 consecutive trials, the mapping of rewards onto tasks reversed with a chance of 25% on each subsequent trial, e.g., whereas before *task X* was the HR and *task Y* the LR task, now *task X* was the LR and *task Y* the HR task. Again, subjects were unaware of this change in reward-contingencies, and needed to learn when such a switch occurred from the reward-feedback provided at the end of each trial.

At the end of the experiment, 15 trials were chosen randomly, and whichever reward was earned in these trials was paid out as a bonus payment to the subjects. One half of these trials was chosen from CR trials, the other from NCR trials, which was communicated to the subjects in order to ensure that both conditions are equally salient. Thus, subjects were motivated to maximize the reward in CR trials, choosing the HR task as often as possible. Given that rewards were randomly chosen in NCR trials, they had no influence over the earned reward in this condition.

This reward manipulation was chosen to manipulate the degree of control subjects had over the outcome of their choices. In CR trials subjects made choices that were directed at earning as much money as they could, by learning the changing reward contingencies and thus controlling reward outcomes. In NCR trials, subjects were unable to control outcomes through their choices, as there were no contingencies to learn. This allowed us to assess effects of control over outcomes on valuation and implementation processes. A second important reason for manipulating reward 'relevance' instead of reward presence (as in Etzel et al., 2016), was that this allowed us to assess specific reward effects on valuation and implementation processes. When contrasting choices in which subjects could earn a reward, with choices in which no reward is present (e.g. Libet et al., 1983; Soon et al., 2008), any difference between these conditions might arise from unspecific processes merely correlated with the presence of reward, like attentional or motor

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

178 preparation (Kahnt et al., 2014). This is mainly because strong differences in expected outcomes
179 immediately trigger these preparatory processes selectively in rewarded trials. In contrast, when rewards
180 are always present, but only sometimes contingent on choices, reward expectations are much more
181 similar across conditions. In fact, if a subject chose tasks randomly in all trials, the expected value would
182 be identical in both reward conditions. Thus, only specific reward-related effects, like the fact that reward
183 outcomes are a relevant factor for making choice only in CR trials, can explain potential differences
184 between CR and NCR trials.

185 ## Design

186 Subjects performed 5 identical runs of this experiment, with 60 trials each. Each run contained 2 blocks
187 with CR and 2 blocks with NCR trials. The length of each block was between 10 and 14 trials, and all trials
188 were all separated by a long and variable ITI. CR and NCR blocks alternated and block order was
189 counterbalanced across runs for each subject. Each block started with either 'Contingent block now
190 starting' or 'Non-contingent block now starting' presented on screen for 5000ms. This mixed blocked and
191 event-related design minimized cross-talk and interference between the reward conditions, and allowed
192 us to estimate cleaner neural signals.

193 Each run also contained 20% (n=12) catch trials. In these trials, subjects were externally cued which task
194 to perform, by presenting the words 'TASK X' or 'TASK Y' instead of the 'CHOOSE' cue. The delay between
195 cue and task execution was 1000ms in these trials. Catch trials were included to prevent subjects from
196 choosing all tasks in a block at its beginning. For instance, in an NCR block, subjects could theoretically
197 decide upon a whole sequence of tasks at the beginning of that block (e.g. X,X,X,Y,X,Y,Y,X,...), and then
198 only implementing that fixed sequence in each trial. In order to encourage subjects to make a conscious
199 choice in each individual trial, catch trials were included. These trials would frequently disrupt any planned
200 sequence of task choices, making such a strategy less feasible. In order to increase the salience of these

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

catch trials, subjects always received a high reward for correct performance. Catch trials were excluded from all analyses.

Furthermore, we ensured that the reward condition was not correlated with any other design variable (target stimulus, delay duration, button mapping, ITI duration), in order to ensure that estimated neural signals were not confounded. Lastly, multivariate pattern analyses can be biased if signal estimates are not based on trials which are IID. Thus we ensured that conditions of the previous trial were not predictive of the current trial, to make each trial as independent of all other trials as possible.

Training session

Subjects were familiarized with the task in a separate training session outside the MR scanner, lasting about 1h10min. Subjects first learned to perform the two tasks, were then instructed about the reward conditions and lastly performed 3 runs of the full experiment (as described above). This training session was performed to minimize learning effects during the MR session, which can be detrimental to multivariate pattern analyses. Training sessions were scheduled between 1-5 days before the MR session. Just before the start of the MR session, subjects performed 10 trials of the task in the MR scanner, in order to familiarize themselves with the novel environment. These trials were not analyzed.

Additional measures

After completing the MR session, subjects filled in multiple questionnaires. They answered custom questions (e.g., How believable were the instructions? How different were the reward conditions? How difficult was making a choice between the two tasks? How difficult was performing the two tasks? Was one task more difficult than the other? At which point in time did you choose the task to perform in each trial?), and the following questionnaires: behavioral inhibition / activation scale (BISBAS, Carver and White, 1994), need for cognition (NFC, Cacioppo et al., 1984), sensitivity to reward / punishment (SPSRQS, Torrubia et al., 2001), and impulsivity (BIS11, Patton et al., 1995). We also acquired pupil dilation data

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

while subjects performed the experiment in the MR scanner. Pupil dilation data is not the focus of the current paper, and is not reported.

Image acquisition

fMRI data was collected using a 3T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany), with a standard thirty-two-channel radio-frequency head coil. A 3D high-resolution anatomical image of the whole brain was acquired for co-registration and normalization of the functional images, using a T1-weighted MPRAGE sequence (TR = 2250 ms, TE = 4.18 ms, TI = 900 ms, acquisition matrix = 256 × 256, FOV = 256 mm, flip angle = 9°, voxel size = 1 × 1 × 1 mm). Furthermore, a field map was acquired for each participant, in order to correct for magnetic field inhomogeneities (TR = 400 ms, TE_1 = 5.19 ms, TE_2 = 7.65 ms, image matrix = 64 x 64, FOV = 192 mm, flip angle = 60°, slice thickness = 3 mm, voxel size = 3 x 3 x 3 mm, distance factor = 20%, 33 slices). Whole brain functional images were collected using a T2*-weighted EPI sequence (TR = 2000 ms, TE = 30 ms, image matrix = 64 × 64, FOV = 192 mm, flip angle = 78°, slice thickness = 3 mm, voxel size = 3 x 3 x 3 x mm, distance factor = 20%, 33 slices). Slices were orientated along the AC-PC line for each subject.

Statistical Analysis

Data Analysis: Behavior

All behavioral analyses were performed in R (RStudio version 1.1.383, RRID:SCR_000432, www.rstudio.com). We first characterized subjects' performance by computing error rates and reaction times (RT). We tested for potential effects of reward condition on error rates using a Bayesian two-sided paired t-tests (using *ttestBF* from the BayesFactor package in R). Error trials, and trials with RTs <300ms were removed from the data analysis. In order to identify potential effects of task and reward condition on RTs, we performed a Bayesian repeated measures ANOVA (using *anovaBF* from the BayesFactor package in R). This ANOVA included the factors task (X, Y) and reward (CR, NCR), and outputs Bayes Factors

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(BF) for all main effects and interaction terms. We did not expect tasks to strongly affect RTs, but did expect RTs to be lower in the CR condition, as compared to the NCR condition.

The Bayesian hypothesis testing employed here allows quantifying the evidence in favor of the alternative hypothesis (BF10) *and* the null hypothesis (BF01), allowing us to conclude whether we find evidence for or against a hypothesized effect, or whether the current evidence remains inconclusive (Rouder, Speckman, Sun, Morey, and Iverson, 2009). Unfortunately, in classical frequentist hypothesis testing we are unable to provide evidence for the null hypothesis in a similar way (Wagenmakers, 2007). In line with previous research (e.g. Andraszewicz et al., 2015; Mertens and De Houwer, 2016), we considered BFs between 1 and 0.3 as anecdotal evidence, BFs between 0.3 and 0.1 as moderate evidence, and BFs smaller than 0.1 as strong evidence against a hypothesis. BFs between 1 and 3 were considered as anecdotal evidence, BFs between 3 and 10 as moderate evidence, and BFs larger than 10 as strong evidence for a hypothesis. Although our conclusions are based solely on the BFs, we also provide frequentists statistical test outcomes for the interested reader.

Given that subjects were free to choose between the two tasks, some subjects might have shown biases to choosing one of the two tasks more often (although that would not have led to a higher overall reward, if anything biases should lower overall rewards). In order to quantify biases, we computed the proportion of trials in which subjects chose task X, separately for the CR and NCR conditions, and tested whether this value differed from 50% using a two-sided Bayesian t-test. The output BF was interpreted in the same way as in the previous analysis.

Choices in CR trials were assessed two-fold. First, we quantified how well subjects performed the probabilistic reversal learning task. If subjects were reliably able to determine which of the two tasks was currently the HR task, they should have chosen that task more often than expected by chance (50%). Thus the proportion of HR task choices in CR trials is our main measure of how successful subjects were in performing the task. This measure was compared to chance level using a one-sided Bayesian t-test.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Furthermore, we expected the proportion of HR choices to be higher in CR, than in NCR trials (where it should be 50%). This was tested using a paired one-sided Bayesian t-test.

Second, we assessed whether subjects were able to learn and update reward contingencies in the reversal learning task. Reinforcement learning (RL) theory suggest that such learning can take place by comparing received rewards with expected rewards, which are computed from the reward history (Sutton and Barto, 1990; Collins et al., 2017). Discrepancies between actual and expected rewards (reward prediction errors, RPE) are thought to signal surprise in the brain and to guide adjustment of behavior (Daw and Doya, 2006), a process which relies on dopaminergic signals in the midbrain (Pessiglione et al., 2006; Schultz, 2016). Here, we fitted a RL model to the choice data of each subject (separately for CR and NCR trials) in order to assess the learning process. Fitted RL models used simple delta-rule learning (as implemented in the *rlfit* package in Matlab, https://github.com/jmxpearson/rlfit). For each task choice *c* the expected reward *Q(c)* was learned from the reward history by comparing the expected and observed rewards at trial *t*:

$$Q_{t+1}(c) = Q_t(s) + \alpha \, X \, \delta_t$$

with $\delta_t = r_t - Q_t(c)$ being the RPE, and α being the learning rate. Choices were generated following a softmax choice function (as implemented in the rlfit package). The parameters were fitted over n = 10 iterations, with random starting values in each iteration. Learning rates were fitted with constraints [0, 1]. In order to assess the model fit, we also estimated a 'null' model for each subject. In this model, we again estimated expected outcomes and RPEs using the same algorithm described above, only fixing the learning rate to 0. The null model thus assumed that subjects do not learn changing reward contingencies, and we expected our RL model to outperform this null model. Model fit was assessed using the AIC and BIC (Burnham and Anderson, 2004). We also assessed an alternative 'hybrid' model, in which learning rates are allowed to vary on a trial-by-trial basis, instead of being fixed for each subject (Bai et al., 2014). It has been argued that such a model better captures behavior in probabilistic reversal learning tasks. In

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

our experiment the simple delta-rule learning model outperformed the more complex hybrid model (as assessed using AIC and BIC), and results from the hybrid model were not assessed further.

For each subject, the learning rate was extracted from the best-fitting model. We expected learning rates to be higher in CR than in NCR trials. In CR trials, the specific reward contingencies changed frequently, and thus subjects needed to update their contingency representations frequently as well. The learning rate in CR trials was also expected to correlate with successful task performance (% high reward choices), given that the reversal learning task can only be performed well if the represented reward contingencies change over time. In NCR trials, we expected learning rates to be low and uncorrelated with choice performance, because reward outcomes were randomly chosen and there were no contingencies to learn. Choices in NCR trials were assessed by testing whether subjects were able to choose tasks randomly in these trials. For this purpose, we computed the distribution of run lengths for each subject, i.e., the number of trials subjects chose to consecutively perform the same task. If subjects chose tasks randomly, this distribution can be expected to follow an exponential distribution (cf. Arrington and Logan, 2004; Soon et al., 2008). The average run length was computed for each subject, separately for CR and NCR trials, and compared to the expected run length under random choice behavior. We expected subjects to show longer runs in CR than in NCR trials, given that the probabilistic reward reversal learning task encourages subjects to perform the same task repeatedly. This was again tested using a one-sided Bayesian t-test.

Data Analysis: fMRI

fMRI data analysis was performed using Matlab (version R2014b 8.4.0, RRID:SCR_001622, The MathWorks) and SPM12 (RRID:SCR_007037, www.fil.ion.ucl.ac.uk/spm/software/spm12/). Raw data was imported according to BIDS standards (RRID:SCR_016124, http://bids.neuroimaging.io/). In order to assess which brain regions contained information about reward outcomes and task choices, raw data was unwarped, realigned and slice time corrected. It was then entered into a first level general linear model

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

318 analysis (GLM, Friston et al., 1994), and subsequently into a multivariate pattern analysis (MVPA, Cox and
319 Savoy, 2003; Kriegeskorte et al., 2006; Haxby, 2012; Haynes, 2015). In order to assess which brain regions
320 represented reward-learning signals, raw data was unwarped, realigned, slice time corrected, normalized,
321 and smoothed. It was then entered into a GLM, adding reward prediction errors as a regressor. Results
322 were analyzed using a mass-univariate approach. Full details of the analyses can be found below.

323 *Neural processing of reward*

324 Multivariate decoding of reward outcomes

325 In a first step, we assessed whether we can replicate previous findings demonstrating contingency effects
326 on reward processing (Tricomi et al., 2004). For this purpose, we estimated a GLM for each subject. For
327 each of the 5 runs we added regressors for each combination of reward value (high vs low) and
328 contingency (CR vs NCR). All regressors were locked to the feedback onset, the duration was set to 0.
329 Regressors were convolved with a canonical haemodynamic response function (as implemented in
330 SPM12). Estimated movement parameters were added as regressors of non-interest to this and all other
331 GLMs reported here.

332 *Baseline decoding*: In a next step, we performed a decoding analysis on the parameter estimates of the
333 GLM. A support-vector classifier (SVC, see Cox and Savoy, 2003; Mitchell et al., 2004; Kamitani and Tong,
334 2005), as implemented in *The Decoding Toolbox* (Hebart et al., 2014), was used using a fixed regularization
335 parameter (C = 1). We performed searchlight decoding (Kriegeskorte et al., 2006; Haynes et al., 2007),
336 which looks for information in local spatial patterns in the brain and makes no a prior assumptions about
337 informative brain regions. A sphere with a radius of 3 voxels was defined around each measured voxel,
338 and parameter estimates for high rewards (both in CR and NCR trials), and for low rewards (again, both in
339 CR and NCR trials) were extracted within that sphere, separately in each run. 4 out of 5 runs were used to
340 train the SVC to distinguish the neural patterns of high and low rewards. Classifier performance was then
341 tested on the remaining, independent run. This procedure was repeated until each run was left out once,

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

resulting in a 5-fold cross-validation and countering potential problems with overfitting. Mean prediction accuracy was calculated across all folds and written into the center voxel of the sphere. This was repeated for each measured voxel in the brain, resulting in a 3D accuracy map. These maps were computed for each subject, normalized to a standard space (Montreal Neurological Institute template as implemented in SPM12), and smoothed (Gaussian kernel, FWHM = 6mm) in order to account for potential differences in information localization across subjects. Group analyses were performed using a random effects model on the accuracy maps, using voxel-by-voxel t-tests against chance level (50%). The chance level was subtracted from all reported accuracy values. A statistical threshold of $p<0.0001$ (uncorrected) at the voxel level, and $p<0.05$ (family-wise error corrected) at the cluster level was applied to all analyses. This threshold is sufficient to rule out inflated false-positive rates in fMRI analyses (Eklund et al., 2016). Any regions surpassing this threshold were used as masks for the following decoding analyses (an approach previously used by Loose et al., 2017). One might argue that identifying strong, outcome-related signals using a method as sensitive as MVPA is trivial. But please note that we are not mainly interested in identifying reward-related signals per se, but rather focus on their modulation through outcome contingency, which is much more interesting. The baseline reward decoding is likely partly driven by underlying univariate signal differences, and we do not claim that results reflect differences in response patterns only. We chose to run this analysis as described to ensure we can compare results especially with the task-related analyses. We will base our conclusions mainly on comparing *differences* between the baseline and other analyses (see below), so this comparison does not constitute a case of double dipping. Lastly, this analysis is sensitive to differences in outcome value, but might possibly also identify brain regions related to unspecific preparatory (e.g., attentional) processes. Although preparatory processes should be identical in CR and NCR trials, due to the fact that the same high and low rewards were given in both conditions, we cannot fully exclude such effects either if subjects were generally more motivated to

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

perform CR than NCR trials. The underlying cause of any observed effects remain differences in reward outcomes however.

Differences in reward outcome coding: Although the baseline decoding analysis should have the maximum power to detect any outcome-related brain regions, results do not allow us to conclude whether outcome processing differed between CR and NCR trials. For this purpose, we repeated the decoding analysis, now only using CR trials, and only NCR trials, respectively. If contingent rewards indeed enhance encoding of reward outcomes in the brain, we should see higher accuracies in the CR than in the NCR decoding analysis. Please note, that we only used half the number of trials as before, thus considerably reducing the signal-to-noise ratio in these analyses. We thus expected lower statistical power and smaller effects.

Similarities in in reward outcome coding: Previous work demonstrated that not all brain regions show a contingency-related modulation of value signals (Elliott et al., 2004), and we thus tested whether some brain regions encoded reward outcomes invariantly across the contingency conditions. We trained a classifier to discriminate between high and low reward outcomes in the CR condition, and tested its performance in the NCR condition, and vice versa. This resulted in two accuracy maps per subject, which were averaged and then entered into a group analysis just like in the previous analyses. Importantly, only brain regions where patterns do not differ across both contingency conditions will show above-chance accuracies in this analysis. This so-called cross classification analysis can be used to identify brain regions in which outcome representations are invariant with respect to the contingency manipulation employed here (see also Kaplan et al., 2015), thus providing positive evidence for contingency-invariant coding of reward outcomes.

Neural correlates of reward-learning signals

While the previous analyses investigated the neural correlates of processing the hedonic value of reward outcomes, here, we directly assessed whether reward-learning signals are affected by reward contingency. Reward prediction errors (RPE) act as learning signals in our reversal learning task

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(Matsumoto et al., 2007; Daw et al., 2011). They indicate the need to update the internal model of the current task-reward associations (e.g. task X = high reward task). In order to identify brain regions encoding this important reward signal, we used a model-based fMRI approach (O'Doherty et al., 2007; Forstmann and Wagenmakers, 2015). In model-based fMRI, a computational model fitted to behavioral data is used to construct regressors, which are then used to estimate GLMs on fMRI data. This approach links brain and behavior in a mechanistic framework and has been used successfully in a number of different settings (for an overview see Forstmann and Wagenmakers, 2015). We used the reinforcement learning models fitted to the behavioral data, and computed trial-by-trial RPEs from the best fitting model of each subject. We then estimated two separate GLMs, one for CR trials and one for NCR trials, on normalized and smoothed raw data. For each of the 5 runs, we added one regressor (duration = 0) locked to the onset of the feedback screen of each trial. Prediction errors should be strongest at this point in time. We added the trial-by-trial RPEs as a parametric modulator, allowing us to identify brain regions correlating with RPE signals. As before, regressors were convolved with a canonical haemodynamic response function. For each subject, a t-contrast map was computed to identify regions reflecting RPEs. These maps were then entered into a group level random effects analysis (within-subjects ANOVA with the factor contingency (CR, NCR)) in order to identify brain regions where prediction errors were modulated by reward contingency. Results were thresholded at $p < 0.001$ (uncorrected) at the voxel level, $p < 0.05$ (FWE corrected) at the cluster level.

Multivariate decoding of tasks

All analyses described above aimed at assessing effects of reward contingency on reward processing. Now, we turn to also test whether any such potential effects could be demonstrated on the implementation of chosen behavior in the brain. For this purpose, we assessed which brain regions encoded the chosen tasks. Two GLMs were estimated for each subject, one modelling task-related brain activity at the time of decision-making, and one modelling activity during a subsequent maintenance phase. It has been shown

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that formation and maintenance of intentions rely on partly dissociable brain networks (Bunge et al., 2003; Gilbert, 2011), and our design allowed us to estimate independent signals related to both epochs as they were separated by a variable inter-trial-interval.

In the first GLM ($GLM_{maintenance}$), for each of the 5 runs we added regressors for each combination of chosen task (task X, task Y) and reward contingency (CR, NCR). All 4 regressors were locked to the cue onset, the duration was set to cover the whole delay period. Please note that due to the jittered delay period duration, the modelled signals were dissociated from the task execution and feedback presentation. These boxcar regressors were then convolved with a canonical haemodynamic response function. This model is highly similar to the model used in (Wisniewski et al., 2016), where subjects were also free to choose one of two different tasks in each trial, making current results highly comparable to this previous study. In sum, this model estimated task-specific brain activity during intention maintenance, i.e. while subjects had to represent their intention to perform a specific chosen task, without yet being able to prepare a specific motor response. A second GLM was estimated ($GLM_{decisiontime}$), in order to extract task-specific brain activity at the time subjects made their choice which of the two tasks to perform. Note that although the cue suggested that subjects should make a task choice at that point in time, there is no strong way of controlling the exact point in time at which choices were made. In fact, choices could have been made earlier than the presentation of the choice cue. It has been shown before that under free choice conditions, subjects choose a task as soon as all necessary information to make a choice is available (Hampton and O'Doherty, 2007; Wisniewski et al., 2015b). In this experiment, this time point is the feedback presentation of the previous trial. At this point, subjects can judge whether they e.g. chose the HR or LR task and determine which of the two tasks to perform in the next trial. We used this approach successfully in a previous experiment (Wisniewski et al., 2015b), again making current results highly comparable with these previous findings. All further task decoding analyses were performed on both GLMs.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

437 *Baseline decoding*: The task decoding analyses followed the same logic as the reward outcome analyses
438 described above. We first performed a searchlight decoding analysis (radius = 3 voxels, C = 1), contrasting
439 parameter estimates for tasks X and Y in all trials (CR and NCR combined). This analysis has the maximum
440 power to detect any brain regions containing task information, which can be notoriously difficult
441 (Bhandari et al., 2018). Resulting accuracy maps were normalized, smoothed (6mm FWHM), and entered
442 into a random effects group analysis (t-test vs chance level, 50%). Results were thresholded at $p<0.001$
443 (uncorrected) at the voxel level, and $p<0.05$ (family-wise error corrected) at the cluster level. Again,
444 regions surpassing this threshold were used to define functional regions-of-interest for the following
445 decoding analyses (see Loose et al., 2017).
446 *Differences in task coding*: In order to assess whether task coding is modulated by reward contingency,
447 we repeated the decoding analysis separately for CR and NCR trials. If contingent rewards indeed increase
448 task shielding in the brain, we should see higher accuracies in the CR than in the NCR decoding analysis.
449 This effect should be especially pronounced if both tasks are similar and easily confused, which is the case
450 in our experiment. Please note, that we again only used half the number of trials as before, reducing the
451 signal-to-noise ratio in these analyses. We thus expected lower statistical power and smaller effects.
452 *Similarities in task coding*: Some previous work suggests that tasks are encoded in a context-invariant
453 format in the brain (Zhang et al., 2013; Wisniewski et al., 2016), and we directly tested whether this was
454 also true in this experiment. Using a cross-classification (xclass) approach, we trained a classifier on CR
455 trials and then tested it on NCR trials (and vice versa). And brain regions showing above chance decoding
456 accuracies in this analysis provides positive evidence of task coding that is invariant with respect to
457 contingent vs non-contingent reward outcomes. Please note that this analysis also ensures that task-
458 related signals are not confounded by potential differences in e.g. cognitive load or expected reward
459 across the CR and NCR conditions, as classifiers are trained and tested only *within* one contingency
460 condition.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Region of interest analyses: We also assessed task information in a number of a priori defined regions of interest (ROI). First, we attempted to replicate results from one of our previous experiments (Wisniewski et al. 2015). There, the dmPFC has been found to encode task choices at the time of decision-making. We extracted this functional ROI, and tested whether we could replicate the finding in this independent and larger sample. Although the overall design differed considerably (e.g. 3 vs 2 tasks, changing reward outcomes vs changing task difficulty), both studies used the same object-categorization task. Second, two previous experiments found task information to be maintained in the fronto-parietal cortex in a context invariant fashion (Loose et al. 2017; Wisniewski et al. 2016). In one paper, task coding was invariant with respect to freely chosen vs. externally cued tasks (Wisniewski et al. 2016), while in the other paper, task coding was invariant with respect to high vs. low control demands (Loose et al. 2017). If we were to show that the regions identified in these two experiments also encode tasks invariantly across reward contingency conditions, that would provide additional evidence for general, context invariant task coding in the fronto-parietal cortex. We thus extracted functional ROIs from both papers (Wisniewski et al. 2016: left parietal cortex, left PFC, Brodman area 8; Loose et al. 2017: left parietal cortex, left PFC), and tested this hypothesis in this independent data-set. For all ROIs defined, we extracted accuracy values for all voxels within the ROI, which were then averaged. One-sided Bayesian t-tests across subjects were performed to assess whether they were above chance.

Control analyses: In order to further corroborate the reliability of our results, we performed a number of control analyses. It has been pointed out before, that RT effects might partly explain task decoding results (Todd et al., 2013), although others were unable to show any such effects (Woolgar et al., 2014; Wisniewski et al., 2015b). Given that we expected RTs to differ across reward conditions, we decided to conservatively control for RTs effects. First, we repeated the GLM estimation, only adding reaction times as an additional regressor of non-interest. We then repeated the main decoding analyses, and tested

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

484 whether accuracy values differed significantly. If RTs indeed explain our task decoding results, we should
485 see a reduction in decoding accuracies when RT effects were regressed out of the data.
486 Furthermore, it is possible that some subjects exhibit excessive error rates or have a strong bias to choose
487 one task more often than the other. High error rates might decrease the signal-to-noise ratio and thus
488 affect observed results. Very strong choice biases might have a similar effect, in extreme cases subjects
489 might have performed only one of the two tasks in a given run (although this was unlikely). In order to
490 ensure that we had enough trials to estimate each regressor, we first excluded subjects with excessively
491 high error-rates (more than 1.5*IQR above average), and then excluded subjects with strong choice biases
492 (more than 1.5*IQR above average). We then tested whether each regressor in all remaining subjects
493 could be estimated from at least 6 trials. If a regressor could only be estimated from fewer trials, that run
494 was excluded from the analysis. Subjects in which more than 1 run was thusly excluded were altogether
495 excluded from the analysis. These criteria were highly similar to the criterion used in (Wisniewski et al.,
496 2015b), which proved an effective control. After excluding these subjects, we repeated the main analyses
497 on the remaining subjects and tested whether they differed from the analysis including all subjects.
498 Two further control analyses were performed to confirm the validity of the decoding procedure used.
499 First, we performed a ROI decoding analysis on a brain region that is not related to task-performance in
500 any way, expecting accuracies to be at chance level. We chose the primary auditory cortex for this
501 purpose, defined using the WFU_pickatlas tool (https://www.nitrc.org/frs/?group_id=46, RRID:
502 SCR_007378). Second, we tested whether our chance level was indeed 50%, or whether it was biased. For
503 this purpose, we performed a permutation analysis (as implemented in the Decoding Toolbox). We
504 repeated the baseline decoding analysis 1000 times for each subject, only randomly assigning the test
505 labels in each of the 1000 permutations. A null distribution was calculated from these permutations
506 separately for each subject, and the mean accuracy value of the null distribution served as an empirical
507 estimate of the chance level. In order to test whether the estimated chance level deviated from 50%, we

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

508 performed a two-sided Bayesian t-test. Additional exploratory analyses were performed to assess possible
509 correlations between behavioral measures, questionnaires, and fMRI results (Figure 2–1).

510 Results

511 Behavioral results

512 We first assessed the effects of tasks (X, Y) and reward condition (CR, NCR) on error rates and reaction
513 times (RT). The average error rate across all subjects was 5.89% (SEM = 0.74%). Thus, subjects were able
514 to perform the task accurately. There was no evidence for an effect of reward condition on error rates
515 (Bayes Factor (BF10) = 0.88, t(34) = 1.96, p = 0.06). Error trials were removed from all further analyses. A
516 repeated-measures ANOVA on the reaction times (RT) including the factors task and reward condition
517 revealed no main effect of reward (BF01 = 31.95, F(1,34) = 0.38, p = 0.53, Figure 2 A). This is likely due to
518 the fact that subjects had a long time to prepare the execution of the task, which minimized potential
519 contingency-related differences in RTs. There was a strong main effect of task however (BF10 > 150,
520 F(1,34) = 3.78, p = 0.05), with task X (RT_X = 1415ms, SEM = 29ms) being faster than task Y (RT_Y = 1467ms,
521 SEM = 35ms). Please note, that this cannot be simply due to a difficulty difference between the two S-R-
522 mappings called task X and task Y, as the specific S-R-mappings were counter-balanced across subjects.
523 Given the long delay phase, subjects should have had enough time to prepare both tasks well, and we
524 were somewhat surprised to see this RT difference. This results might reflect the encoding sequence in
525 the learning phase. Subjects might have learned the S-R-mapping labelled X first, and then learned the S-
526 R-mapping labelled Y second. If the second task is mainly encoded by how it differs from the first, this
527 might lead to a RT difference (see also Lien et al., 2005). There was no evidence for an interaction between
528 task and reward (BF10 = 0.26, F(1,34) = 6.63, p = 0.01).
529 We then assessed whether subjects showed choice biases towards one of the two tasks, which might
530 indicate stable preferences for specific tasks and might in turn affect fMRI analyses (see below). In order

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

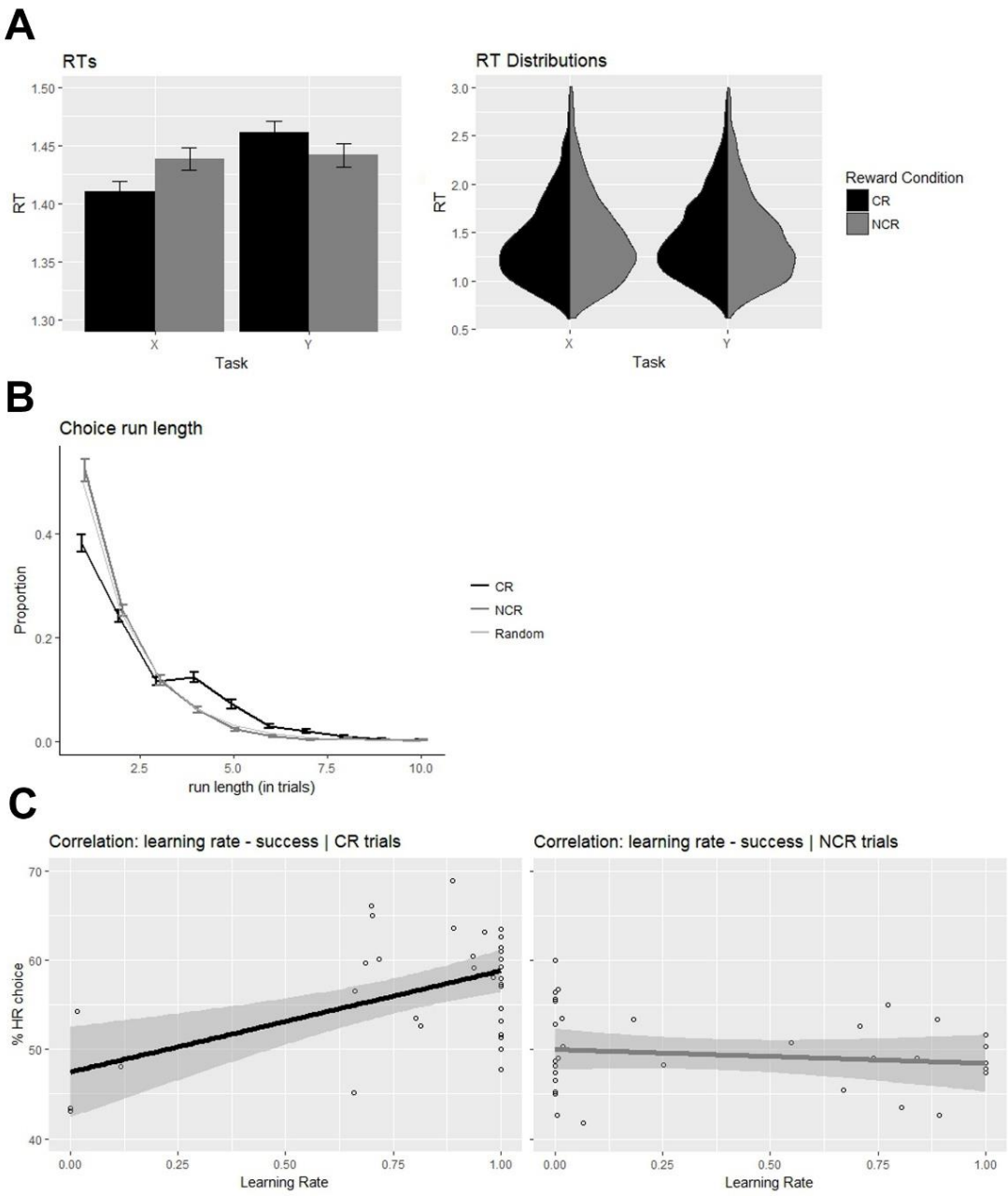


Figure 2. **Behavioral Results. A.** Reaction Times (RT). The bar graph depicts the average reaction times for each combination of task and reward condition. Contingent (CR) trials are shown in black, non-contingent (NCR) trials are shown in grey. The violin plot depicts the RT distributions of the same data. **B.** Choice run length. This plot depicts the distribution of run lengths (the number of consecutive trials in the same task). Data from CR trials is shown in black, data from NCR trials is shown in grey. The expected distribution if choices were completely random is depicted in light grey. All error bars depict the SEM. **C.** Correlation of learning rate and success. Learning rates were extracted form a fitted RL model. Success was measured as % HR task choices. In CR trial, subjects who learned the changing reward contingencies quickly, were more successful. In NCR, no such correlation was observed. Each dot represents one subject, and linear functions were fitted to the data (lines). Further information on correlations between performance and additional questionnaire measures can be found in Supplementary Figure 1.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to quantify any potential choice biases, we computed the percentage of task X choices for both reward conditions separately. Subjects chose task X in 52.14% (SEM = 1.44%) of the CR trials, and 52.29% (1.72%) of the NCR trials. These values did not differ from 50% in the CR condition (BF10 = 0.48, $t(34) = 1.47$, $p = 0.14$), and NCR condition (BF10 = 0.40, $t(34) = 1.32$, $p = 0.19$). There was also no difference between the two reward conditions (BF01 = 5.45, $t(34) = 0.14$, $p = 0.88$), indicating that subjects did not exhibit strong choice biases in this experiment.

Next, we measured subjects' success in solving the reversal learning task presented in CR trials, by computing the percentage of high-reward (HR) task choices for each subject. If they were unable to learn which of the two tasks was the HR task, this value should be 50%. Higher values indicate increasing success in performing the reversal learning task. We hypothesized that subjects chose HR tasks more often in CR, as compared to NCR trials. Subjects chose the HR task in 56.40% (SEM = 1.15%) of the CR trials, which was above chance level (BF10 >150, $t(34) = 5.56$, $p < 0.001$). They chose the HR task in 49.47% (SEM = 0.84%) of the NCR trials, which did not differ from the chance level (BF01 = 4.59, $t(34) = 0.62$, $p = 0.53$). Importantly, we found strong evidence for our hypothesis that subjects chose HR tasks more often in the CR, than in the NCR condition (BF10 > 150, $t(34) = 5.44$, $p < 0.001$). These findings demonstrate that subjects indeed chose tasks strategically in the CR condition, in order to maximize their reward outcome.

We then described the learning process in the CR trials in more details by fitting a reinforcement learning (RL) model (Sutton and Barto, 1990, see Materials and Methods for more details) to the choice data of each subject, and extracting the estimated learning rate (α). We expected subjects to show high learning rates in CR trials, reflecting the fact that subjects frequently needed to update which of the two tasks yielded higher reward outcomes. We compared fitted models in both CR and NCR trials to a null model, in which the learning rate was fixed to 0, assuming that subjects never learned about the reward contingencies in this experiment. Model fit was assessed using the AIC and BIC (Burnham and Anderson, 2004). As expected, the RL model provided a better fit to the data than the null model in both CR trials

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(AIC_{RL_CR}=129.97, AIC_{NULL_CR}=159.54, BIC_{RL_CR}=132.71, BIC_{NULL_CR}=159.54), as well as NCR trials (AIC_{RL_NCR}=158.70, AIC_{NULL_NCR}=158.90, BIC_{RL_NCR}=132.71, BIC_{NULL_NCR}=158.90). Given that reward contingencies changed frequently in the CR trials, we expected learning rates to be higher in CR than in NCR trials. We found strong evidence in favor of this hypothesis (α_{CR}: mean = .78, median = .96, sd = .33, min/max = <.001/1; α_{NCR}: mean = .36, median = .06, sd = .41, min/max = <.001/1; BF10 > 150, t(34) = 4.63, p < 0.001). We then correlated estimated learning rates with successful task performance (% HR task choices), again using a Bayesian framework for correlation estimation (using *bayes.cor.test* form the BayesianFirstAid package in R). Specifically, we estimated the probability of the correlation being above 0 (p(r>0)), and also estimated 95% credible intervals (95% CI), which indicates the range of values within which the correlation falls with a 95% probability. If this interval did not include 0, we interpreted the correlation as either positive or negative. The estimated learning rate in CR trials was indeed correlated with successful task performance (% HR task choices), r = .44 (95% CI = [.026, .74], p(r>0) = .97, Figure 2 C), linking our computational modelling more closely to behavior. As a control analysis, we also correlated learning rate in NCR with proportion of HR task choices in NCR trials. As expected, we found no correlation, r=-.12 (95% CI = [-.46, .21], p(r>0) = .21). Classically estimated correlations confirmed these results, r = .56, p < 0.001, and r = -.12, p = 0.46, respectively. These results indicate that successful subjects were able to learn about changing reward contingencies more quickly, and also demonstrate that subjects treated both reward conditions differently.

Lastly, in NCR trials we expected subjects to choose tasks randomly, as their choices had no effect on reward outcomes (see Materials and Methods for more details). In order to test this, we computed the run length for each subject, i.e. the average number of consecutive trials in the same task (Arrington and Logan, 2004). The average run length was then compared to the expected theoretical distribution if choices were fully random (Figure 2 B). The average run length in NCR trials was 1.95 trials (SEM = 0.07 trials), which did not differ from the expected 'random-choice' run length (BF01 = 4.85, t(34) = 0.52, p =

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

591 0.60). Subjects in this experiment thus did not exhibit repetition bias, which has been reported previously
592 for free-choice tasks (Arrington and Logan, 2004). The average run length in CR trials was 2.54 trials (SEM
593 = 0.08 trials), which was longer than in NCR trials (BF10 >150, $t(34) = 5.91$, $p < 0.001$), demonstrating that
594 subjects stayed longer in the same task. This is a viable strategy in the reversal-learning task they
595 performed. Once they identified which was the HR task, repeatedly performing that task maximized
596 reward outcomes.

597 Reward-related brain activity

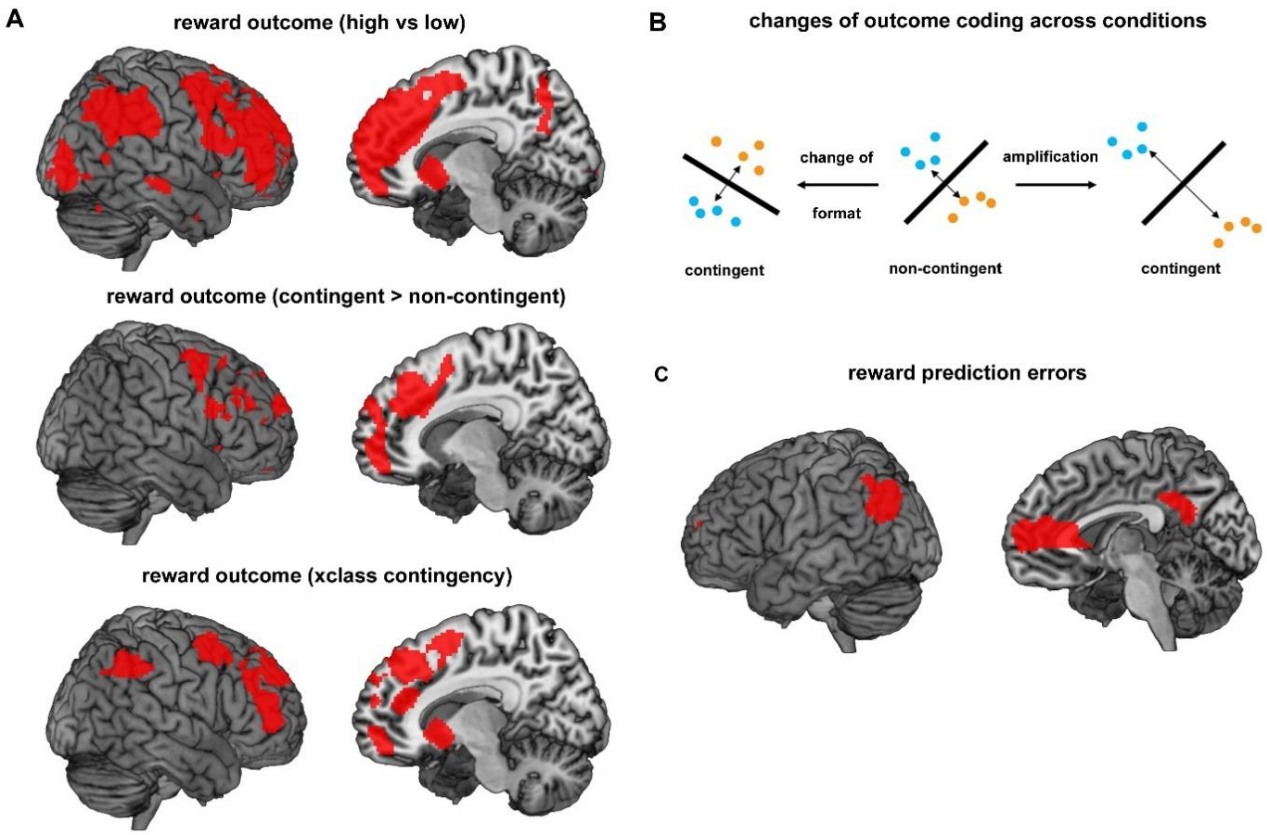
598 Multivariate decoding of reward outcome values

599 One of our main goals was to assess whether reward contingency affects valuation processes in the brain.
600 In a first analysis, we aimed to extend previous findings demonstrating an effect of reward contingency
601 on the processing of its hedonic value (Elliott et al., 2004). For this purpose, identified brain regions
602 encoding outcome values (high vs low) at the time of feedback presentation. We found an extensive
603 network to encode outcome values including subcortical brain regions, as well as large parts of the
604 prefrontal and parietal cortex (Figure 3 A). Please note that this contrast might not *only* capture specific
605 reward value signals, it might also reflect effects caused by differences in reward outcomes, like attention
606 or motor preparation. We explicitly assessed whether reaction times affected outcome coding (see Todd
607 et al., 2013), and found no effect (Figure 3-1). Subsequently, we assessed whether these outcome signals
608 were modulated by reward contingency, hypothesizing that contingent rewards showed stronger
609 decoding results than non-contingent rewards. For this purpose, we repeated the decoding analysis
610 described above, now separately for CR and NCR trials, respectively. The two resulting accuracy maps
611 were entered into a within-subjects ANOVA, and a contrast was computed identifying brain regions with
612 higher accuracies in CR than in NCR trials. Using small-volume correction ($p < 0.001$ uncorrected, $p < 0.05$
613 FWE corrected), we assessed which of the brain regions identified in the *baseline* analysis also showed
614 stronger value coding for contingent rewards. We found the striatum, bilateral lateral PFC, dACC, anterior

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

medial PFC, and IPS to show stronger reward value coding for contingent rewards, as compared to non-contingent rewards. In a last step, we directly assessed whether there were brain regions that encoded reward values in a contingency-invariant fashion, using a cross-classification approach. Here, we trained a classifier to distinguish high from low rewards only on CR trials, and then tested its performance on NCR trials, and vice versa. This allowed us to identify brain regions in which outcome values are encoded invariantly across the two contingency conditions, i.e. where neural patterns do not differ across contingency conditions (Kaplan et al., 2015). We found the striatum, lateral and medial PFC, dACC, and IPS to encode rewards in a contingency invariant form. This pattern of results suggests that the neural code for different reward values did not change across contingency conditions, yet value signals were still stronger in CR than in NCR trials. This is compatible with an increased gain or amplification of value representations through contingency (Figure 3 B), where representations do not change but become more separable in neural state space (see Waskom et al., 2014 for a similar argument).

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



627

628 **Figure 3: Reward-related brain activity. A.** Multivariate decoding of reward outcome value. Above:
629 baseline decoding. Depicted are regions that encoded the value of reward outcomes (high vs. low). The
630 regions identified were used as masks for the following analyses. Results are displayed at $p < 0.05$ (FWE
631 corrected). Middle: regions with a stronger coding of reward value in contingent (CR) than in non-
632 contingent (NCR) trials. Below: regions encoding reward values in similar formats in both contingency
633 conditions, as tested using a cross-classification (xclass) analysis. We also repeated this analysis, explicitly
634 controlling for the effect of reaction times, and results can be found in Supplementary Figure 2. **B.**
635 Amplification vs change of format of neural coding. Most regions identified in A showed both stronger
636 decoding in CR trials, and similar formats across both contingency conditions. This is compatible with an
637 amplification or gain increase of neural codes. In the middle, a hypothetical example of a pattern decoding
638 is depicted. High reward trials are depicted as blue, low reward trials as orange dots. The classifier fits a
639 decision boundary to separate the two distributions. If this code changes between the two contingency
640 conditions (left), decoding is still possible at similar accuracy levels as before, but a classifier trained on
641 NCR trials will be unsuccessful in classifying CR trials. If this code is amplified in the CR condition however
642 (right), the same classifier can will be successful in both conditions. Accuracies increase, as the two
643 distributions become more separable. **C.** Brain regions correlating with reward prediction error signals (in
644 both CR and NCR trials).

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Learning signals: Reward prediction errors

In the previous analysis, we assessed which brain regions directly encoded different reward outcomes in individual trials. We now turn to identifying brain regions supporting reward-based learning processes across multiple trials. We used the fitted RL models (see above) to extract trial-by-trials reward prediction errors (RPEs), which signal the need to adapt one's behavior (O'Reilly et al., 2013). Following a model-based neuroscience approach (Forstmann and Wagenmakers, 2015), we identified brain regions in which activity correlated with RPEs. These learning signals should be strongest at the time of decision-making (in our case the reward feedback presentation, see Materials and Methods for more details), and we found the left parietal cortex and anterior medial PFC to correlate with RPEs in CR trials (Figure 3 C). In NCR trials, we found anterior cingulate and anterior medial prefrontal cortex to encode RPEs. We statistically assessed the difference between these two results, using a within-subjects ANOVA with the factor 'model' (2 levels). We found no significant differences ($p < 0.001$ (uncorrected) at the voxel level, $p < 0.05$ (FWE corrected) at the cluster level), and thus decided to combine both conditions to increase statistical power. Running the same analysis over all trials (CR and NCR) again revealed the left parietal cortex (overlapping with the region identified in Analysis 1), ACC and anterior medial PFC, but also the precuneus. These regions thus signal discrepancies between expected and received rewards during feedback presentation, indicating the need to adapt behavior in the subsequent trial.

These brain regions could either signal general surprise, as RPEs are the difference between expected and received rewards (O'Reilly et al., 2013). They could also signal the need to update an internal model of our environment. Our findings are more in line with the former option. Any region signaling the need to update the internal model of the environment should be specifically involved only in CR trials (where updating is required), and not in NCR trials (where updating is not needed). In order to test this, we identified subjects that only showed high learning rates in CR and low learning rates in NCR trials (n=19). For these subjects, prediction errors only signaled the need to update their internal model. Results

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

669 showed that for this subset of subjects, only the anterior medial PFC correlated with RPEs ($p < 0.001$
670 uncorrected at the voxel, and $p < 0.05$ FWE corrected at the cluster level). This seems to suggest that the
671 anterior medial PFC was involved in model updating, while the left parietal cortex and precuneus signaled
672 general surprise. Given that the sample size was considerably smaller in this analysis, results should be
673 interpreted with caution however.

674 ## Multivariate decoding of tasks

675 *Baseline decoding analysis*: The previous analysis demonstrated that reward contingency indeed affected
676 the neural processing of the hedonic value of reward outcomes, and possibly also related learning signals.
677 In the following analysis we assessed whether these effects propagated to the implementation of chosen
678 behavior, i.e. the coding of chosen tasks as well. For this purpose, we first estimated a GLM modelling
679 task-related neural activity during the maintenance of chosen tasks, from the onset of the 'choose' cue to
680 the onset of the task execution screen. (see Materials and Methods for more details, and Haynes et al.
681 (2007) for a similar approach). During this time, subjects needed to maintain their intention to perform
682 one of the two tasks. We performed a searchlight decoding analysis contrasting task X and task Y,
683 combining both CR and NCR trials in order to maximize the power to detect any brain regions containing
684 task information (see Loose et al., 2017 for a similar approach). Please note that during this time subjects
685 cannot prepare specific motor responses yet, but they can use this time to retrieve the current S-R-
686 mapping. We found two brain regions to contain task information, the left posterior parietal cortex (mean
687 accuracy = 4.61%, SEM = 0.65%), spanning over the midline into the right parietal cortex, and the right
688 anterior middle frontal gyrus (aMFG, mean accuracy = 4.66%, SEM = 0.89%, see Figure 4 A, Table 1).
689 Interestingly, the parietal cluster identified in this analysis partly overlapped with the parietal cluster
690 found to encode reward prediction errors in the previous analysis, suggesting that the left parietal cortex
691 is involved in both reward-learning and task processing.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

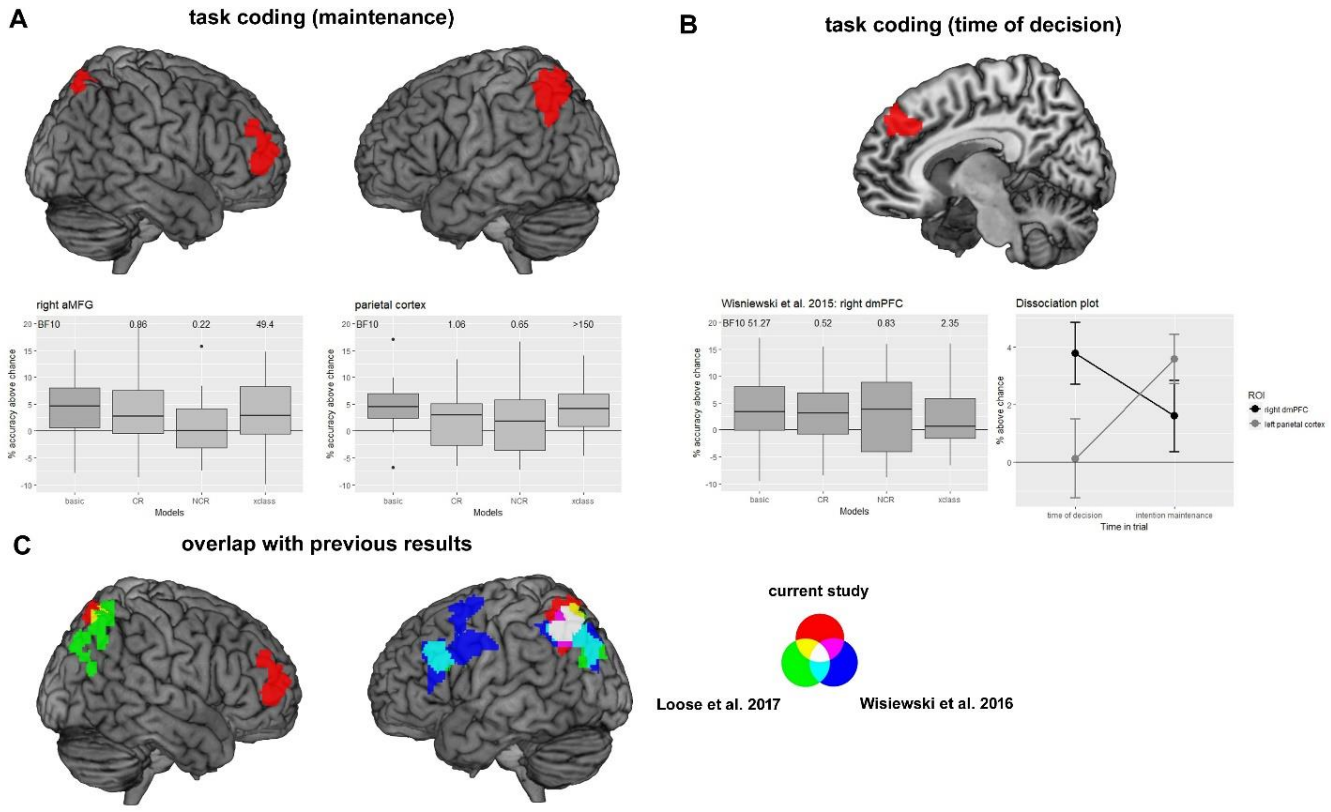


Figure 4: **Task coding**. **A.** Task coding during maintenance. Results from the baseline decoding analysis are depicted above. Two clusters passed the significance threshold, one in the parietal cortex, and one in the right anterior MFG. These two clusters were then used as ROIs, and accuracies were extracted for the contingent (CR), non-contingent (NCR), and cross-classification (xclass) task decoding analyses. Results can be seen below. Above the boxplots, Bayes factors (BF10) of a t-test vs. chance level are shown. Please note, that we do not report BF10 for the baseline analysis, as this analysis was used to define the ROIs, and running additional statistical tests on this data would constitute double dipping. **B.** Task coding at the time of decision-making. Above the ROI in the right dmPFC used in this analysis from Wisniewski et al. (2015) is depicted. This study demonstrated that the right dmPFC encodes tasks at the time of decision-making. The box plot depicts results from our data in this ROI, for all four analyses performed (baseline, CR, NCR, xclass). We largely replicate these previous findings. The dissociation plot depicts a double dissociation between two ROIs (right dmPFC, as defined using data from Wisniewski et al., 2015, and the left parietal cortex, as defined using data from Wisniewski et al., 2016), and two time points in the trial (time of decision-making, maintenance). It can be seen that the dmPFC only encodes tasks at the time of decision-making, while the left parietal cortex only encodes tasks during the maintenance phase. All error bars represent SEM. **C.** Overlap with previous results. Results from the current study (red) are overlain on previous findings from Wisniewski et al. 2016 (blue), and Loose et al. 2017 (green). All results are based on task decoding analyses (searchlight decoding, radius = 3 voxels, C = 1, chance level = 50%), albeit with different specific tasks being contrasted in each study. Despite this fact, all three studies find task information around the intraparietal sulcus. Findings in the PFC are less consistent. We further assessed task information encoded throughout the multiple-demand network, results can be found in Supplementary Figure 3.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1: Baseline task decoding

				MNI coordinates (peak)		
Brain region	**Side**	**Cluster size**	**Mean accuracy (SEM)**	**X**	**Y**	**Z**
parietal cortex	Bilateral	2427	4.61% (0.65%)	-10	-60	60
anterior MFG	Right	955	4.66% (0.89%)	32	58	18

Results are shown for a statistical threshold of p<0.001 (uncorrected) at the voxel level and p<0.05 (FWE corrected) at the cluster level.

Differences in task coding: In a next step, we assessed whether tasks were encoded with a higher accuracy in CR, than in NCR trials, similar to what we found for reward outcomes. Previous research demonstrated higher decoding accuracies in rewarded, as compared to non-rewarded tasks (Etzel et al., 2016). We built functional ROIs from the two regions identified in the baseline analysis, and extracted the average accuracy values for the task decoding analyses performed on CR trials only, and NCR trials only. Please note that these two analyses use only half as many trials as the baseline analysis, and the signal-to-noise-ratio can be expected to be lower. We found no task information in the parietal cortex in these two analyses (CR: 1.29%, SEM = 0.91%, BF10 = 1.06, t(34) = 1.59, p = 0.06; NCR: 1.73%, SEM = 1.44%, BF10 = 0.64, t(34) = 1.23, p = 0.11), and found no evidence for stronger task coding in CR than in NCR trials (BF10 = 0.16, t(34) = 0.09, p = 0.53). A similar pattern of results was found in the right aMFG (CR: 1.79%, SEM = 1.37%, BF10 = 0.85, t(34) = 1.44, p = 0.07; NCR: 0.48%, SEM = 1.35%, BF10 = 0.22, t(34) = 0.25, p = 0.40; CR > NCR: BF10 = 0.40, t(34) = 0.84, p = 0.20). Thus, we find no evidence for an effect of reward contingency on task representations, despite the fact that behavior clearly differed between the two reward conditions, and that contingency has been found to modulate the coding of reward outcomes. In order to assess whether the lack of evidence for differences in task coding might stem from a lack in

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

734 statistical power, we performed an additional control analysis. We again performed two separate task
735 decoding analysis, only using high reward and low reward trials (instead of CR and NCR trials), respectively.
736 We then tested whether decoding accuracies differed between these two conditions. Importantly, this
737 analysis has a similar statistical power, as the same number of trials is used. And indeed, we found task
738 coding to differ between these two conditions even at the whole brain level ($p < 0.001$ uncorrected at the
739 voxel, and $p < 0.05$ FWE corrected at the cluster level). Please note that this comparison might confound
740 effects of reward value with attentional processes. Nevertheless, this shows that our analysis approach is
741 able to identify differences in task coding in this dataset, although it fails to do so for our reward
742 contingency manipulation.
743 *Similarities in task coding*: We also directly tested whether task representations were invariant across the
744 two reward conditions, using a cross-classification approach. We trained a classifier to distinguish tasks in
745 CR trials, and tested its performance in NCR trials, and vice versa. In this analysis, accuracies can only be
746 above chance if task coding is invariant across both conditions. Results indicate than both the parietal
747 cortex (4.03%, SEM = 0.76%, BF10 > 150), as well as the right aMFG (3.71%, SEM = 1.16%, BF10 = 49.39)
748 show this type of contingency-invariant task coding. We further tested whether accuracies in the cross-
749 classification differed from the baseline accuracies, finding moderate evidence for an absence of any
750 differences (parietal cortex BF01 = 4.34, $t(34) = 0.71$, $p = 0.47$, aMFG BF01 = 3.94, $t(34) = 0.84$, $p = 0.40$).
751 These results thus show that the parietal cortex and aMFG encode tasks using a general, reward-
752 contingency-invariant format.
753 *ROI analyses and replications*: We also tested for task information in several a-priori ROIs, taken from two
754 previous experiments (Loose et al. 2017, Wisniewski et al. 2016), which tested for effects of cognitive
755 control, and free choice on task coding, respectively. Both previous studies found the left parietal cortex
756 to be involved in context-invariant task coding, and we thus set out to replicate these previous results
757 here. We extracted the ROIs reported in these two studies, and extracted decoding accuracies in each of

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

these ROIs, for all 4 analyses performed here (baseline, CR, NCR, xclass). We were able to replicate Loose and colleagues' left parietal results (baseline BF10 = 133.69, $t(34) = 3.89$, $p < 0.001$; CR BF10 = 0.68, $t(34) = 1.23$, $p = 0.10$; NCR BF10 = 0.54, $t(34) = 1.11$, $p = 0.13$; xclass BF10 = 33.17, $t(34) = 3.33$, $p = 0.001$). Although somewhat weaker, we also replicated their right parietal results (baseline BF10 = 8.49, $t(34) = 2.72$, $p = 0.004$; CR BF10 = 0.77, $t(34) = 1.37$, $p = 0.08$; NCR BF10 = 0.14, $t(34) = 0.28$, $p = 0.61$; xclass BF10 = 8.10, $t(34) = 2.70$, $p = 0.005$). However, we were unable to detect task information in left PFC (baseline BF10 = 0.49, $t(34) = 1.03$, $p = 0.15$; CR BF10 = 0.21, $t(34) = 0.23$, $p = 0.40$; NCR BF10 = 0.44, $t(34) = 0.93$, $p = 0.17$; xclass BF10 = 0.29, $t(34) = 0.54$, $p = 0.29$), which is in line with the original paper, where PFC findings were also somewhat less robust. Additionally, we were able to replicate Wisniewski and colleagues' left parietal finding (baseline BF10 = >150, $t(34) = 4.20$, $p < 0.001$; CR BF10 = 0.80, $t(34) = 1.40$, $p = 0.08$; NCR BF10 = 0.47, $t(34) = 1.00$, $p = 0.16$; xclass BF10 = 87.28, $t(34) = 3.72$, $p < 0.001$), as well as left BA8 (baseline BF10 = 9.3, $t(34) = 2.77$, $p = 0.004$; CR BF10 = 0.39, $t(34) = 0.83$, $p = 0.20$; NCR BF10 = 0.36, $t(34) = 0.76$, $p = 0.22$; xclass BF10 = 3.09, $t(34) = 2.22$, $p = 0.16$), but not the left PFC (baseline BF10 = 0.59, $t(34) = 1.17$, $p = 0.12$; CR BF10 = 0.37, $t(34) = 0.78$, $p = 0.21$; NCR BF10 = 0.16, $t(34) = 0.15$, $p = 0.56$; xclass BF10 = 0.38, $t(34) = 0.81$, $p = 0.21$). Thus, three studies with similar overall designs but considerable differences in the specific tasks used consistently find invariant task coding in the parietal, but not in the prefrontal cortex.

Furthermore, Wisniewski et al. 2015 found task information at the time of decision-making in the right dorso-medial PFC (Figure 4 B). In order to replicate this finding, we repeated all 4 task decoding analysis, only looking at the time of decision-making instead of intention maintenance (which was the reward feedback presentation in this experiment, see Materials and Methods for more details). The right dmPFC, as identified by Wisniewski and colleagues, was found to encode tasks also in the current study (baseline 3.76%, SEM = 1.07%, BF10 = 51.27, $t(34) = 3.51$, $p < 0.001$, Figure 4 B). This was despite the fact that there were considerable differences in the overall experimental design of these two studies (e.g. 2 class vs. 3 class decoding, changing reward outcomes vs. changing task difficulty). We found anecdotal evidence for

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

782 contingency-invariant task coding in this region (xclass 2.03%, SEM = 0.98%, BF10 = 2.35, t(34) = 2.07, p =
783 0.02), although the baseline and xclass analyses did not differ (BF10 = 1.64, t(34) = 1.63, p = 0.11).
784 Interestingly, the dmPFC was also found to encode reward outcome values, with its outcome signal being
785 amplified by our contingency manipulation (Figure 3 A). This region thus simultaneously encoded both
786 reward outcomes and the choices informed by these outcomes, highlighting its role in linking value to
787 intention processing in the brain. Additionally, we found a double dissociation in task coding between the
788 right dmPFC and left parietal cortex (Figure 4B), with the former only encoding tasks at the time of
789 decision-making, and the latter only encoding tasks during intention maintenance. Please note that due a
790 jittered inter-trial-interval, the decision-time and intention maintenance could be investigated
791 independently. This dissociation was assessed statistically by performing an ANOVA on the accuracy
792 values, using the factors 'time in trial' (time of decision vs intention maintenance) and 'ROI' (right dmPFC
793 vs left parietal cortex). We found moderate evidence for a time x ROI interaction (BF10 = 5.39, F(1,34) =
794 10.49, p = 0.04). Furthermore, the right dmPFC only encoded tasks at the time of decision (BF10 = 51.27,
795 t(34) = 3.51, p < 0.001), but not during intention maintenance (BF10 = 0.68, t(34) = 1.28, p = 0.10). The left
796 parietal cortex only encoded tasks during intention maintenance (BF10 > 150, t(34) = 4.20, p < 0.001), but
797 not at time of decision (BF10 = 0.19, t(34) = 0.09, p = 0.46). This double dissociation thus suggests a
798 temporal order of task processing in the brain, with the medial PFC transiently encoding chosen tasks at
799 the time of decision-making, and the left parietal cortex then maintaining that information until the tasks
800 can be executed. Lastly, we also assessed task information throughout the multiple demand network
801 (Duncan, 2010; Woolgar et al., 2015), and found tasks to be encoded in a contingency-invariant format
802 (Figure 4-1).
803 *Control analyses*: In order to provide further support for our main results, we decided to perform a
804 number of additional control analyses. First, we controlled for potential effects of RTs on task decoding
805 results. It has been pointed out before, that task information in the brain can at least partly be explained

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

806 through RT effects (Todd et al., 2013). Although others have found no such effects (Woolgar et al., 2014),
807 we decided to conservatively control for RT effects nonetheless, especially given that we found RT
808 differences between tasks (see above). We thus repeated the task decoding analyses, only first regressing
809 RT-related effects out of the data. We used the parietal and aMFG ROIs defined in the baseline analysis
810 and tested whether task information was still present after controlling for potential RT effects. We still
811 found the parietal cortex to encode tasks (4.61%, SEM = 0.65%, BF10 > 150, $t(34) = 6.99$, $p < 0.001$), and
812 also found the task coding to be reward-invariant (4.03%, SEM = 0.76%, BF10 > 150, $t(34) = 5.24$, $p <$
813 0.001). The same was true for the aMFG (4.66%, SEM = 0.89%, BF10 > 150, $t(34) = 5.19$, $p < 0.001$; and
814 3.71%, SEM = 1.16%, BF10 = 23.38, $t(34) = 3.18$, $p = 0.001$; respectively). Results in the baseline and xclass
815 analysis were equal in both regions, BFs10 >= 3.24, $ts(34) < 0.67$, $ps > 0.25$. These results thus mirror the
816 main analysis above, showing that RT-related variance cannot explain task decoding results in our
817 experiment.
818 Although overall error rates were low and choice biases were largely absent, it was still possible that
819 individual subjects showed excessively high error rates or strong choice biases, affecting task decoding
820 results. The influence of individual subjects should be relatively small given our large sample size, but we
821 still repeated the main analyses, excluding subjects with excessively high error rates and excessively
822 strong choice biases. Additionally, we excluded subjects in which regressors could not be estimated from
823 a sufficient number of trials (see Materials and Methods for more details). Using these highly conservative
824 exclusion criteria, we removed an additional 12 subjects from the sample, leading to a sample size of 23
825 subjects. Even though statistical power was considerably lower because of the smaller sample size, we
826 were still able to detect task information in the parietal cortex (5.20%, SEM = 0.79%, BF10 >150, $t(22) =$
827 6.54, $p < 0.001$), which was again reward-invariant (3.81%, SEM = 0.96%, BF10 = 96.61, $t(22) = 3.93$, $p <$
828 0.001), and the same was true for the aMFG (5.03%, SEM = 1.09%, BF10 >150, $t(22) = 4.60$, $p < 0.001$, and

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3.71%, SEM = 1.39%, BF10 = 7.34, t(22) = 2.66, p = 0.006, respectively). Therefore, neither error rates, nor choice biases can explain the reported task decoding results.

In order to validate the decoding procedure, we also extracted task decoding accuracies from a region not involved in performing this task, the primary auditory cortex. As expected, we found accuracies not to differ from chance level in this region (-0.36%, SEM = 0.93%, BF01 = 7.22, t(34) = 0.38, p = 0.64), showing that the task decoding analysis was not biased towards positive accuracy values. Lastly, we empirically estimated the chance level of our decoding analysis using permutation tests, in order to rule out a biased chance level. The estimated chance level was 49.98%, which did not differ from the theoretical chance level of 50% (BF01 > 150, t(34999) = 0.41, p = 0.67). Thus, comparing our decoding accuracies against a chance level of 50% was valid.

Discussion

Here, we investigated the effects of control over choice outcomes on outcome valuation and choice implementation. Subjects performed a probabilistic reward reversal learning task, in which they had control over the outcomes of their choices. They also performed a free choice task with non-contingent reward outcomes, in which outcomes were not under their direct control. Although we found reward contingency to modulate outcome valuation, we found no effects on choice implementation. Furthermore, we found two main brain regions to be crucial for encoding tasks and reward outcomes: the right dmPFC and the left parietal cortex (around the IPS). The dmPFC was found to encode chosen tasks at the time of decision-making, and simultaneously encoded reward outcome values, emphasizing its role in linking value-related with intentional control processes. While the parietal cortex encoded reward-prediction errors at the time of decision-making, it encoded chosen tasks during a subsequent maintenance phase. We found a double dissociation between both regions, with the dmPFC encoding tasks only at the time of decision-making, and the parietal cortex only during intention maintenance.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Control over choice outcomes affects outcome valuation but not choice implementation

Much previous research on the effects of reward motivation on cognition investigated the effects of reward prospect (Jimura et al., 2010; Dreisbach and Fischer, 2012). These findings demonstrated that positive reinforcement improves cognition, as compared to no reinforcement at all. However, an equally important and often overlooked property of reinforcement is the degree of control we have in reaching it. Sometimes, an action will cause outcomes in a fairly clear way (e.g. hitting a light switch), other times, that link will be less close (e.g. refreshing your Facebook timeline). Previous work on non-human primates has shown that the strength of such action-outcome contingencies modulates the neural processing of reward outcomes (Izquierdo et al., 2004; Chudasama et al., 2013). Our results show that this is also true in humans (see also Tricomi et al., 2004), and that neural representations of outcome values (and correlated processes) are amplified by reward contingency. Although somewhat weaker, evidence for reward learning signals points in the same direction. This is in line with predictions from gain-theories of motivation. It has been suggested that rewards increase the gain of subcortical dopaminergic neurons (Tobler et al., 2005), making them more sensitive to changes in rewards (see also Ikeda and Hikosaka, 2003; Thurley et al., 2008). We directly demonstrate such gain increases, in subcortical dopaminergic regions and beyond.

Importantly, in order for this value signal to lead to actual rewards, chosen behavior has to be implemented as intended first (see also Ruge et al., 2010). One might thus expect contingency to lead to stronger task shielding and coding (Dreisbach and Wenke, 2011), as the costs of confusing the two highly similar tasks are potentially high. However, we found no evidence for such effects. On the contrary, we found evidence for a similar or invariant coding of tasks across both contingency conditions. This finding informs current debates on the nature of task coding in the brain (Wisniewski, 2018). On the one hand, some have argued for flexible task coding especially in the fronto-parietal cortex (Woolgar et al., 2015; Qiao et al., 2017), often based on the multiple-demand network theory (Duncan, 2010). This account

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

predicts that task coding should be stronger when task demands are high (Woolgar et al., 2015), or when correct performance is rewarded (Etzel et al., 2016). Despite our efforts to replicate these findings in our data-set, we found no evidence for an influence of reward contingency on task coding. This was despite the fact that behavior differed between these conditions and that value-related signals were affected by reward contingency. One might argue that our analysis had insufficient statistical power to detect true effects, though we believe this to be unlikely. First, we decided to have a relatively large sample size (n=35). Second, additional control analyses showed that other analyses, matched for statistical power, do show significant results.

On the other hand, others have argued that the same task representations could be used in multiple different situations (i.e. 'multiplexing' of task information), and that this allows us to flexibly react to novel and changing demands (Botvinick and Cohen, 2014). Multiplexing predicts that task information should be invariant across different contexts (Levine and Schwarzbach, 2017), which has been shown previously (Zhang et al., 2013; Wisniewski et al., 2016; Loose et al., 2017). Here, we replicate and extend these findings, by showing that tasks are encoded in an outcome-contingency-invariant format in frontal and parietal brain regions, strengthening the idea of multiplexing of task information in the brain. One possible alternative explanation for this finding might be that subjects were highly trained in performing the two tasks, and were at their performance ceiling. This might make a modulation of task coding too small to detect. Although we cannot fully exclude this interpretation, we want to point out that contingency did have robust effects on behavior. Also, most related previous experiments trained their subjects, those that found effects (Woolgar et al., 2015; Etzel et al., 2016) and those that did not (Wisniewski et al., 2016). We thus believe this alternative explanation to be unlikely. Overall, our task decoding results are in line with the idea of multiplexing of task information in the brain. Future research will have to test more directly which environmental conditions lead to multiplexing of task information in the brain, and which do not.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The roles of dmPFC and parietal cortex in value-related and task-related processes

The dmPFC is a key region for decision-making in dynamic environments. It is supports effort-based foraging choices (Wisniewski et al., 2015b), and here we extend this finding by showing its involvement in a different task with different outcomes (reward reversal learning). The dmPFC is important for cognitive control, supporting rule and action selection (Rowe et al., 2008), working memory (Taylor et al., 2004), and processing uncertainty (Volz et al., 2003). It has further been associated with valuation processes, anticipating both positive and negative outcomes (Jensen et al., 2003; Knutson et al., 2003), and encoding reward prediction errors (Vassena et al., 2014). In this experiment, we demonstrated that the dmPFC is specifically involved in encoding tasks only at the time at which a choice is made, other regions later maintain that choice outcome until it can be executed. We also demonstrated the dmPFC to encode outcome values at the same time. Please note that we do not claim this value signal to only represent the magnitude of reward outcomes, it might also represent related processes (e.g. attention). Nevertheless, the cause of this effect are different outcome values, and this highlights the importance of dmPFC in linking valuation to strategic decision-making, providing an explanation to how it might support goal-directed behavior (Viard et al., 2011).

The second key region identified in this experiment was the left parietal cortex, especially around the IPS. This brain region encodes prediction errors (Daw and Doya, 2006; Matsumoto et al., 2007; Katahira et al., 2015), which might signal model updating (Behrens et al., 2007; Walton et al., 2007; Rutledge et al., 2010). Alternatively, it has been suggested that the parietal cortex signals surprise, and does not reflect model updating (O'Reilly et al., 2013). Our findings are more in line with surprise signaling, the only brain region possibly involved in model updating in our experiment was the anterior medial PFC (see also (Braem et al., 2013). The parietal cortex is also a key region for cognitive control (Ruge et al., 2009), and working memory (Christophel et al., 2017). It is part of the multiple demand network (Duncan, 2010; Fedorenko et al., 2013), a set of brain regions characterized by their high flexibility to adapt to changing demands.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Previous work on non-human primates demonstrated that the prefrontal cortex flexibly switches between representing different control-related information within single trials (Sigala et al., 2008; Stokes et al., 2013). Our results show that the parietal cortex in humans exhibits similar flexibility. It switches between encoding control-related and value-related variables within single trials. This provides compelling evidence for the flexibility of the parietal cortex in adapting to rapidly changing task demands. In the future, it will be interesting to assess whether and how the parietal cortex links value-related and control-related variables. Given its involvement in foraging behavior (Sugrue, 2004), the previous choice and outcome history likely affects current choice representations in this brain region. Future experiments will shed more light on how exactly our choice history shapes our current choices.

Conclusion

In this experiment, we assessed whether controlling outcomes affects outcome valuation and choice implementation in the brain. By comparing choices that are informed by expected outcomes as well as choices that are not, we linked largely parallel research on 'free choice' (Libet et al., 1983) and value-based decision-making (Hampton and O'Doherty, 2007), which has been long overdue. While we found strong effects on outcome valuation, we found no such effects on choice implementation. Our results further highlight the importance of both the dmPFC and parietal cortex in bridging valuation and executive processes in the brain. Both regions have been involved in processing task choices and their reward outcomes, flexibly switching between encoding value-related and task-related information.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Andraszewicz S, Scheibehenne B, Rieskamp J, Grasman R, Verhagen J, Wagenmakers E-J (2015) An Introduction to Bayesian Hypothesis Testing for Management Research. J Manag 41:521–543.

Arrington CM, Logan GD (2004) The cost of a voluntary task switch. Psychol Sci 15:610–615.

Bai Y, Katahira K, Ohira H (2014) Dual learning processes underlying human decision-making in reversal learning tasks: functional significance and evidence from the model fit to human behavior. Front Psychol 5 Available at: https://www.frontiersin.org/articles/10.3389/fpsyg.2014.00871/full [Accessed May 31, 2018].

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221.

Bhandari A, Gagne C, Badre D (2018) Just above Chance: Is It Harder to Decode Information from Human Prefrontal Cortex Blood Oxygenation Level-dependent Signals? J Cogn Neurosci:1–26.

Botvinick MM, Cohen JD (2014) The Computational and Neural Basis of Cognitive Control: Charted Territory and New Frontiers. Cogn Sci 38:1249–1285.

Braem S, King JA, Korb FM, Krebs RM, Notebaert W, Egner T (2013) Affective Modulation of Cognitive Control is Determined by Performance-Contingency and Mediated by Ventromedial Prefrontal and Cingulate Cortex. J Neurosci 33:16961–16970.

Bunge SA, Kahn I, Wallis JD, Miller EK, Wagner AD (2003) Neural Circuits Subserving the Retrieval and Maintenance of Abstract Rules. J Neurophysiol 90:3419–3428.

Burnham KP, Anderson DR (2004) Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociol Methods Res 33:261–304.

Cacioppo JT, Petty RE, Chuan Feng Kao (1984) The Efficient Assessment of Need for Cognition. J Pers Assess 48:306.

Carver CS, White TL (1994) Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS scales. J Pers Soc Psychol 67:319.

Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J-D (2017) The Distributed Nature of Working Memory. Trends Cogn Sci 21:111–124.

Chudasama Y, Daniels TE, Gorrin DP, Rhodes SEV, Rudebeck PH, Murray EA (2013) The Role of the Anterior Cingulate Cortex in Choices based on Reward Value and Reward Contingency. Cereb Cortex 23:2884–2898.

Collins AGE, Ciullo B, Frank MJ, Badre D (2017) Working Memory Load Strengthens Reward Prediction Errors. J Neurosci 37:4332–4342.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

974 Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and
975 classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19:261–270.

976 Daw ND, Doya K (2006) The computational neurobiology of learning and reward. Curr Opin Neurobiol
977 16:199–204.

978 Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-Based Influences on Humans'
979 Choices and Striatal Prediction Errors. Neuron 69:1204–1215.

980 Domenech P, Redouté J, Koechlin E, Dreher J-C (2018) The Neuro-Computational Architecture of Value-
981 Based Selection in the Human Brain. Cereb Cortex 28:585–601.

982 Doya K (2008) Modulators of decision making. Nat Neurosci 11:410–416.

983 Dreisbach G, Fischer R (2012) The role of affect and reward in the conflict-triggered adjustment of
984 cognitive control. Front Hum Neurosci 6 Available at:
985 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3533233/ [Accessed October 30, 2014].

986 Dreisbach G, Wenke D (2011) The shielding function of task sets and its relaxation during task switching.
987 J Exp Psychol Learn Mem Cogn 37:1540–1546.

988 Duncan J (2010) The multiple-demand (MD) system of the primate brain: mental programs for intelligent
989 behaviour. Trends Cogn Sci 14:172–179.

990 Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have
991 inflated false-positive rates. Proc Natl Acad Sci 113:7900–7905.

992 Elliott R, Newman JL, Longe OA, William Deakin JF (2004) Instrumental responding for rewards is
993 associated with enhanced neuronal response in subcortical reward systems. NeuroImage
994 21:984–990.

995 Etzel JA, Cole MW, Zacks JM, Kay KN, Braver TS (2016) Reward Motivation Enhances Task Coding in
996 Frontoparietal Cortex. Cereb Cortex 26:1647–1659.

997 Fedorenko E, Duncan J, Kanwisher N (2013) Broad domain generality in focal regions of frontal and
998 parietal cortex. Proc Natl Acad Sci 110:16616–16621.

999 Forstmann BU, Wagenmakers E-J (2015) An Introduction to Model-Based Cognitive Neuroscience.
1000 Springer.

1001 Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RS (1994) Statistical parametric maps
1002 in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210.

1003 Gilbert SJ (2011) Decoding the Content of Delayed Intentions. J Neurosci 31:2888–2894.

1004 Hampton AN, O'Doherty JP (2007) Decoding the neural substrates of reward-related decision making
1005 with functional MRI. Proc Natl Acad Sci 104:1377–1382.

1006 Haxby JV (2012) Multivariate pattern analysis of fMRI: The early beginnings. NeuroImage 62:852–855.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1007 Haynes J-D (2015) A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives.
1008 Neuron 87:257–270.

1009 Haynes J-D, Sakai K, Rees G, Gilbert SJ, Frith C, Passingham RE (2007) Reading Hidden Intentions in the
1010 Human Brain. Curr Biol 17:323–328.

1011 Hebart MN, Görgen K, Haynes J-D (2014) The Decoding Toolbox (TDT): A versatile software package for
1012 multivariate analyses of functional imaging data. Front Neuroinformatics 8:88.

1013 Ikeda T, Hikosaka O (2003) Reward-Dependent Gain and Bias of Visual Responses in Primate Superior
1014 Colliculus. Neuron 39:693–700.

1015 Izquierdo A, Suda RK, Murray EA (2004) Bilateral Orbital Prefrontal Cortex Lesions in Rhesus Monkeys
1016 Disrupt Choices Guided by Both Reward Value and Reward Contingency. J Neurosci 24:7540–
1017 7548.

1018 Jensen J, McIntosh AR, Crawley AP, Mikulis DJ, Remington G, Kapur S (2003) Direct Activation of the
1019 Ventral Striatum in Anticipation of Aversive Stimuli. Neuron 40:1251–1257.

1020 Jimura K, Locke HS, Braver TS (2010) Prefrontal cortex mediation of cognitive enhancement in rewarding
1021 motivational contexts. Proc Natl Acad Sci 107:8871–8876.

1022 Kahnt T, Park SQ, Haynes J-D, Tobler PN (2014) Disentangling neural representations of value and
1023 salience in the human brain. Proc Natl Acad Sci 111:5000–5005.

1024 Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci
1025 8:679–685.

1026 Kaplan JT, Man K, Greening SG (2015) Multivariate cross-classification: applying machine learning
1027 techniques to characterize abstraction in neural representations. Front Hum Neurosci 9
1028 Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4373279/ [Accessed December 10,
1029 2015].

1030 Katahira K, Matsuda Y-T, Fujimura T, Ueno K, Asamizuya T, Suzuki C, Cheng K, Okanoya K, Okada M
1031 (2015) Neural basis of decision making guided by emotional outcomes. J Neurophysiol
1032 113:3056–3068.

1033 Knutson B, Fong GW, Bennett SM, Adams CM, Hommer D (2003) A region of mesial prefrontal cortex
1034 tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI.
1035 NeuroImage 18:263–272.

1036 Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl
1037 Acad Sci 103:3863–3868.

1038 Levine SM, Schwarzbach J (2017) Decoding of auditory and tactile perceptual decisions in parietal
1039 cortex. NeuroImage 162:297–305.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1040 Libet B, Gleason CA, Wright EW, Pearl DK (1983) Time of conscious intention to act in relation to onset of
1041 cerebral activity (readiness-potential) the unconscious initiation of a freely voluntary act. Brain
1042 106:623–642.

1043 Lien M-C, Ruthruff E, Remington RW, Johnston JC (2005) On the limits of advance preparation for a task
1044 switch: do people prepare all the task some of the time or some of the task all the time? J Exp
1045 Psychol Hum Percept Perform 31:299–315.

1046 Loose LS, Wisniewski D, Rusconi M, Goschke T, Haynes J-D (2017) Switch-Independent Task
1047 Representations in Frontal and Parietal Cortex. J Neurosci 37:8033–8042.

1048 Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction
1049 errors of action values. Nat Neurosci 10:647–656.

1050 Mertens G, De Houwer J (2016) Potentiation of the startle reflex is in line with contingency reversal
1051 instructions rather than the conditioning history. Biol Psychol 113:91–99.

1052 Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang X, Just M, Newman S (2004) Learning to
1053 decode cognitive states from brain images. Mach Learn 57:145–175.

1054 Mobbs D, Hassabis D, Yu R, Chu C, Rushworth MFS, Boorman ED, Dalgleish T (2013) Foraging under
1055 Competition: The Neural Basis of Input-Matching in Humans. J Neurosci 33:9866–9872.

1056 O'Doherty JP, Hampton A, Kim H (2007) Model-Based fMRI and Its Application to Reward Learning and
1057 Decision Making. Ann N Y Acad Sci 1104:35–53.

1058 O'Reilly JX, Schüffelgen U, Cuell SF, Behrens TEJ, Mars RB, Rushworth MFS (2013) Dissociable effects of
1059 surprise and model update in parietal and anterior cingulate cortex. Proc Natl Acad Sci
1060 110:E3660–E3669.

1061 Patton JH, Stanford MS, Barratt ES (1995) Factor structure of the Barratt impulsiveness scale. J Clin
1062 Psychol 51:768–774.

1063 Peirce JW (2007) PsychoPy—Psychophysics software in Python. J Neurosci Methods 162:8–13.

1064 Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors
1065 underpin reward-seeking behaviour in humans. Nature 442:1042–1045.

1066 Qiao L, Zhang L, Chen A, Egner T (2017) Dynamic Trial-by-Trial Recoding of Task-Set Representations in
1067 the Frontoparietal Cortex Mediates Behavioral Flexibility. J Neurosci 37:11037–11050.

1068 Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian tests for accepting and rejecting
1069 the null hypothesis. Psychon Bull Rev 16:225–237.

1070 Rowe J, Hughes L, Eckstein D, Owen AM (2008) Rule-Selection and Action-Selection have a Shared
1071 Neuroanatomical Basis in the Human Prefrontal and Parietal Cortex. Cereb Cortex N Y NY
1072 18:2275–2285.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1073 Rubinstein JS, Meyer DE, Evans JE (2001) Executive control of cognitive processes in task switching. J Exp
1074 Psychol Hum Percept Perform 27:763–797.

1075 Ruge H, Braver T, Meiran N (2009) Attention, intention, and strategy in preparatory control.
1076 Neuropsychologia 47:1670–1685.

1077 Ruge H, Müller S, Braver T (2010) Anticipating the consequences of action: An fMRI study of intention-
1078 based task preparation. Psychophysiology 47:1019–1027.

1079 Rutledge RB, Dean M, Caplin A, Glimcher PW (2010) Testing the Reward Prediction Error Hypothesis
1080 with an Axiomatic Model. J Neurosci 30:13525–13536.

1081 Schultz W (2016) Dopamine reward prediction-error signalling: a two-component response. Nat Rev
1082 Neurosci 17:183–195.

1083 Sigala N, Kusunoki M, Nimmo-Smith I, Gaffan D, Duncan J (2008) Hierarchical coding for sequential task
1084 events in the monkey prefrontal cortex. Proc Natl Acad Sci 105:11969–11974.

1085 Soon CS, Brass M, Heinze H-J, Haynes J-D (2008) Unconscious determinants of free decisions in the
1086 human brain. Nat Neurosci 11:543–545.

1087 Soon CS, He AH, Bode S, Haynes J-D (2013) Predicting free choices for abstract intentions. Proc Natl Acad
1088 Sci 110:6217–6222.

1089 Sperduti M, Delaveau P, Fossati P, Nadel J (2011) Different brain structures related to self- and external-
1090 agency attribution: a brief review and meta-analysis. Brain Struct Funct 216:151–157.

1091 Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic Coding for Cognitive Control
1092 in Prefrontal Cortex. Neuron 78:364–375.

1093 Sugrue LP (2004) Matching Behavior and the Representation of Value in the Parietal Cortex. Science
1094 304:1782–1787.

1095 Sutton RS, Barto AG (1990) Time-derivative models of Pavlovian reinforcement. In: Learning and
1096 computational neuroscience: Foundations of adaptive networks (Gabriel M, Moore J, eds), pp
1097 497–537. Cambridge, MA, US: The MIT Press.

1098 Taylor SF, Welsh RC, Wager TD, Luan Phan K, Fitzgerald KD, Gehring WJ (2004) A functional
1099 neuroimaging study of motivation and executive function. NeuroImage 21:1045–1054.

1100 Thurley K, Senn W, Lüscher H-R (2008) Dopamine Increases the Gain of the Input-Output Response of
1101 Rat Prefrontal Pyramidal Neurons. J Neurophysiol 99:2985–2997.

1102 Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive Coding of Reward Value by Dopamine Neurons.
1103 Science 307:1642–1645.

1104 Todd MT, Nystrom LE, Cohen JD (2013) Confounds in multivariate pattern analysis: Theory and rule
1105 representation case study. NeuroImage 77:157–165.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1106 Torrubia R, Ávila C, Moltó J, Caseras X (2001) The Sensitivity to Punishment and Sensitivity to Reward
1107 Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. Personal
1108 Individ Differ 31:837–862.

1109 Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of Caudate Activity by Action Contingency. Neuron
1110 41:281–292.

1111 Vassena E, Krebs RM, Silvetti M, Fias W, Verguts T (2014) Dissociating contributions of ACC and vmPFC in
1112 reward prediction, outcome, and choice. Neuropsychologia 59:112–123.

1113 Viard A, Doeller CF, Hartley T, Bird CM, Burgess N (2011) Anterior Hippocampus and Goal-Directed
1114 Spatial Decision Making. J Neurosci 31:4613–4621.

1115 Volz KG, Schubotz RI, von Cramon DY (2003) Predicting events of varying probability: uncertainty
1116 investigated by fMRI. NeuroImage 19:271–280.

1117 Wagenmakers E-J (2007) A practical solution to the pervasive problems of p values. Psychon Bull Rev
1118 14:779–804.

1119 Walton ME, Croxson PL, Behrens TEJ, Kennerley SW, Rushworth MFS (2007) Adaptive decision making
1120 and value in the anterior cingulate cortex. NeuroImage 36:T142–T154.

1121 Waskom ML, Kumaran D, Gordon AM, Rissman J, Wagner AD (2014) Frontoparietal Representations of
1122 Task Context Support the Flexible Control of Goal-Directed Cognition. J Neurosci 34:10743–
1123 10755.

1124 Wisniewski D (2018) Context-Dependence and Context-Invariance in the Neural Coding of Intentional
1125 Action. Front Psychol 9 Available at:
1126 https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02310/full [Accessed January 10,
1127 2019].

1128 Wisniewski D, Goschke T, Haynes J-D (2016) Similar coding of freely chosen and externally cued
1129 intentions in a fronto-parietal network. NeuroImage 134:450–458.

1130 Wisniewski D, Reverberi C, Momennejad I, Kahnt T, Haynes J-D (2015a) The Role of the Parietal Cortex in
1131 the Representation of Task–Reward Associations. J Neurosci 35:12355–12365.

1132 Wisniewski D, Reverberi C, Tusche A, Haynes J-D (2015b) The Neural Representation of Voluntary Task-
1133 Set Selection in Dynamic Environments. Cereb Cortex 25:4715–4726.

1134 Woolgar A, Afshar S, Williams MA, Rich AN (2015) Flexible Coding of Task Rules in Frontoparietal Cortex:
1135 An Adaptive System for Flexible Cognitive Control. J Cogn Neurosci:1–17.

1136 Woolgar A, Golland P, Bode S (2014) Coping with confounds in multivoxel pattern analysis: What should
1137 we do about reaction time differences? A comment on Todd, Nystrom & Cohen 2013.
1138 NeuroImage 98:506–512.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

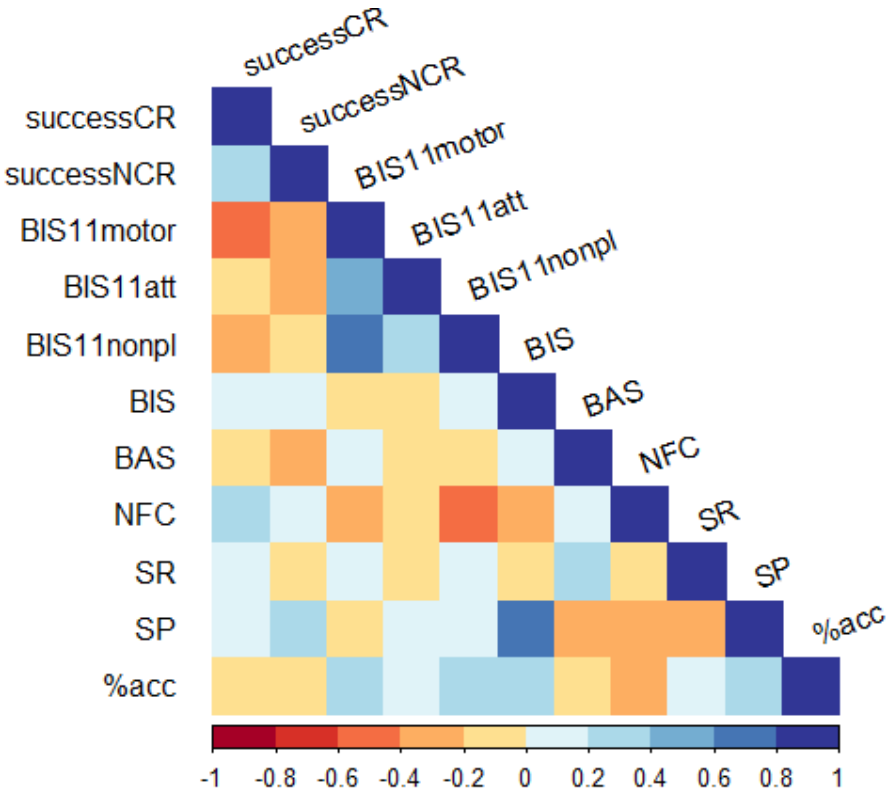
1139 Zhang J, Kriegeskorte N, Carlin JD, Rowe JB (2013) Choosing the Rules: Distinct and Overlapping
1140 Frontoparietal Representations of Task Rules for Perceptual Decisions. J Neurosci 33:11852–
1141 11862.

1142

1143

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Material



Supplementary Figure 1: Correlation analysis. An additional exploratory analysis was performed to correlate performance, questionnaire measures, and decoding accuracies (baseline task decoding, from the parietal cortex cluster). Depicted are all pairwise correlations between % high reward choices in CR trials (successCR), % high reward chocies in NCR trials (successNCR), motor impulsivity (BIS11motor), attentional impulsivity (BIS11att), non-planning impulsivity (BIS11nonpl), behavioral inhibition (BIS), behavioral approach (BAS), need for cognition (NFC), sensitivity to reward (SR), sensitivity to punishment (SP), and decoding accuracies in the baseline task decoding analysis in the parietal cortex (%acc). The plot was generated using the *corrplot* package in R.

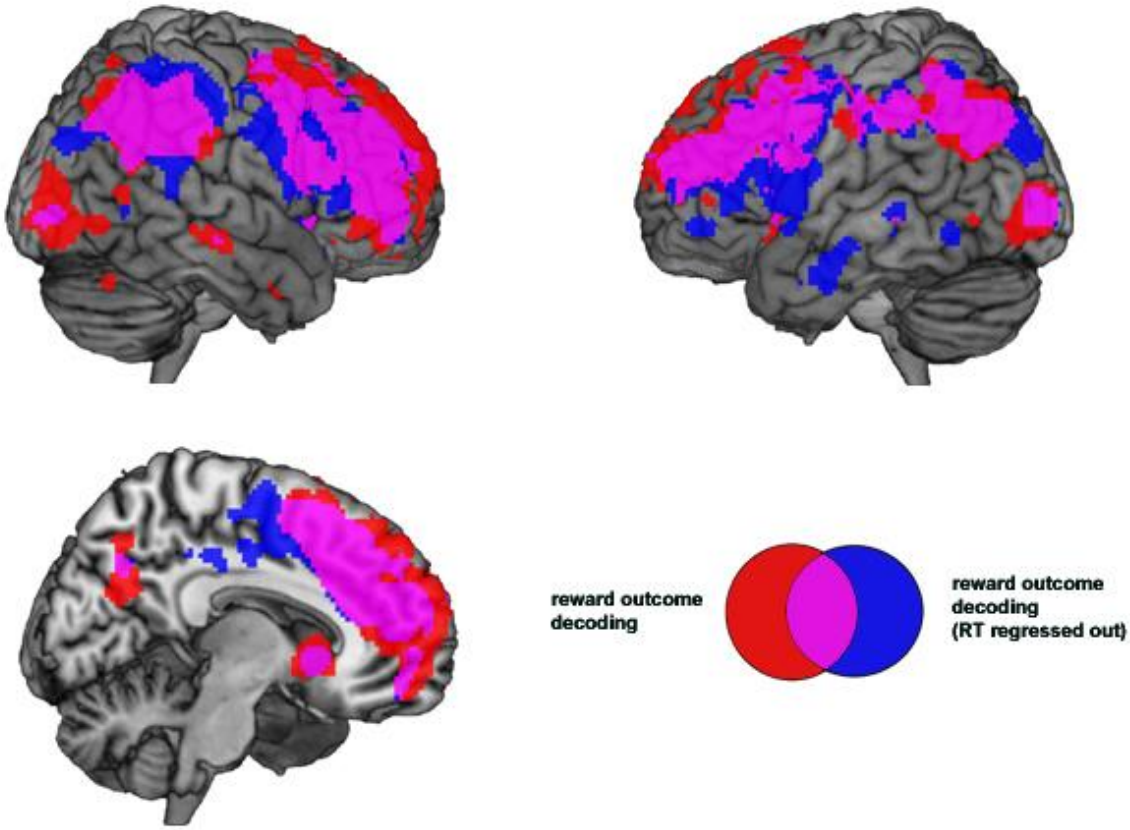
Despite this descriptive approach, we also tested the strength of these correlations in a Bayesian framework (using *bayes.cor.test* form the BayesianFirstAid package in R). Although our conclusions are based on this correlation analysis, we also report classically estimated correlations and corresponding p-values for the interested reader. We expected successful performance to be correlated with higher need for cognition, lower impulsivity, and higher sensitivity to reward. We also expected task coding to be related to task performance, with better performance related to higher accuracies. Higher accuracies could also be related to lower impulsivity, higher sensitivity to reward, and higher need for cognition. Successful performance was correlated with impulsivity, as measured using the BIS11, r = -.34 (95%CI = [-.62 -.024]; classical estimation r = -.33, p = 0.052), with impulsive subjects being less successful in

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

performing the reversal learning task. The BIS11 further splits impulsivity into three components: attentional, motor, and non-planning impulsivity. The observed correlation was mostly driven by motor impulsivity (r = -.45, 95%CI = [-.70 -.15]; r = -.47, p = 0.004), but not by non-planning (r = -.19, 95%CI = [-.52 .14]; r = -.20, p = 0.24) or attentional impulsivity (r = -.11, 95%CI = [-.45 .02]; r = -.11, p = 0.51). There was no correlation of success with either sensitivity to reward (r = .04, 95%CI = [-.29 .38]; r = .06, p = 0.71), or the need for cognition (r = .26, 95%CI = [-.07 .56]; r = .26, p = 0.11), despite the fact the need for cognition seems to be associated with reward decision-making (Sandra and Otto 2018). A qualitatively similar pattern was evident for decoding accuracies, extracted during intention maintenance from the parietal cortex. Correlations with impulsivity (r = -.27, 95%CI = [-.57 .07]; r = -.32, p = 0.053), sensitivity to reward (r = -.04, 95%CI = [-.38 .31], r = .17, p = 0.30), and need for cognition (r = .09, 95%CI = [-.24 .41]; r = -.24, p – 0.16) were at least similar numerically to the correlations with task success. Given that the evidence was somewhat weaker in this analysis, results should be interpreted with care however. Overall, task performance and to a lesser degree decoding accuracies seem to be most strongly related to impulsivity, and not to sensitivity to reward or need for cognition. This unexpected link to impulsivity should be addressed directly in future research.

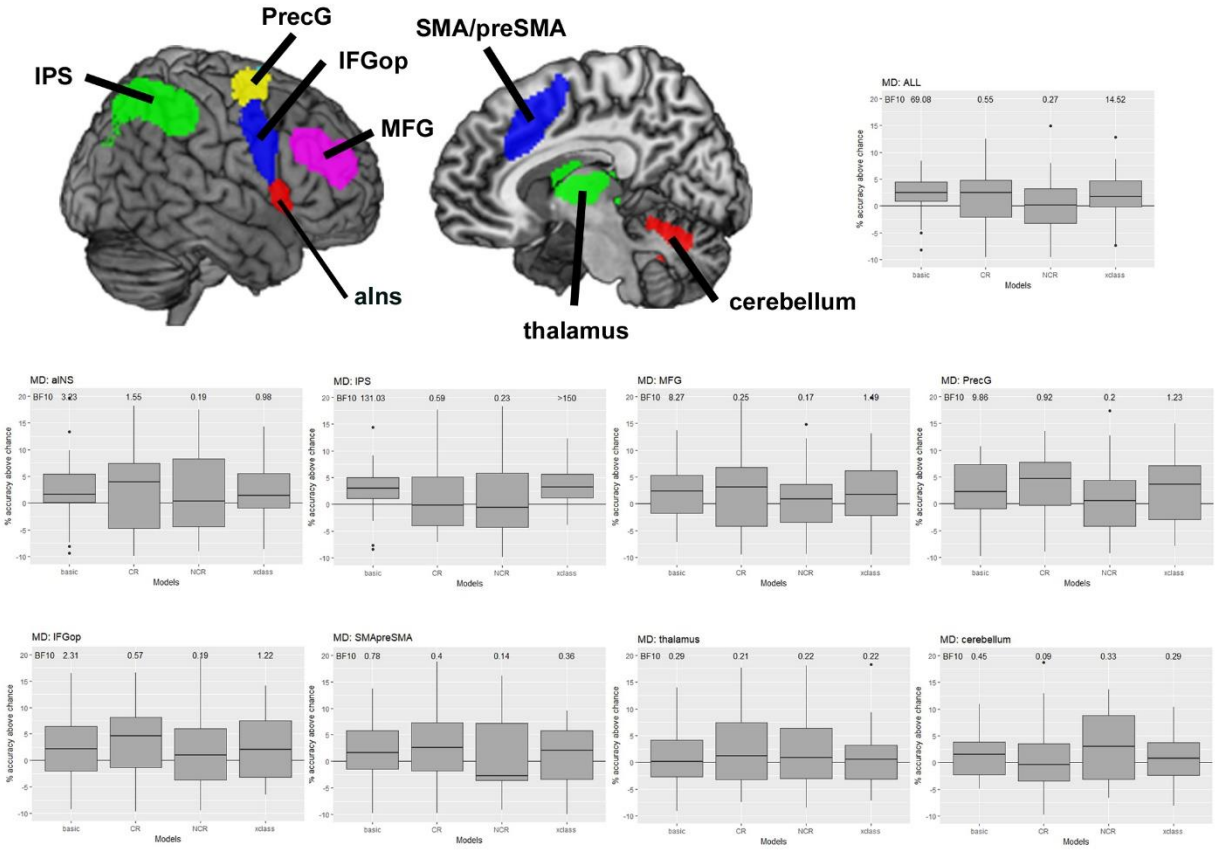
Sandra, D.A., & Otto A.R. (2018) Cognitive Capacity Limitations and Need for Cognition Differentially Predict Reward-Induced Cognitive Effort Expenditure. *Cognition,* 172: 101–6.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Supplementary Figure 2. Controlling RT-effects in reward outcome coding. We repeated the reward outcome decoding analysis, using a similar first-level GLM to estimate signals (4 regressors: each combination of high vs low reward, contingent vs non-contingent reward, locked to feedback onset). Additionally, we added regressors of non-interest capturing RT-related variance in the data. The rest of the analysis was identical to the reward outcome decoding analysis presented in the main body of the text. Results from the reward outcome decoding analysis (red), and the same analysis with RT-related effects regressed out of the data (blue) are depicted. As can be seen, the overlap between both analyses is substantial. Results depicted at $p < 0.05$ (FWE, corrected at the voxel level). This indicates that controlling for RT did not strongly alter our results.

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



1191

1192 **Supplementary Figure 3.** Task information in the multiple demand (MD) network. Depicted are task
1193 decoding results in the bilateral functional ROIS provided by Fedorenko, Duncan, & Kanwisher (2013),
1194 specifically the anterior insula (aINS), cerebellum, inferior frontal gyrus pars opercularis (IFGop),
1195 intraparietal sulcus (IPS), middle frontal gyrus (MFG), pre-central gyrus (precG), supplementary and pre-
1196 supplementary motor area (SMA/preSMA), as well as thalamus. Averaging across all MD regions, we
1197 found strong evidence for the presence of task information (2.23%, SEM = 0.61%, BF10 = 69.08, t(34) =
1198 3.63, $p < 0.001$, Figure 1). We then tested whether accuracies were higher in CR trials than in NCR trials,
1199 using the same analysis as used for the regions identified in the main task decoding analysis. We found no
1200 evidence for a higher accuracy in CR, as compared to NCR trials (BF10 = 0.37, t(34) = 0.68, p = 0.24).
1201 Furthermore, we found task coding to be contingency-invariant, using a cross-classification approach
1202 (2.02%, SEM = 0.67%, BF10 = 14.52, t(34) = 2.97, p = 0.002). Accuracies in the baseline and cross-
1203 classification analysis did not differ (BF10 = 5.11, t(34) = 0.40, p = 0.68). This suggests that the MD network
1204 encodes tasks in a contingency-invariant fashion, and shows that the current context does not affect task
1205 coding in the MD network. This is despite the clear effects contingency has on the coding of reward
1206 outcomes.

1207 Looking at individual MD regions, we found task information in the aINS (2.25%, SEM = 1.00%, BF10 =
1208 3.23, t(34) = 2.24, p = 0.01), IPS (2.83%, SEM = 0.72%, BF10 = 131.02, t(34) = 3.88, $p < 0.001$), MFG (2.44%,

bioRxiv preprint doi: https://doi.org/10.1101/375642; this version posted January 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SEM = 0.90%, BF10 = 8.26, t(34) = 2.71, p = 0.005), precentral gyrus (2.48%, SEM = 0.87, BF10 = 9.86, t(34) = 2.79, p = 0.004), but not in the cerebellum (0.85%, SEM = 0.90%, BF10 = 0.44, t(34) = 0.94, p = 0.17), IFGop (2.11%, SEM = 1.02%, BF10 = 2.31, t(34) = 2.06, p = 0.02) SMA/preSMA (1.48%, SEM = 1.07%, BF10 = 0.77, t(34) = 1.37, p = 0.08), and thalamus (0.58%, SEM = 1.06%, BF10 = 0.29, t(34) = 0.54, p = 0.29). None of these regions showed a higher accuracy in CR than in NCR trials (BFs10 <= 0.60, ts(34) < 1.19, ps > 0.12). However, in all of those regions the accuracy in the baseline and xclass analyses was equal (BFs10 >= 3.47, ts(34) < 1.00, ps > 0.32). In sum, we did not find our reward manipulation to affect task coding in the MD network. We did find contingency-invariant task information in this network however. Also, not all parts of the MD network seemed to be encoding tasks in our experiment.

Fedorenko E, Duncan J, Kanwisher N. 2013. Broad domain generality in focal regions of frontal and parietal cortex. P Natl Acad Sci USA. 110:16616–16621.