Published in final edited form as:
Neuroimage. 2022 February 15; 247: 118821. doi:10.1016/j.neuroimage.2021.118821.

Similar network compositions, but distinct neural dynamics underlying belief updating in environments with and without explicit outcomes

Vincenzo G. Fiore[a,b,*], **Xiaosi Gu**[a,b,c,*]

[a]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, United States

[b]Center for Computational Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10027, United States

[c]Nash Family Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, United States

Abstract

Classic decision theories typically assume the presence of explicit value-based outcomes after action selections to update beliefs about action-outcome contingencies. However, ecological environments are often opaque, and it remains unclear whether the neural dynamics underlying belief updating vary under conditions characterized by the presence or absence of such explicit value-based information, after each choice selection. We investigated this question in healthy humans ($n = 28$) using Bayesian inference and two multi-option fMRI tasks: a multi-armed bandit task, and a probabilistic perceptual task, respectively with and without explicit value-based feedback after choice selections. Model-based fMRI analysis revealed a network encoding belief updating which did not change depending on the task. More precisely, we found a confidence-building network that included anterior hippocampus, amygdala, and medial prefrontal cortex (mPFC), which became more active as beliefs about action-outcome probabilities were confirmed by newly acquired information. Despite these consistent responses across tasks, dynamic causal modeling estimated that the network dynamics changed depending on the presence or absence of trial-by-trial value-based outcomes. In the task deprived of immediate feedback, the hippocampus increased its influence towards both amygdala and mPFC, in association with increased strength in the confidence signal. However, the opposite causal relations were found (i.e., from both mPFC and amygdala towards the hippocampus), in presence of immediate outcomes. This finding revealed an asymmetric relationship between decision confidence computations, which were based on similar computational models across tasks, and neural implementation, which varied depending on the availability of outcomes after choice selections.

[*]Corresponding authors: vincenzo.fiore@mssm.edu (V.G. Fiore), xiaosi.gu@mssm.edu (X. Gu).

Credit author statement
VGF designed the study and the computational models, ran the analyses, wrote the manuscript. XG designed the study, wrote the manuscript.

Supplementary materials
Supplementary material associated with this article can be found, in the online version, at doi: 10.1016/j.neuroimage.2021.118821.

Keywords

Confidence; Anterior hippocampus; Amygdala; Medial prefrontal cortex; Dynamic causal modeling

1. Introduction

In classic decision-making theories, the presence of explicit outcomes (e.g. based on values) after a choice selection is considered a crucial source of information for belief updating and behavioral adaptability, e.g. by triggering prediction error signals (Glimcher, 2011; Rangel et al., 2008). This line of research has yielded fruitful results, most prominently in the identification of the neurocomputational mechanisms underlying reinforcement learning (Dabney et al., 2020; Schultz et al., 1997) and belief updates in changing environments (Behrens et al., 2007; McGuire et al., 2014; Soltani and Izquierdo, 2019). In real life, however, decisions are often made in opaque environments where outcomes can be sporadic or temporarily inaccessible. Despite this opacity, we can still form and update beliefs about how likely our chosen actions are to deliver what we need or want, based on other sources of information (Ma and Jazayeri, 2014; Pouget et al., 2013). For instance, previous studies have reported that, in probabilistic environments without explicit outcomes, the anterior hippocampus monitors the entropy in sequences of visual or auditory events in order to generate expectations regarding future stimuli and guide behavior accordingly (Harrison et al., 2006; Krug et al., 2014; Strange et al., 2005; Tobia et al., 2012). However, it is yet to be determined whether action-outcome belief updating, under different conditions of access to immediate feedback, relies on different neural patterns or network dynamics.

In Bayesian terms, beliefs are represented as probability distributions associating one's actions with one or more known outcomes (Fleming and Daw, 2017; Kording and Wolpert, 2006; Payzan-LeNestour and Bossaerts, 2011). These distributions of probabilities are taken into account in decision-making (Orban and Wolpert, 2011; Sanders et al., 2016) and are encoded by the activity of neuronal populations (Ma et al., 2006; Pouget et al., 2013; Rich et al., 2015). In this sense, decision confidence (*c*), and its complementary decision uncertainty (Adler and Ma, 2018; Atiya et al., 2020; Meyniel et al., 2015a, 2015b), respectively describe the estimated subjective probability that a chosen action will produce a desired outcome, given prior beliefs [formally: p(desired outcome | priors, action)] and the complementary probability (*1-c*) it will produce any outcome but the desired one. In continuous decision making, these estimates are updated on the basis of new evidence, generating posterior beliefs, which are available for future choice selections. Under conditions in which sensory stimuli are unambiguous and reliable (Ma and Jazayeri, 2014), confirming accumulating evidence leads to narrow distributions and precise action-outcome beliefs, whereas conflicting information leads to wide distributions and imprecision (Meyniel et al., 2015a, 2015b; Payzan-LeNestour et al., 2013; Pouget et al., 2016).

Here, we aimed at investigating the neural dynamics underlying the update of these estimated probabilities or beliefs, with or without immediate value-based outcomes. To this end we modified (Fig. 1; cf. Fiore et al., 2021) a probabilistic perceptual task, which does not provide value based feedbacks after choice selections (Adams et al., 2018; Huq et al., 1988; Phillips and Edwards, 1966), and a non-stationary multi-armed bandit task, which provides stochastic value-based feedbacks after each choice (O'Doherty et al., 2017; Robbins, 1952; Sutton and Barto, 1998). We used a multi-option design instead of the traditional two-option setup to allow for non-binary decision-making (Churchland and Ditterich, 2012; Churchland et al., 2008; Tajima et al., 2019), thus preventing the participants from using any exclusion/confirmation criterion as a strategy for decision-making. The tasks also employed three categorical types of data and identical volatility (i.e. average number of trials before reversal), therefore conflating noise and surprise (McGuire et al., 2014; Nassar et al., 2016, 2019) and providing participants with discrete evidence for the belief update. This design was conceived to foster nuanced transitions between trials characterized by high model-estimated subjective confidence (i.e., near certainty in one's mind that a chosen selection will yield a desired outcome) and those characterized by low model-estimated subjective confidence, or high uncertainty (i.e., all available actions are estimated to yield the desired outcome with a probability close to chance). To allow a comparison across tasks, we employed two similar Bayesian learner models to estimate subject-specific, trial-by-trial, belief updating and associated choice-related confidence, replicating the actual behavior recorded in healthy volunteers ($N = 28$) being scanned with functional magnetic resonance imaging (fMRI). Finally, in consideration of unavoidable structural differences characterizing the two tasks, due to the presence or absence of the explicit value-based outcomes, we also considered the possibility that different decision-making processes might have controlled the choice selections in the two tasks. Therefore, we compared the performance of our Bayesian learner model with those of three more computational models, in which decision processes were determined by heuristics, volatility monitoring or prediction-error learning, respectively.

Previous investigations have revealed a number of brain regions involved in the computations of belief updating and the associated decision confidence (Meyniel and Dehaene, 2017; Morriss et al., 2018; Pouget et al., 2016). For instance, the medial prefrontal cortex (mPFC) (Bang and Fleming, 2018; Matsumoto and Tanaka, 2004; Yoshida and Ishii, 2006) has been shown to be responsible for processing self-monitoring, action evaluation and choice confidence in goal-oriented behavior. At the subcortical level, the anterior hippocampus (aHip) has been implicated in monitoring expectations, predictability and entropy reduction in changing environments (Harrison et al., 2006; Rigoli et al., 2019; Strange et al., 2005), whereas the amygdala (Amg) has been associated with value representation and risk estimation in stochastic environments (Bechara et al., 1999; Dolan, 2007; Jung et al., 2018). Consistent with this literature, we found a neurocircuitry encompassing mPFC, hippocampus and amygdala, which supported belief updating across both tasks and became more active as decision confidence increased. We then used dynamic causal modeling (DCM) (Friston et al., 2003; Stephan et al., 2010; Zeidman et al., 2019) to estimate the directed influence or effective connectivity among the nodes of this neurocircuitry and highlight differences associated with the two tested task environments.

2. Materials and methods

2.1. Participants

We recruited 28 healthy volunteers (16 females), age 24.8 ± 7.0. Participants taking any medication, or with a history of mental disorder or drug abuse were excluded from the study. One subject was excluded from all fMRI analysis involving the Cards task, due to excessive movement. The study was approved by the Institutional Review Board at the University of Texas, Dallas and University of Texas Southwestern Medical Center. Informed written consent was obtained from all subjects and all participants were informed that they could withdraw from the study at any point.

2.2. Experimental design: 3-options continuous choice tasks

In a first task (Fig. 1A), termed *Beads task* (modified from: Huq et al., 1988; Phillips and Edwards, 1966), a new visual stimulus (a red, blue or green *bead*) was presented at each trial, adding a confirming (e.g. a red bead after another red one) or conflicting (e.g. a blue bead following a green one) visual stimulus in a sequence of colored beads. Participants had 2 s to decide from which of three jars displayed on the monitor the bead had been drawn. The jars were illustrated on screen as containing beads of three colors in a ratio of 80%−10%−10%. After each button press, the selected jar would be highlighted with a black rectangle, for the remaining time on the clock allowed for the choice selection, plus 0.5 s. The last five extracted beads were always present on screen (Fig. 1A), and the participants could make the first choice selection starting from the 5th bead extraction. Between trials, a gray square would appear to conceal the new extracted bead for a variable time of 2 to 3.5 s. Participants had access to immediate value-based outcomes during a training session, which would show that each correct guess would result in accumulating 100 points. However, no outcome was provided during the MRI task and the participants were made aware that the accumulated points would be disclosed only at the end of the task.

In the second task, termed *Cards task* (Fig. 1B) (modified from: O'Doherty et al., 2001; Robbins, 1952), the participants had 2 s to select among three cards presented on the screen and characterized by three geometric figures (randomly selected among triangle, square, circle, star and diamond). Each card was assigned a different predominant value among the three possible outcomes of 100, 10 and 0 and after selecting a card, the screen would display a stochastic outcome, highlighted in green (variable duration: 1–2.5 s), with assigned probability of 80%, for the predominant value, and 10% for each of the remaining outcomes. A second message on screen also signaled the amount won on white screen (fixed interval: 0.5 s), followed by a fixation cross (variable interval: 1–2.5 s), which would precede a new trial. The total amount of points accumulated, per block, was always displayed in the lower part of the screen.

For both tasks, participants were compensated with $1 for every 500 points, selecting the points accumulated during one random block per task. Both tasks consisted in 3 blocks of 71 trials each, so the participants were told that the maximum amount of bonus they could gain consisted in about $15 dollars from each task. Three identical sequences -one per block- were used for all subjects, for both the bead colors and the card-value associations.

Task order and block order were counterbalanced across subjects. The choice selections were recorded using a magnet compatible response box, consisting of three buttons arranged horizontally, so to map the three available choices on the screen. The participants were allowed to hold the response box with both hands and generally used both thumbs to perform their choice selections. Finally, the participants were instructed that the computer would randomly change the jar from which it extracted beads, or the card-value associations. The pace of these pseudo-random changes was determined in an interval of 6 ± 1 trials for both tasks (in total across the three blocks: 10 reversals took place after 5 trials, 15 after 6 trials and 8 after 7 trials) and it was independent of the performance of the participants (cf. Fig. 1D,E and Fig. 2). The participants were not instructed explicitly about the number of trials characterizing the interval among reversals, but they familiarized with the task structure by running a training session, consisting of an entire block of trials, outside of the scanner. The distribution of probabilities of the task related events (i.e., beads extracted from a jar or value-based outcomes yielded after a card selection) was identical across tasks (80%−10%−10%). This feature, jointly with the use of three categorical evidence and the relative stability of the pace of reversals, was meant to avoid computations of large vs small prediction errors (cf. McGuire et al., 2014), or sudden changes in the environment volatility (cf. Behrens et al., 2007). By this means the task design was aimed at focusing the participants’ decision process on the probabilistic perceptual evidence, in the Beads task (Adams et al., 2018; Huq et al., 1988; Phillips and Edwards, 1966), and on value-based feedbacks, in the Cards task (O’Doherty et al., 2017; Robbins, 1952; Sutton and Barto, 1998).

2.3. Bayesian learner model

We used two similar computational models to estimate: 1) in the Beads tasks, the trial-by trial subjective probability assigned to each jar, as the source of extraction of the latest bead visible on screen; 2) in the Cards task, the trial-by trial subjective probability that each card would be associated with the highest chance to yield 100 points. We then used the model-estimated probability assigned to the selected jar or card, per trial, to determine subject- and trial-specific choice-confidence (*c*).

To update the estimated subjective probabilities the model relied on a Markov chain: in each trial *t*, the estimates defined at time *t-1* were incrementally updated into the new probabilities, depending on the latest available evidence (*e*) and the subject-specific assumptions about the likelihood (λ) of events in the environment. This update process can be summarized as:

$$P(\mathrm{Jar}_t \mid \mathrm{Jar}_{t-1}, e, \lambda) \tag{1.1}$$

$$P(\mathrm{Card}_{t+1} \mid \mathrm{Card}_t, e, \lambda) \tag{1.2}$$

For instance, if one assumed the environments were characterized by low stochasticity (i.e., vast majority of beads in the jars are of a single color and the vast majority of the value-based feedbacks after a card selection match the assigned value), she would quickly adapt to any change of bead color or card selection outcome, i.e., the behavior would rely largely on the latest presented evidence. Conversely, an environment assumed to be characterized by high stochasticity would result in a slow adaptation, as conflicting evidence would have to accumulate to be sufficient to trigger a change in subjective estimates and choice selections. In Bayesian terms, the probability of an event to occur, assuming a hypothesis is true, defines the likelihood. In the Beads task, the likelihood represents the three events can occur after a bead is extracted from a jar: e.g., P (bead$_{\text{blue}}$, bead$_{\text{red}}$, bead$_{\text{green}}$ | Jar$_{\text{blue}}$). Similarly, in the Cards task, three events can occur, after each choice selection, e.g., P (outcome$_{100}$, outcome$_{10}$, outcome$_{0}$ | Card$_{100}$). Participants were not aware of the exact distribution of colored beads in the jars or outcome stochasticity associated with card selections. Therefore, we considered each subject would rely on their own assumptions about these likelihoods, which would be kept constant through the task, as the task instructions explicitly mentioned the probability distributions would not vary depending on performance or other measures. For simplicity, the model assumed the participants relied on likelihoods characterized by a dominant event for each hypothesis [e.g., P(bead$_{\text{blue}}$ | Jar$_{\text{blue}}$) or P(outcome$_{100}$ | Card$_{100}$)], and two, equally probable, secondary events [e.g., P(bead$_{\text{red}}$ | Jar$_{\text{blue}}$) = P(bead$_{\text{green}}$ | Jar$_{\text{blue}}$) = ((1 - P(bead$_{\text{blue}}$ | Jar$_{\text{blue}}$))/2) or P(outcome$_{10}$ | Card$_{100}$) = P (outcome$_{0}$ | Card$_{100}$) = ((1-P(outcome$_{100}$ | Card$_{100}$))/2)]. Finally, we also assumed that dominant events associated with each of the three hypotheses, per task, were identical [i.e., P(bead$_{\text{blue}}$ | Jar$_{\text{blue}}$) = P(bead$_{\text{red}}$ | Jar$_{\text{red}}$) = P(bead$_{\text{green}}$ | Jar$_{\text{green}}$) or P(outcome$_{100}$ | Card$_{100}$) = P(outcome$_{10}$ | Card$_{10}$) = P(outcome$_{0}$ | Card$_{0}$)]. For instance, given the structure of the tasks, an ideal player would assign ~80% probability to all dominant events and ~10%, to each of the two secondary events. As described, participants assuming a high probability for the dominant event (i.e., close to deterministic environments) would quickly update their beliefs, whereas participants assuming equally distributed events would show a slow update of prior beliefs. This update process of prior probabilities into posterior probabilities was carried out following a standard Bayesian rule (see step by step examples of computations in the supplementary materials):

$$P\left(Jar_j\right)_{\mathrm{t}} \propto P\left(bead_{\mathrm{t}} \mid Jar_j\right)P\left(Jar_j\right)_{\mathrm{t}-1} \tag{2.1}$$

$$P\left(Card100_j\right)_{\mathrm{t}+1} \propto P\left(outcome_{\mathrm{t}} \mid Card100\right)P\left(Card100_j\right)_{\mathrm{t}} \tag{2.2}$$

These computations were used to estimate the trial-by-trial subjective probabilities assigned to the three jars (j), as the source of the latest extracted bead, or to the three cards (j), as the most likely to yield 100 points. Finally, these three dimensional probability estimates were transformed -after the calculation of the error required for parameter regression- to avoid any selection to reach ~0% probability, which would have hindered future incremental updates, as follows: $\frac{\max\left[.05,\ Jar_{j,t}\right]}{\sum \max\left[.05,\ Jar_{j,t}\right]}$ and $\frac{\max\left[.05,\ Card100_{j,t+1}\right]}{\sum \max\left[.05,\ Card100_{j,t+1}\right]}$.

We used a Monte Carlo method (i.e., random search within the space of parameters) to estimate the value of the dominant probability in the likelihood, per each task, that would better match real choice selections expressed by each of the 28 participants. For the regression, we coupled the trial-by-trial model-estimated distribution of probabilities across the three available choices with the actual choice selection of each participant. This value (c) was used to generate an error score per trial, which was computed as $|\log(c)|$. The regression method tested 10^3 randomly generated values for the parameter λ, in a $\left[\frac{1}{3}1\right]$ interval, searching for the value that minimized the total error across all sequences, per task (Fig. 1c).

2.4. Alternative computational models

The performance of the Bayesian learner model was compared with three computational models. These were all based on a similar Markovian principle of trial-by-trial update but relied on significantly different computational processes for the simulation of the participants' choice behavior. The comparison among the models was deemed necessary for two reasons: first, to control for the possibility that the participants were not in fact relying on Bayesian inference and confidence estimations to perform their choice selections. Second to control for the possibility that different, non-comparable, computational processes would be responsible for the choice behaviors in the two tasks.

2.5. Heuristic model

In a first alternative model, we considered the possibility that the participants used simple heuristics to guide their actions, as these computations would make complex considerations about choice confidence less relevant. We employed a mechanism that would increment or decrement, at each trial, the probability associated with a jar or a card by a subject-specific constant H:

$$Jar_{j,\,t} = \min\left(1, \max\left(0, Jar_{j,\,t-1} + H\right)\right) \tag{3.1}$$

$$Card100_{j,\,t+1} = \min\left(1, \max\left(0, Card100_{j,\,t} + H\right)\right) \tag{3.2}$$

where H is a vector ($[-\frac{h}{2}, h, -\frac{h}{2}]$), so that in the Beads task the jar of the same color of the latest extracted bead would increase its estimated probability by a value of h (with a maximum value of 1), whereas the remaining two jars would decrease their probability by $-\frac{h}{2}$ (with a minimum value of 0). Similarly, in the Cards task, a 100 point outcome would trigger an increase of probability for the selected card, decreasing the probabilities assigned to the non-selected cards, whereas the opposite process was employed for 10 points and 0 points outcomes. The value of h was regressed for each subject to find the value that would reduce the model-estimated error.

2.6. Volatility model

In a second model, we considered the possibility that the participants were able to monitor the volatility of the task environments to determine when a reversal occurred. This in turn could be used to estimate, on a trial-by-trial basis, which information could be ignored or which might require behavioral adaptation, depending on the estimated probability that a reversal might have just occurred or not. This model employed computations similar to the ones already described for the Bayesian learner model, with the key difference of relying on a dynamic pace of update, rather than a fixed one (cf. Behrens et al., 2007). To this end, we assumed that the subjective likelihoods (λ in Eqs. (1.1) and 1.2) would vary trial-by-trial, thus varying the evidence-to-prior beliefs ratio and generating a dynamic belief update process. The changes in the λ values were determined as a function of the estimated probability assigned to the occurrence of a reversal $P(R_t \mid R_{t-1},\ V)$, where the value of V was fixed per subject and represented a subjective volatility. The task structures did not allow to compute this value on the basis of changes in the stochasticity of the environments (cf. Cavenaghi et al., 2021), or thanks to the presence of large or small errors (cf. McGuire et al., 2014). Therefore, the model assumed that reversal could be monitored in terms of the subjective likelihood that an event at trial t could be found to differ from an event at trial t-1 (see step-by-step update computations in the supplementary materials).

$$P(R)_{\mathrm{t}} \propto P(Bead_{\neg it} \mid Bead_{it-1})P(R)_{\mathrm{t}-1} \tag{4.1}$$

$$P(R)_{\mathrm{t}+1} \propto P(\text{ Oucome }_{\neg \mathrm{it}} \mid \text{Outcome }_{it-1})P(R)_{\mathrm{t}} \tag{4.2}$$

Note two significant differences in these computations when comparing the two tasks. The estimation of the probability of a reversal was updated before a choice selection in the Beads task and after a choice selection in the Cards task. Furthermore, a repeated event (which would mark a decrease in the estimated probability for a reversal to occur, P(R)) in the Beads task simply consisted in the display of two consecutive beads of the same color, any other combination of colored beads determined an increase in the estimated P(R). Conversely, in the Cards task, we considered a combination of action selection and outcomes, so that a repeated event consisted of two consecutive outcomes of 100 points, under the condition that the choice selections had been repeated as well. Any other combination of choice selections and events increased the estimated P(R). The trial-by-trial P(R) was then used (scaled in a $\left[\frac{1}{3}1\right]$ interval) to determine the trial-by-trial likelihood (λ) of the task-related dominant events, which in turn allowed for the computations described for the Bayesian learner model to be adjusted dynamically affecting belief updating. High estimated P(R) resulted in increased weight on the new evidence, whereas low estimated volatility resulted in increased weight on prior beliefs. Finally, as described for the Bayesian learner model, the P(R) was filtered $\frac{\max[.05,\ \mathrm{R}_t]}{\sum \max[.05,\ \mathrm{R}_t]}$.

2.7. Reinforcement learning model

In a third model, we considered the possibility that participants relied on prediction error estimations to guide their actions. Thus, we used a reinforcement learning approach, characterized by the learning rule:

$$\mathrm{r_t} = \mathrm{r_{t-1}} + \alpha \mathrm{V} \tag{5}$$

where r represents the expected reward or outcome, per trial, α is the learning rate, estimated in each subject, and V is the prediction error, computed with a standard Rescorla-Wagner learning rule. In the Beads task, where no explicit reward is provided, we assumed that a bead at trial t of the same color of the chosen jar at trial t-1 would be considered by the participant as an outcome of 100 (i.e. confirming the previous choice was correct).

Conversely, an inconsistency between the present bead color and the previous choice of jar color would be encoded as an outcome of 0. In both tasks, choice selection was then determined using a linear transformation (cf. Behrens et al., 2007):

$$f(r, \gamma) = \max[0, \min[1, (\gamma(r - 0.5) + 0.5)]] \tag{6}$$

The parameter γ, which was also estimated in each subject, represents whether the participant was risk averse (>1) or risk prone (<1) in their choice selections.

Finally, we compared the predictions of the models in terms of BIC score. Limited to the model comparison, to allow for comparable BIC scores, we used a softmax transformation of the trial-by-trial estimated probabilities (τ=0.1 for all models) and relied on the maximum likelihood (controlled by the Matlab function *fmincon*) for the parameter regression, across models. We estimated the likelihood twice, with independent measures for a linear and a logarithmic computation of the trial-by-trial error (cf. supplementary Fig. 1). The Bayesian learner outperformed all the other tested models in both tasks (Table 1), under both conditions of error estimation, and was therefore used for the fMRI and DCM analysis.

2.8. fMRI data acquisition and preprocessing

Functional MRI data were acquired using a Philips 3-Tesla MR scanner at the Advanced Imaging Research Center at University of Texas Southwestern Medical Center. The anatomical scan sequence (multiecho MPRAGE) was carried out with a resolution of 1 mm, Multi Parametric Maps. Functional images (EPI) were acquired with a resolution of 3.4 × 3.4 × 4 mm, repetition time of 2000 milliseconds, echo time of 25 milliseconds, 38 axial slices, flip angle=90°, and a field of view of 240 mm. We used standard Statistical Parametric Mapping algorithms (SPM12, Wellcome Department of Imaging Neuroscience; www.fil.ion.ucl.ac.uk/spm/) for data preprocessing, including motion realignment to the first volume, coregistration to the participant’s anatomical scan, MNI normalization, and spatial smoothing, using an isotropic 8-mm full-width at half-maximum (FWHM) Gaussian kernel.

2.9. Model-based fMRI

We considered choice confidence encompasses several steps in a process, each responsible for part of the probability estimation in a choice selection (Ma and Jazayeri, 2014; Meyniel et al., 2015b), from sensory estimation to motor execution (Orban and Wolpert, 2011; Wolpert and Landy, 2012), or outcome evaluation (Bach et al., 2011; Meyniel et al., 2015a), when at all available. Thus, for the GLM analysis, we used default SPM12 functions to observe BOLD signals associated with the on-sets of choice selections (i.e. button presses). This is consistent with our computational definition of confidence (Meyniel et al., 2015b), which was meant to investigate this phenomenon in association with a choice, i.e. the estimated probability a selected action would deliver a desired result. We convolved a canonical hemodynamic function (HRF), which is a synthetic hemodynamic response function composed of two gamma functions (Friston et al., 1998, 1994) in SPM, with task regressors related to all onset of choice selections. In each task, we used a GLM to identify the relationship between the -parametrically modulated- task events and the hemodynamic response. For this analysis, we used the RTs as trial-by-trial covariates and choice confidence (c) as parametric modulator. The three blocks were concatenated in both tasks. Whole brain activations were determined using a threshold of $P<.005$, uncorrected (Fig. 3A,B).

2.10. Dynamic causal modeling (DCM)

For the DCM estimation, we relied on SPM12 default functions. Our model based GLM revealed three nodes in the network subserving the signal of confidence (aHip, Amg, mPFC; Fig. 3A,B). We extracted fMRI time series from individual ROIs, using their principal eigenvariates: we relied on group-level anatomical maps manually generated for the Amg and the aHip (available at: https://neurovault.org/collections/9720/), and we used spherical ROIs (8 mm radius) for the mPFC. These were centered on task-specific, group-averaged, peaks activity, at the following coordinates: mPFC (bead: [−9, 56, −1]; card: [−9, 56, −1]). We also included the visual cortex as a network input region (cf. Friston et al., 2003; Stephan et al., 2007, 2008), with ROIs centered on task-specific peaks of activity, at the following coordinates: bead: [−18 −94 2] [18 −94 2]; card: [−18 −91 11] [18 −91 11]. The time series extracted in the visual cortex ROIs were derived from a baseline BOLD activity, before the use of covariates and parametric modulators. In the Beads task, we found significant bilateral BOLD response to the signal of confidence, whereas the response recorded in the Cards task was limited to the ROIs in the left hemisphere. Thus, we focused the DCM analysis on the left hemisphere only.

For the network architectures, we restricted the DCM analysis to those architectures that would better inform about changes in directional relationships among the active ROIs. This was not meant to try to exhaust all possible model structures, but rather to aim at a good balance between accuracy and complexity, thus affording a sufficient generalizability (Pitt and Myung, 2002; Stephan et al., 2010). Therefore, we compared neural architectures that fulfilled two criteria: 1) they presented different targets for the modulatory signals; and 2) they were computationally comparable in terms of the number of free parameters (i.e. comparable number of node-to-node and modulatory connections (cf. Yu et al., 2020). These two requirements led to develop eight models (Fig. 4), characterized by variations in only one of the three key matrices that define network architectures in a DCM analysis (Friston et al., 2003; Penny et al., 2004; Stephan et al., 2010). In particular, we used a fixed, fully connected, A-matrix across all models, so to define a common baseline network of connections, where the nodes could propagate information among one another. We also used a fixed C-matrix across all models, thus defining a constant target for the “driving input”: in this case the presence of the visual stimuli would directly affect the activity of the node representing the visual cortex (e.g. see: Gu et al., 2015a, 2015b). Finally, we generated eight B-matrices, defining eight configurations of targets for the modulatory input (the trial-by-trial signal of confidence, c), so allowing context- and time-dependent variations of the effective connectivity in the targeted connections of the baseline A-matrix. We assumed that these modulatory inputs would affect self-connectivity for all four nodes and the connectivity between visual cortex and the main three ROIs, bidirectionally, for all models. However, the modulatory signal would target the connectivity between the main three ROIs in one direction only. Thus, the eight models did not differ in the overall number of targets of modulatory inputs (10 fixed and 3 variable), but in the combination of selected targets (i.e., 2 possible directions for 3 pairs of nodes, as illustrated in Fig. 4, where the targets of the modulatory signal are highlighted in red).

We used a Bayesian family-wise random effect approach to estimate the directed connectivity between each pair of nodes within the two network triplets, grouping the eight models each time into two families of four models, each family characterized by one common modulated directed connectivity. This method allowed comparing, for instance, the four models with an aHip-to-mPFc modulated connectivity (with varying connectivity between aHip and Amg and between Amg and mPFC, Fig. 4) vs the four models characterized by an mPFC-to-aHip modulated connectivity (and varying connectivity for the remaining couples of nodes). The method allowed to assess the likelihood a modulatory signal would affect the information flow in a specific direction, within a single pair of nodes, independent of the presence of a single winning model that would have emerged in a single model comparison.

2.11. Data availability

Single subject behavior, datacode of the 4 models described, GLM results and associated DCM estimations are fully available in the form of a G-node repository [https://gin.g-node.org/G-Node – link provided upon acceptance].

3. Results

Both tasks posed a significant challenge to the participants, due to the fast pace of the reversals, combined with the presence of three options for the choice selections. The choice selections expressed by the participants indicated they were able to track the correct extraction jar in the Beads task with high accuracy (74.8%±7.4), whereas the Cards task resulted in a lower performance, albeit significantly above chance level, in the attempt to track card-value association reversals (accuracy: 55.36%±7.6). This difference is likely due to the need to explore the deck of cards after each reversal, which required a few probing trials. Indeed, if this analysis of behavioral accuracy is limited to the three trials before each reversal, the gap between the two tasks is significantly reduced (79.5%±13.7 and 71.67%±11.5, for the bead and Cards task, respectively), indicating the participants eventually adapted to the continuous changes in the environments with high accuracy, across tasks. Differences between tasks are also apparent when comparing the behaviors in terms of the probability of changing choice selections in the trials following each reversal (Fig. 2). In both tasks the peak of changes in choice selections takes place during the initial trials after a reversal, with a delay for the Cards task, due to task structure (cf. Fig. 2A,B). In the Beads task (Fig. 2C,D), a comparison between trials characterized by either confirming or conflicting beads (respectively, a bead color at trial *t* of the same or different color, compared with the bead at trial *t-1*) revealed higher probability of a change of selections in association with conflicting beads, in comparison with confirming ones, limited to the fifth, sixth and seventh trial after a reversal (Fig. 2C,D). In the Cards task (Fig. 2E,F), high outcomes (i.e., 100 points) were rarely (<5%, on average) followed by a change in selection, irrespective of the trial position after the reversal. Similarly, low outcomes (i.e., 10 or 0 points) were associated with a high probability (>65%) of triggering a change in choice selections across all trials, despite the significant decrease in the total number of low outcomes, as the participants usually identified the optimal selections a few trials after the reversals.

These behavioral differences suggested the presence or absence of explicit outcomes significantly affected the choice behavior of the participants and they indicated that different computational processes might have been guiding the behaviors in the two tasks. Therefore, we compared the Bayesian learner model with three alternative computational models to control for the possibility that choice selections in either of the two tasks could be better explained using simple heuristics, a method to monitor the volatility of the environment or prediction-error based learning. Model comparison and parameter recovery (Table 1 and supplementary Table 1 and 2) confirmed the presence of qualitative differences in the participants' behaviors across tasks, as similarities emerged between the Bayesian learner and either the Heuristic models or the Reinforcement learning model, in the Beads and Cards task, respectively. Importantly, the Bayesian learner model significantly outperformed any other tested model in matching the participants' behavior, in both tasks, thus allowing to use this model for a comparison across tasks. In particular, the Bayesian learner model represented subject-specific behavioral differences in terms of variations of the parameter controlling the estimated likelihoods. Within-subject comparison revealed the Beads task was characterized by significantly lower λ values (i.e., slower pace of belief update) in comparison with the Cards task (Beads task: λ =0.869±.065; Cards task: λ =0.92±.074; d(27)= 3.35, p=.0024; Fig. 1C). These values resulted in optimized Bayesian inferences that provided a mean behavioral prediction accuracy of 83.5%±6.08% for the Beads task and 72.23%±7.41% for the Cards task (chance≈33%).

A correlation analysis revealed that the estimated confidence (c) was negatively correlated with the reaction times (RTs) in the Beads task for 27 participants out of 28, whereas in the Cards task only five participants showed any significant correlation (cf. Fig. 1D,E; see supplementary Table 3). Next, we included RTs values as covariates and we examined the neural activations associated with the estimated confidence (c) as parametric modulators in separate GLMs for each task (Fig. 1D,E). Consistent with previous literature (Meyniel and Dehaene, 2017; Morriss et al., 2018; Payzan-LeNestour et al., 2013; Pouget et al., 2016), we found that confidence was encoded in the mPFC, aHip and Amg, bilaterally, in the Beads task, and limited to the left hemisphere, in the Cards task (Fig. 3, visible at cluster size>50, P<.005, uncorrected, cf. Supplementary Fig. 2A). A within subject comparison of β values extracted from these ROIs across tasks confirmed the strong similarities of BOLD neural responses, irrespective of the presence of explicit value-based outcomes, as no significant difference was found across tasks. Consistent responses across tasks were also found as decreased activations in association with the signal of confidence (or, symmetrically, as increased activations in association with the complementary probability of 1-c), highlighting BOLD activity in the anterior insular, dorsal anterior cingulate and dorsolateral prefrontal cortex (cf. Supplementary Fig. 3A,B). Finally, a whole brain contrast between the two tasks revealed increased activity (peak: [27 47 20]) in the right lateral frontopolar cortex (Koechlin and Hyafil, 2007; Mansouri et al., 2017) as well as in the putamen, bilaterally (peaks [24−4 5] and [−30 2 8]), when comparing Bead vs Cards task BOLD activity (visible at cluster size>50, *P*<.005, uncorrected, Supplementary Fig. 2B), but this result would not survive FWE correction.

Next, we used DCM to investigate the directional dependencies among confidence-encoding regions. Family-wise model comparison was used to test pair-wise directed connectivity among Amg, aHip and mPFC, in the left hemisphere across tasks, grouping the eight tested models (Fig. 4) into two competing families of four models, depending on the tested connectivity. Limited to the tested model architectures, this analysis revealed that the modulated effective connectivity associated with the signal of confidence were found to change depending on value-based outcome availability. In the Beads task, when immediate value-based outcome was absent, confidence primarily modulated the connections from the aHip to other regions (exceedance probability, left hemisphere: aHip-to-mPFC: 93% and aHip-to-Amg: 90%; Fig. 5A; cf. converging results across hemispheres, supplementary Fig. 4), with no clear directionality for the remaining pair-wise analysis (exceedance probability, left hemisphere: mPFC-to-Amg: 59%). In contrast, in the Cards task, which provided immediate value-based feedback to each choice, the signal of confidence primarily modulated mPFC-to-aHip, mPFC-to-Amg, and Amg-to-aHip connectivity (exceedance probability, left hemisphere: 99% and 95%, and 89%, respectively; Fig. 5B; cf. model exceedance probabilities for both tasks, supplementary Fig. 5). The diverging results did not allow to run t-test comparisons for the weight of modulatory connectivity estimated in the two tasks (Stephan et al., 2010).

4. Discussion

Humans live in constantly changing environments that are often opaque, as explicit outcomes following a choice behavior (e.g., hedonic or value-based) are not always available. To compare belief updating in the presence or absence of immediate explicit outcomes, we developed a multi-option probabilistic perceptual task, and a multi-option non-stationary armed bandit task. Then, we estimated belief updating in a healthy control population ($N = 28$) relying on a Bayesian learner model. The efficacy of these computations in replicating the target human behavior was compared against three alternative computational constructs, to test whether choice selections could be better explained by heuristic decisions, volatility estimates, or prediction-error based learning. This comparison indicated the processes based on the Bayesian inference provided a more accurate explanation of the choice selections expressed in both tasks, outperforming alternative explanations, and allowing for a comparison of neural activity and network dynamics. Interestingly, model comparison also revealed similarities for the Heuristics model and the Bayesian learner model, limited to the Beads task, and for the Reinforcement learning model and Bayesian learner model, limited to the Cards task (cf. Table 1 and supplementary Table 1 and 2). This result suggests that: first, the absence of an immediate feedback favored simplified decision-making processes over prediction-error based ones. However, the non-linear belief updating granted by the Bayesian inference computations provided an advantage in replicating the human behavior, in comparison with linear updates (i.e., heuristics), as the same information in a probabilistic perceptual task can have a varying effect (cf. Fig. 2), depending on the context. Second, the presence of a value-based feedback after each choice selection favored value-based decision-making processes relying on prediction error computations. However, the use of categorical outcomes in a multi-armed bandit task favored again the Bayesian inference, suggesting the outcomes were computed as discrete evidence in favor or against a hypothesis, rather than at their face values.

Our main findings revealed two important aspects of belief updating in uncertain environments with or without immediate outcomes. First, we identified a network that responded to the signal of decision confidence (amygdala, anterior hippocampus, mPFC), regardless of the presence of immediate value-based outcomes. Second, within the limits offered by the network architectures that were tested, DCM analysis suggested that the network dynamics changed as a function of feedback availability. Taken together, these findings revealed important changes across different Marrian levels of analysis (Marr and Poggio, 1976), as the presence of immediate value-based outcomes impacted the neural implementation of belief updating in its confidence-building component, despite the identical computational mechanisms across conditions of feedback access.

Existing literature on decision-making primarily focuses on choices made in environments with immediate explicit feedback, usually value-based outcomes, as it is usually assumed that previously experienced outcomes following chosen actions are needed to generate subjective values and to guide future choices (Berridge and Kringelbach, 2015; O'Doherty et al., 2017; Rangel et al., 2008). This approach has been highly successful in accounting for different forms of conditioning, habitual and goal directed behavior and in uncovering their underlying neural substrates (Balleine et al., 2007; Balleine and O'Doherty, 2010; Dezfouli et al., 2014). The algorithmic formalisation offered by the reinforcement-learning framework (Sutton and Barto, 1998) further expanded the domain of decision-making investigations in accessible environments, indicating reward prediction-error are encoded by dopamine signals (Schultz, 2002; Schultz et al., 1997), and driving model-based analysis of neural activity (Daw et al., 2011, 2005; Dolan and Dayan, 2013; Lee et al., 2014). Nevertheless, many real-life decisions are made in the absence of immediate, value-based outcomes, where agents need to form beliefs based on other sources of information. Partially addressing this issue, previous studies on perceptual decision making have explored how people make choices based only on sensory evidence, in the absence of outcomes (Hanks and Summerfield, 2017; Heekeren et al., 2008). These studies focus on attention processes and perceptual uncertainty, as sensory inputs are characterized by ambiguity or noise, and have highlighted the roles played by hippocampus and mPFC in assessing sensory predictability and the subsequent choice confidence (Bang and Fleming, 2018; Harrison et al., 2006; Rahnev et al., 2016; Strange et al., 2005). Instead, here we used tasks deprived of sensory uncertainty, aiming at investigating the neural dynamics responsible for the update of action-outcome contingencies in the presence and absence of immediate outcomes. Both our two multi-option tasks, with and without immediate value-based feedback, were characterized by simple sets of rules, categorical and discrete evidence (i.e. the three feedback values or the three bead colors), and easy-to-compute distributions of probabilities, reducing stimulus uncertainty as well as second-order uncertainty (Bang and Fleming, 2018; Fleming and Daw, 2017).

DCM analysis indicates that the directional influence among brain regions encoding decision confidence changed across environments. In the presence of immediate outcomes, mPFC and amygdala drove other regions during confidence encoding; but in the absence of these outcomes, the hippocampus showed directional influence towards mPFC and amygdala. Confidence estimations in the Beads task (no outcome) relied on each colored bead in a sequence as discrete evidence, where a sequence of beads of the same color signaled that the environment was likely going through a stable phase, therefore allowing decision confidence to increase. We speculate that the hippocampus was engaged first, so to monitor the predictability of visual stimuli and signal the reduction in entropy of the task environment (Harrison et al., 2006; Rigoli et al., 2019; Strange et al., 2005), as belief-confirming evidence was accumulating. Subsequently, the mPFC and amygdala received this information from the hippocampus (Gluth et al., 2015) and were engaged to increase the estimated confidence (Bang and Fleming, 2018; Matsumoto and Tanaka, 2004; Yoshida and Ishii, 2006) that current choice selections would yield in the future a currently inaccessible desired outcome (Bechara et al., 1999; Dolan, 2007; Jung et al., 2018). Differently, in the Cards task, where stochastic numeric outcomes were immediately available, evidence accumulation was based on expected values. Thus, we speculate that the mPFC (De Martino et al., 2013; Koechlin and Hyafil, 2007) and amygdala (Bechara et al., 1999; Dolan, 2007; Jung et al., 2018) became the driving force in calculating these value-based signals and assigning them to the available choices. This information was then passed to the hippocampus, to monitor the stability or entropy of the environment.

It is important to highlight a few limitations associated with the interpretation of the behavior recorded in the two tasks. Despite the described core of similarities, the two tasks differ in a few key features: first, the chronicle of the latest 5 outcomes is only externalized in the Beads task; second, the Beads task is observation-based, whereas the Cards task is action-based; third, the Cards task presents an increased exploration cost, in comparison with the Beads task; finally, it can be argued that different computational models might reveal that the participants relied on different cognitive processes to guide choice selections in two tasks. Concerning the first three points, in both tasks the update of choice-confidence and uncertainty are based on the observation of the latest available data, therefore reducing the impact of the chronicle externalization. In the Cards task these observations are limited to the choice selections performed in the previous trials, but they either confirm or conflict with existing beliefs by the same quantity in both tasks, due to the organization of evidence into three distinct categories. After a reversal, the Cards task forces the participants to explore the deck to find a new optimal choice, and this behavior was captured by the Bayesian learner model in terms of estimated λ values, and associated differences in Bayesian inference and belief update. We propose that these differences, while important, have a limited impact on the interpretation of our findings, as suggested by the fact that a model based only on Bayesian inferences for belief updates was able to replicate the target behavior at high level of accuracy, across tasks. Finally, the model comparison indicates that the Bayesian inference perspective provides the most effective tool, among those tested, to capture key features of the behavior across tasks. Our comparison is of course limited to the few tested models, but these cover the computational approaches that are most commonly used in association with the choice selections performed in similar environments. Namely, Bayesian inference approach, in a probabilistic perceptual task (Adams et al., 2018; Huq et al., 1988; Phillips and Edwards, 1966) or a reinforcement learning approach, in presence of non-stationary environments characterized by stochastic outcomes (O'Doherty et al., 2017; Robbins, 1952; Sutton and Barto, 1998). All considered, we suggest that these limitations call for further investigations, but they do not hinder the interpretation of the key findings described in this study.

In conclusion, our findings shed a new light on pervasive computational and neural mechanisms underlying belief formation and update. These results represent also an important step to inform future investigations into the breakdown of belief update processes, such as those observed in addiction (Gowin et al., 2013; Ognibene et al., 2019; Verdejo-Garcia et al., 2018), mood disorders (Bishop and Gagne, 2018; Huys et al., 2015), as well as across several psychiatric disorders (Hoven et al., 2019).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Prof. Karl Friston for his comments and kind suggestions in shaping this manuscript. We thank Jennifer Jung for her help in setting up the code for both tasks (Psychtoolbox library in Matlab), AnnCathrin V. Guertler, Chandana C. Tatineni, and Ju-Chi Yu, for their help in collecting the data. The authors also acknowledge the computational resources and staff expertiseprovided by Scientific Computing at the Icahn School of Medicine at Mount Sinai. This research was supported by internal funding from the University of Texas at Dallas, both authors' previous institution. VGF is funded by the Mental Illness Research, Education, and Clinical Center (MIRECC VISN 2) at the James J. Peter Veterans Affairs Medical Center, Bronx, NY. XG is supported by National Institute on Drug Abuse (R01DA043695, R21DA049243) https://www.drugabuse.gov/. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors have declared that no competing interests exist.

Data availability

Single subject behavior, datacode of the 4 models described, GLM results and associated DCM estimations are available in the form of a G-node repository [https://doi.org/10.12751/g-node.kduj58].

Abbreviations:

aHip	anterior hippocampus
Amg	amygdala
DCM	dynamic causal modelling
mPFC	medial prefrontal cortex

References

Adams RA, Napier G, Roiser JP, Mathys C, Gilleen J, 2018. Attractor-like dynamics in belief updating in schizophrenia. J. Neurosci 38, 9471–9485. [PubMed: 30185463]

Adler WT, Ma WJ, 2018. Comparing Bayesian and non-Bayesian accounts of human confidence reports. PLoS Comput. Biol 14, e1006572. [PubMed: 30422974]

Atiya NAA, Zgonnikov A, O'Hora D, Schoemann M, Scherbaum S, Wong-Lin K, 2020. Changes-of-mind in the absence of new post-decision evidence. PLoS Comput. Biol 16, e1007149. [PubMed: 32012147]

Bach DR, Hulme O, Penny WD, Dolan RJ, 2011. The known unknowns: neural representation of second-order uncertainty, and ambiguity. J. Neurosci 31, 4811–4820. [PubMed: 21451019]

Balleine BW, Delgado MR, Hikosaka O, 2007. The role of the dorsal striatum in reward and decision-making. J. Neurosci 27, 8161–8165. [PubMed: 17670959]

Balleine BW, O'Doherty JP, 2010. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35, 48–69. [PubMed: 19776734]

Bang D, Fleming SM, 2018. Distinct encoding of decision confidence in human medial prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A 115, 6082–6087. [PubMed: 29784814]

Bechara A, Damasio H, Damasio AR, Lee GP, 1999. Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. J. Neurosci 19, 5473–5481. [PubMed: 10377356]

Behrens TE, Woolrich MW, Walton ME, Rushworth MF, 2007. Learning the value of information in an uncertain world. Nat. Neurosci 10, 1214–1221. [PubMed: 17676057]

Berridge KC, Kringelbach ML, 2015. Pleasure systems in the brain. NeuronNeuron 86, 646–664.

Bishop SJ, Gagne C, 2018. Anxiety, Depression, and Decision Making: a Computational Perspective. Annu. Rev. Neurosci 41, 371–388. [PubMed: 29709209]

Cavenaghi E, Sottocornola G, Stella F, Zanker M, 2021. Non Stationary Multi-Armed Bandit: empirical Evaluation of a New Concept Drift-Aware Algorithm. Entropy (Basel) 23.

Churchland AK, Ditterich J, 2012. New advances in understanding decisions among multiple alternatives. Curr. Opin. Neurobiol 22, 920–926. [PubMed: 22554881]

Churchland AK, Kiani R, Shadlen MN, 2008. Decision-making with multiple alternatives. Nat. Neurosci 11, 693–702. [PubMed: 18488024]

Dabney W, Kurth-Nelson Z, Uchida N, Starkweather CK, Hassabis D, Munos R, Botvinick M, 2020. A distributional code for value in dopamine-based reinforcement learning. NatureNature 577, 671–675.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ, 2011. Model-based influences on humans' choices and striatal prediction errors. NeuronNeuron 69, 1204–1215.

Daw ND, Niv Y, Dayan P, 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci 8, 1704–1711. [PubMed: 16286932]

De Martino B, Fleming SM, Garrett N, Dolan RJ, 2013. Confidence in value-based choice. Nat. Neurosci 16, 105–110. [PubMed: 23222911]

Dezfouli A, Lingawi NW, Balleine BW, 2014. Habits as action sequences: hierarchical action control and changes in outcome value. Philos. Trans. R. Soc. Lond. B Biol. Sci 369.

Dolan RJ, 2007. The human amygdala and orbital prefrontal cortex in behavioural regulation. Philos. Trans. R. Soc. Lond. B Biol. Sci 362, 787–799. [PubMed: 17403643]

Dolan RJ, Dayan P, 2013. Goals and habits in the brain. NeuronNeuron 80, 312–325.

Fiore VG, Guertler AV, Yu JC, Tatineni CC, Gu X, 2021. A change of mind: globus pallidus activity and effective connectivity during changes in choice selections. Eur. J. Neurosci

Fleming SM, Daw ND, 2017. Self-evaluation of decision-making: a general Bayesian framework for metacognitive computation. Psychol. Rev 124, 91–114. [PubMed: 28004960]

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R, 1998. Event-related fMRI: characterizing differential responses. NeuroimageNeuroimage 7, 30–40.

Friston KJ, Harrison L, Penny W, 2003. Dynamic causal modelling. NeuroimageNeuroimage 19, 1273–1302.

Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RS, 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp 2, 189–210.

Glimcher PW, 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. U. S. A 108 (Suppl 3), 15647–15654. [PubMed: 21389268]

Gluth S, Sommer T, Rieskamp J, Buchel C, 2015. Effective connectivity between hippocampus and ventromedial prefrontal cortex controls preferential choices from memory. NeuronNeuron 86, 1078–1090.

Gowin JL, Mackey S, Paulus MP, 2013. Altered risk-related processing in substance users: imbalance of pain and gain. Drug Alcohol Depend 132, 13–21. [PubMed: 23623507]

Gu X, Eilam-Stock T, Zhou T, Anagnostou E, Kolevzon A, Soorya L, Hof PR, Friston KJ, Fan J, 2015a. Autonomic and brain responses associated with empathy deficits in autism spectrum disorder. Hum. Brain Mapp 36, 3323–3338. [PubMed: 25995134]

Gu X, Wang X, Hula A, Wang S, Xu S, Lohrenz TM, Knight RT, Gao Z, Dayan P, Montague PR, 2015b. Necessary, yet dissociable contributions of the insular and ventromedial prefrontal cortices to norm adaptation: computational and lesion evidence in humans. J. Neurosci 35, 467–473. [PubMed: 25589742]

Hanks TD, Summerfield C, 2017. Perceptual decision making in rodents, monkeys, and humans. NeuronNeuron 93, 15–31.

Harrison LM, Duggins A, Friston KJ, 2006. Encoding uncertainty in the hippocampus. Neural Netw 19, 535–546. [PubMed: 16527453]

Heekeren HR, Marrett S, Ungerleider LG, 2008. The neural systems that mediate human perceptual decision making. Nat. Rev. Neurosci 9, 467–479. [PubMed: 18464792]

Hoven M, Lebreton M, Engelmann JB, Denys D, Luigjes J, van Holst RJ, 2019. Abnormalities of confidence in psychiatry: an overview and future perspectives. Transl. Psychiatry 9, 268. [PubMed: 31636252]

Huq SF, Garety PA, Hemsley DR, 1988. Probabilistic judgements in deluded and non-deluded subjects. Q. J. Exp. Psychol. A 40, 801–812. [PubMed: 3212213]

Huys QJ, Daw ND, Dayan P, 2015. Depression: a decision-theoretic analysis. Annu. Rev. Neurosci 38, 1–23. [PubMed: 25705929]

Jung WH, Lee S, Lerman C, Kable JW, 2018. Amygdala functional and structural connectivity predicts individual risk tolerance. NeuronNeuron 98, 394–404 e394.

Koechlin E, Hyafil A, 2007. Anterior prefrontal function and the limits of human decision-making. ScienceScience 318, 594–598.

Kording KP, Wolpert DM, 2006. Bayesian decision theory in sensorimotor control. Trends Cogn. Sci 10, 319–326. [PubMed: 16807063]

Krug A, Cabanis M, Pyka M, Pauly K, Walter H, Landsberg M, Shah NJ, Winterer G, Wolwer W, Musso F, Muller BW, Wiedemann G, Herrlich J, Schnell K, Vogeley K, Schilbach L, Langohr K, Rapp A, Klingberg S, Kircher T, 2014. Investigation of decision-making under uncertainty in healthy subjects: a multi-centric fMRI study. Behav. Brain Res 261, 89–96. [PubMed: 24355752]

Lee SW, Shimojo S, O'Doherty JP, 2014. Neural computations underlying arbitration between model-based and model-free learning. NeuronNeuron 81, 687–699.

Ma WJ, Beck JM, Latham PE, Pouget A, 2006. Bayesian inference with probabilistic population codes. Nat. Neurosci 9, 1432–1438. [PubMed: 17057707]

Ma WJ, Jazayeri M, 2014. Neural coding of uncertainty and probability. Annu. Rev. Neurosci 37, 205–220. [PubMed: 25032495]

Mansouri FA, Koechlin E, Rosa MGP, Buckley MJ, 2017. Managing competing goals - a key role for the frontopolar cortex. Nat. Rev. Neurosci 18, 645–657. [PubMed: 28951610]

Marr D, Poggio T, 1976. From Understanding Computation to Understanding Neural Circuitry Massachusetts Institute of Technology, Artificial Intelligence Laboratory, Cambridge.

Matsumoto K, Tanaka K, 2004. The role of the medial prefrontal cortex in achieving goals. Curr. Opin. Neurobiol 14, 178–185. [PubMed: 15082322]

McGuire JT, Nassar MR, Gold JI, Kable JW, 2014. Functionally dissociable influences on learning rate in a dynamic environment. NeuronNeuron 84, 870–881.

Meyniel F, Dehaene S, 2017. Brain networks for confidence weighting and hierarchical inference during probabilistic learning. Proc. Natl. Acad. Sci. U. S. A 114, E3859–E3868. [PubMed: 28439014]

Meyniel F, Schlunegger D, Dehaene S, 2015a. The sense of confidence during probabilistic learning: a normative account. PLoS Comput. Biol 11, e1004305. [PubMed: 26076466]

Meyniel F, Sigman M, Mainen ZF, 2015b. Confidence as Bayesian probability: from neural origins to behavior. NeuronNeuron 88, 78–92.

Morriss J, Gell M, van Reekum CM, 2018. The uncertain brain: a co-ordinate based meta-analysis of the neural signatures supporting uncertainty during different contexts. Neurosci. Biobehav. Rev

Nassar MR, Bruckner R, Gold JI, Li SC, Heekeren HR, Eppinger B, 2016. Age differences in learning emerge from an insufficient representation of uncertainty in older adults. Nat. Commun 7, 11609. [PubMed: 27282467]

Nassar MR, McGuire JT, Ritz H, Kable JW, 2019. Dissociable forms of uncertainty–driven representational change across the human brain. J. Neurosci 39, 1688–1698. [PubMed: 30523066]

O’Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C, 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat. Neurosci 4, 95–102. [PubMed: 11135651]

O’Doherty JP, Cockburn J, Pauli WM, 2017. Learning, reward, and decision making. Annu. Rev. Psychol 68, 73–100. [PubMed: 27687119]

Ognibene D, Fiore VG, Gu X, 2019. Addiction beyond pharmacological effects: the role of environment complexity and bounded rationality. Neural Netw 116, 269–278. [PubMed: 31125913]

Orban G, Wolpert DM, 2011. Representations of uncertainty in sensorimotor control. Curr. Opin. Neurobiol 21, 629–635. [PubMed: 21689923]

Payzan-LeNestour E, Bossaerts P, 2011. Risk, unexpected uncertainty, and estimation uncertainty: Bayesian learning in unstable settings. PLoS Comput. Biol 7, e1001048. [PubMed: 21283774]

Payzan-LeNestour E, Dunne S, Bossaerts P, O’Doherty JP, 2013. The neural representation of unexpected uncertainty during value-based decision making. NeuronNeuron 79, 191–201.

Penny WD, Stephan KE, Mechelli A, Friston KJ, 2004. Comparing dynamic causal models. NeuroimageNeuroimage 22, 1157–1172.

Phillips LD, Edwards W, 1966. Conservatism in a simple probability inference task. J. Exp. Psychol 72, 346–354. [PubMed: 5968681]

Pitt M, Myung I, 2002. When a good fit can be bad. Trends Cogn. Sci 6, 421–425. [PubMed: 12413575]

Pouget A, Beck JM, Ma WJ, Latham PE, 2013. Probabilistic brains: knowns and unknowns. Nat. Neurosci 16, 1170–1178. [PubMed: 23955561]

Pouget A, Drugowitsch J, Kepecs A, 2016. Confidence and certainty: distinct probabilistic quantities for different goals. Nat. Neurosci 19, 366–374. [PubMed: 26906503]

Rahnev D, Nee DE, Riddle J, Larson AS, D’Esposito M, 2016. Causal evidence for frontal cortex organization for perceptual decision making. Proc. Natl. Acad. Sci. U. S. A 113, 6059–6064. [PubMed: 27162349]

Rangel A, Camerer C, Montague PR, 2008. A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci 9, 545–556. [PubMed: 18545266]

Rich D, Cazettes F, Wang Y, Pena JL, Fischer BJ, 2015. Neural representation of probabilities for Bayesian inference. J. Comput. Neurosci 38, 315–323. [PubMed: 25561333]

Rigoli F, Michely J, Friston KJ, Dolan RJ, 2019. The role of the hippocampus in weighting expectations during inference under uncertainty. Cortex 115, 1–14. [PubMed: 30738997]

Robbins H, 1952. Some aspects of the sequential design of experiments. Bull. Am. Math. Soc 58, 527–535.

Sanders JI, Hangya B, Kepecs A, 2016. Signatures of a statistical computation in the human sense of confidence. NeuronNeuron 90, 499–506.

Schultz W, 2002. Getting formal with dopamine and reward. NeuronNeuron 36, 241–263.

Schultz W, Dayan P, Montague PR, 1997. A neural substrate of prediction and reward. ScienceScience 275, 1593–1599.

Soltani A, Izquierdo A, 2019. Adaptive learning under expected and unexpected uncertainty. Nat. Rev. Neurosci 20, 635–644. [PubMed: 31147631]

Stephan KE, Harrison LM, Kiebel SJ, David O, Penny WD, Friston KJ, 2007. Dynamic causal models of neural system dynamics:current state and future extensions. J. Biosci 32, 129–144. [PubMed: 17426386]

Stephan KE, Kasper L, Harrison LM, Daunizeau J, den Ouden HE, Breakspear M, Friston KJ, 2008. Nonlinear dynamic causal models for fMRI. NeuroimageNeuroimage 42, 649–662.

Stephan KE, Penny WD, Moran RJ, den Ouden HE, Daunizeau J, Friston KJ, 2010. Ten simple rules for dynamic causal modeling. NeuroimageNeuroimage 49, 3099–3109.

Strange BA, Duggins A, Penny W, Dolan RJ, Friston KJ, 2005. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Netw 18, 225–230. [PubMed: 15896570]

Sutton RS, Barto AG, 1998. Reinforcement Learning: An Introduction MIT Press, Cambridge, MA.

Tajima S, Drugowitsch J, Patel N, Pouget A, 2019. Optimal policy for multi-alternative decisions. Nat. Neurosci

Tobia MJ, Iacovella V, Hasson U, 2012. Multiple sensitivity profiles to diversity and transition structure in non-stationary input. NeuroimageNeuroimage 60, 991–1005.

Verdejo-Garcia A, Chong TT, Stout JC, Yucel M, London ED, 2018. Stages of dysfunctional decision-making in addiction. Pharmacol. Biochem. Behav 164, 99–105. [PubMed: 28216068]

Wolpert DM, Landy MS, 2012. Motor control is decision-making. Curr. Opin. Neurobiol 22, 996–1003. [PubMed: 22647641]

Yoshida W, Ishii S, 2006. Resolution of uncertainty in prefrontal cortex. NeuronNeuron 50, 781–789.

Yu J−C, Fiore VG, Briggs RW, Braud J, Rubia K, Adinoff B, Gu X, 2020. An insula-driven network computes decision uncertainty and promotes abstinence in chronic cocaine users. Eur. J. Neurosci 52, 4923–4936. [PubMed: 33439518]

Zeidman P, Jafarian A, Corbin N, Seghier ML, Razi A, Price CJ, Friston KJ, 2019. A guide to group effective connectivity analysis, part 1: first level analysis with DCM for fMRI. NeuroimageNeuroimage 200, 174–190.

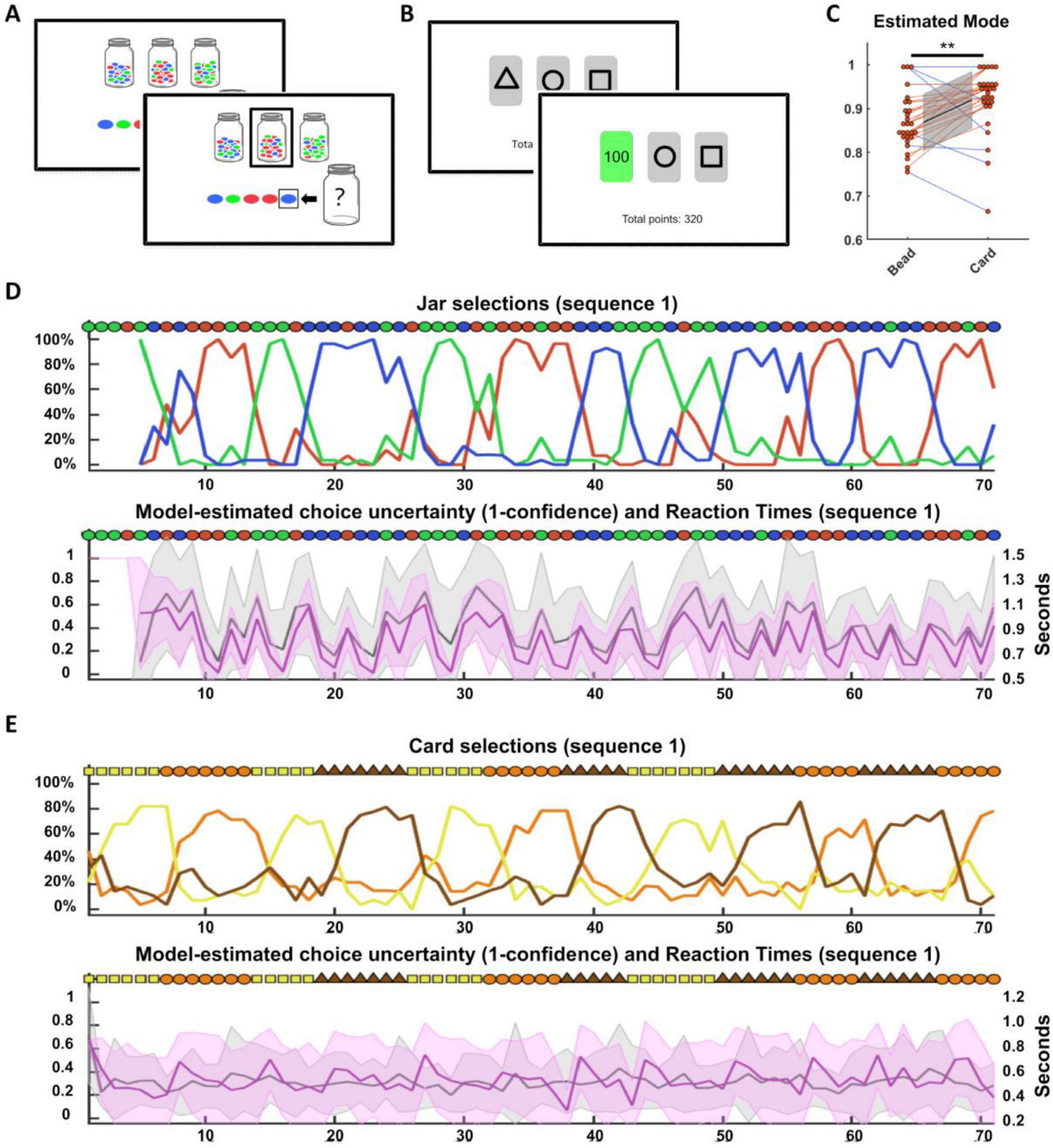

Fig. 1.

Experimental paradigm, choice behavior, and model estimations for trial-by-trial confidence. (A) Beads task. Participants chose from which of three jars, containing predominantly red, blue or green beads (80%−10%−10% ratio), the latest bead was extracted from. The latest five extracted beads were always displayed on screen, and no feedback was provided after each choice selection. (B) Cards task. Participants chose one card among three cards characterized by different geometric figures. Each card yielded 100, 10 or 0 points (immediate value-based outcome) with a probability distribution of 80%−10%−10%. The extraction jar and the card-value associations were changed every 6 ± 1 trials, in three pre-established pseudo-random sequences that were used for all participants, in a counterbalanced order. (C) Subject-specific values of the parameter λ (grouped in intervals of 0.01 in the scatterplot), used in the Bayesian learner models to estimate trial-by-trial subjective choice confidence in the participants ($N = 28$, in both tasks). Different λ values determine different paces of belief update: for instance, the lower the value, the more confirming evidence is required to increase confidence and decrease uncertainty. The color of each extracted new bead and the feedback provided after each card selection was used as discrete new evidence to update the estimated probability distribution or priors of each participant. (D-E) The upper rows of the panels illustrate the trial-by-trial choice selections expressed by the participants (as percentages). The lower rows of the panels illustrate the relation between trial-by-trial reaction times (mean and standard deviation, in gray) and the complementary probability of model-estimated confidence (or uncertainty, mean and standard deviation, in magenta). At the top of both rows an illustration of the sequence of colored bead, extracted in each trial, and an illustration of the sequence of winning cards, per trial (i.e., the geometric symbol characterizing the card associated with the highest reward). These were the actual sequences used for one of the three blocks in the two tasks. **: $p<.005$.

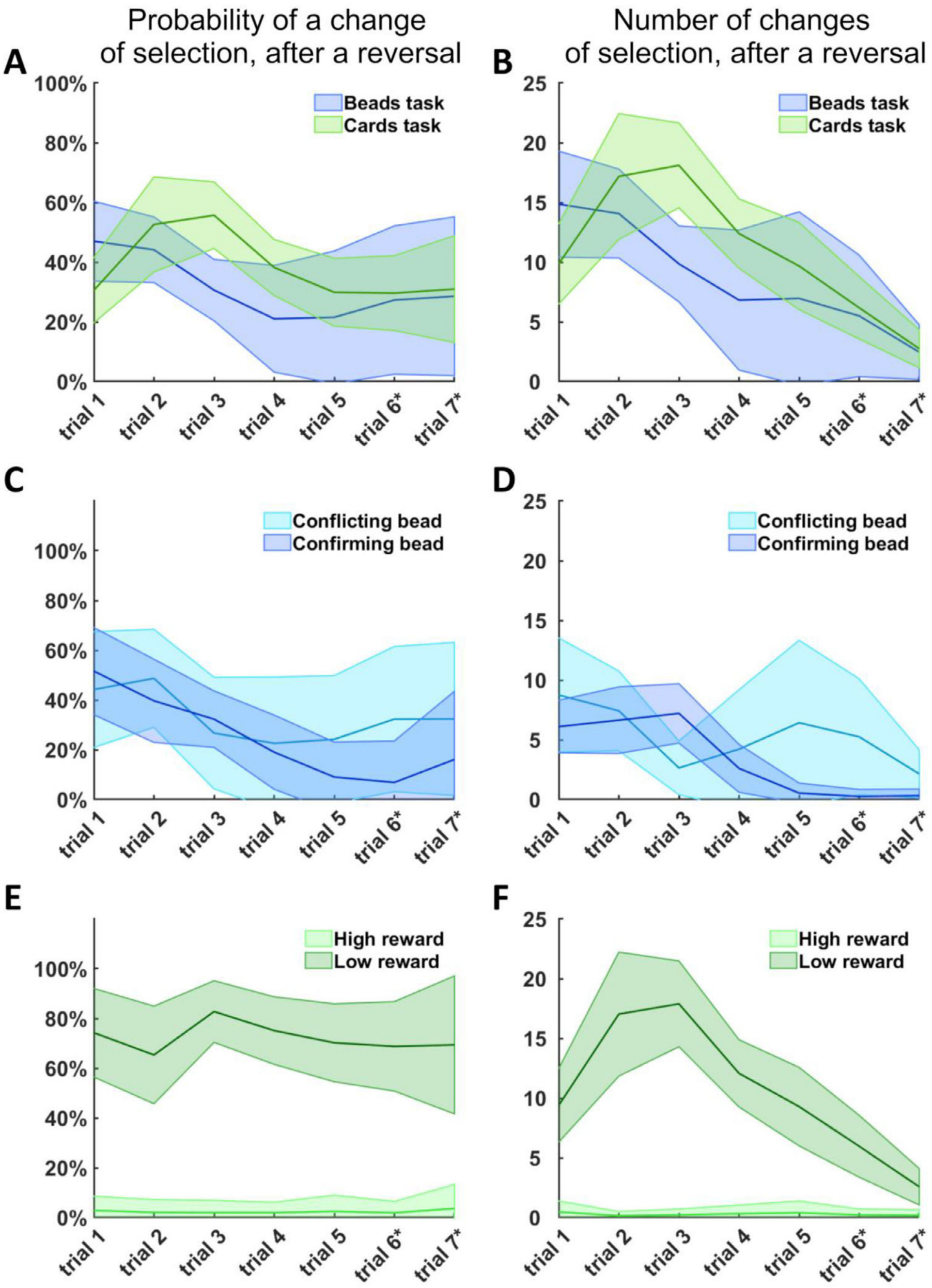

Fig. 2.

Analysis of the Behavior in the trials following a reversal. The error bands represent mean and standard deviation describing the probability of a change of choice selection (A, C, E) and the absolute number of these changes, recorded in the trials after any reversal (B, D, F), across the population of participants (*N* = 28, in both tasks). (A) In the Beads task the peak of changes in choice selections occurs in the initial two trials after a reversal, whereas in the Cards task the peak takes place in the second and third trial after a reversal. This delay is due to the different task structures. In the Beads task choice selections take place after a bead is displayed, so the first evidence that a reversal has occurred can be immediately computed. Conversely, in the Cards task, the first indication that a reversal occurred can only be revealed to the participant after a choice selection, so this information can be used starting from trial 2 after the reversal. (B) Note the decrease in the overall number of changes in choice selections as more evidence is collected after a reversal, consolidating the beliefs. In the Beads task, the highest probabilities (C) of a change in choice selections and the highest numbers of changes (D) were recorded in the initial trials following a reversal, across different types of evidence. Confirming vs conflicting beads were associated with similar probabilities to change a choice selection for the first four trials after a reversal. Starting from the fifth trial, conflicting beads were found to be significantly more likely to trigger a change in policy (trial 5: t(27)=3.27, *p*=.0029; trial 6: t(27)=4.59, *p*<.001; trial 7: t(27)=2.66, *p*=.0128). In the Cards task (E,F), 100 point outcomes (high reward) were followed by a change in choice selection in less than 5% of the events, on average, across all trials following a reversal. Conversely, 10 or 0 point outcomes (low reward) triggered a change in choice selections with high probability (>65%), across trials following a reversal. * Note that, due to the structure of the task, the 6th and 7th trial after a reversal only occur 23 and 8 times, respectively.

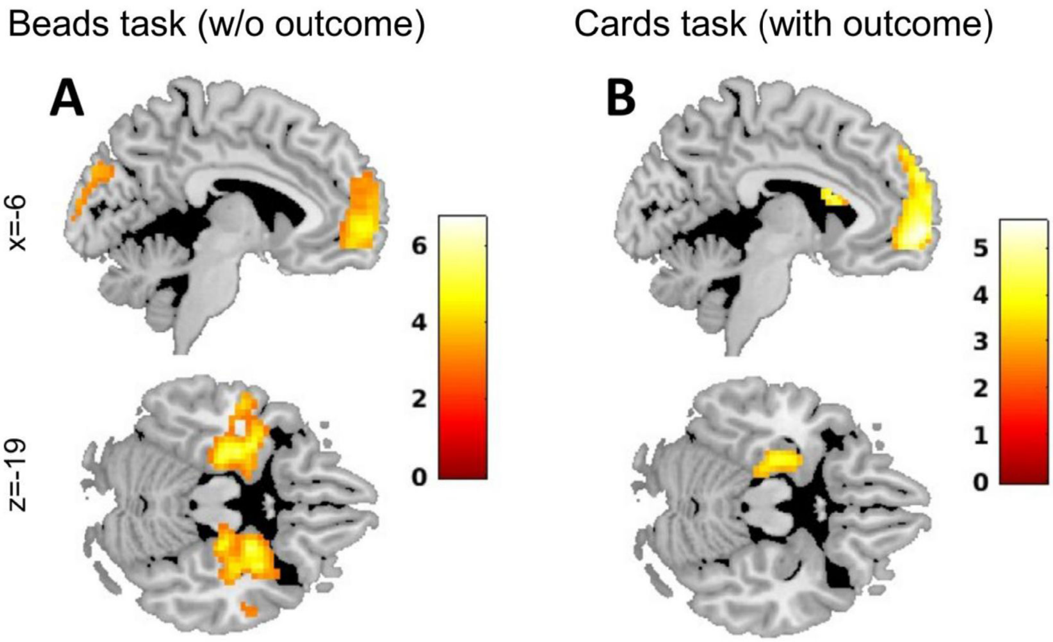

Fig. 3.

Model-based BOLD responses encoding Bayesian confidence. fMRI BOLD response recorded using the model-estimated signal of confidence as parametric modulator and reaction times as covariates, in the Beads task (A; *P*<.005, *k* = 50, uncorrected, *N* = 28), in the Cards task (B; *P*<.005, *k* = 50, uncorrected, *N* = 27). The two panels illustrate the similarities in the neural response across tasks, as they highlight BOLD signal in the medial prefrontal cortex, anterior hippocampus, and amygdala, bilaterally in the Beads task, and in the left hemisphere in the Cards task. To be noted that the BOLD responses of each single task did not resist FWE correction, whereas the joint activity of the two tasks resists P_{FWE}<0.05 correction for all the three regions of interest, limited to the left hemisphere (cf. supplementary figure 2)..

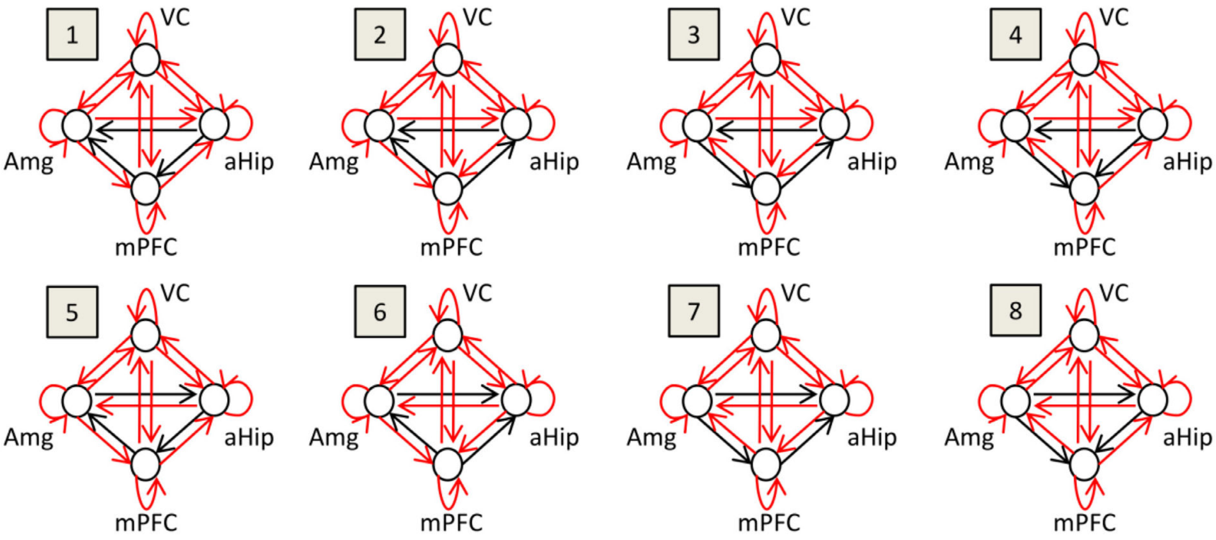

Fig. 4.

Neural network architectures tested in the family-wise DCM. These neural architectures illustrate the eight models used to test the effective connectivity associated with the signal of confidence. The same model structures have been tested in both tasks to ease a comparison of the results. The general architecture of connectivity (*A matrix*) remains unchanged, and it is represented by all the arrows in the network illustrations. Conversely, the targets of the modulatory inputs (*B matrix*), marked only by the red arrows, are different in each model. This differentiation results in the generation of the 8 models, then divided into competing families of 4, to allow family-wise comparisons for each pair of nodes in the ROI triplet of aHip-Amg-mPFC. For instance, to determine whether the directed modulated connectivity from Amg-to-aHip better explains the extracted data, in comparison with the aHip-to-Amg connectivity, the models 1 to 4 are considered in a single family against the models 5 to 8 in a second family. (aHip: anterior hippocampus; Amg: amygdala; mPFC: medial prefrontal cortex; VC: visual cortex).

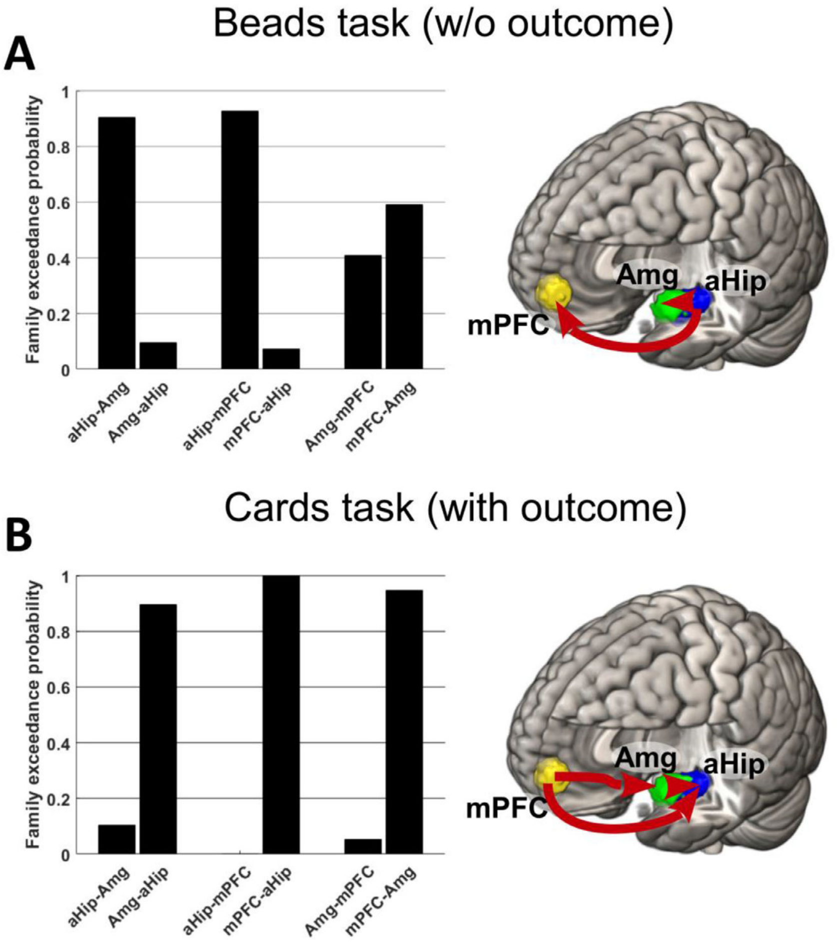

Fig. 5.

DCM results using family-wise comparison. Histograms report the family-wise exceedance probability for each tested pair-wise connection in each task. On the right, arrows connecting the ROIs in the 3d brain rendering illustrate a summary of the results of the pair-wise analysis, highlighting the estimated direction of effective connectivity. Missing connectivity between ROIs represents those pair-wise analyses for which DCM did not provide a conclusive result. In the Beads task, the aHip increases its influence towards Amg and mPFC, as the confidence signal increases (A). Conversely, in the Cards task, mPFC and Amg exert an increasing influence towards the aHip in association with increased signal of confidence (B). aHip: anterior hippocampus; Amg: amygdala; mPFC: medial prefrontal cortex.

Table 1

Model comparison.

Model	Average BIC score Logarithmic error function		Average BIC score Linear error function	
	Beads task	**Cards task**	**Beads task**	**Cards task**
Bayesian Learner	154.49±51.85	257.39± 75.60	70.88±23.83	66.29±14.20
Heuristic	160.43±49.81	410.24± 52.62	100.72±18.10	155.58±22.74
Volatility	250.78±56.28	311.03± 90.83	94.23±19.06	124.84±29.95
Reinforcement Learning	233.84±43.80	277.81± 72.71	141.35±35.02	122.73±27.35

The Bayesian model is characterized by significantly lower BIC scores in comparison with any other tested model, across tasks and across regression method. For the models reporting the closest sets of BIC scores, within-subjects ttest comparisons highlighted significantly lower BIC scores in the Bayesian Learner vs Heuristic model dt(27)= −3.16 *p*=.0038 (Beads task, logarithmic error function) and in the Bayesian Learner vs reinforcement Learning model dt(27)= −3.3 *p*=.0027 (Cards task, logarithmic error function).