

Separate mesocortical and mesolimbic pathways encode effort and reward learning signals

Short title: Separate pathways in multi-attribute learning

Tobias U. Hauser^{a,b,*}, Eran Eldar^{a,b} & Raymond J. Dolan^{a,b}

BIOLOGICAL SCIENCES: Psychological and Cognitive Sciences; Neuroscience

^a Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom

^b Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, United Kingdom

* Corresponding author (t.hauser@ucl.ac.uk)

Correspondence

Tobias U. Hauser
Wellcome Trust Centre for Neuroimaging
University College London
12 Queen Square
London WC1N 3BG
United Kingdom
Phone: +44 / 20 3448 4409
Email: t.hauser@ucl.ac.uk
ORCID: 0000-0002-7997-8137
Twitter: tobiasuhauser

Abstract

Optimal decision-making mandates organisms learn the relevant features of choice options. Likewise, knowing how much effort we should expend can assume paramount importance. A mesolimbic network supports reward learning but it is unclear whether other choice features, such as effort learning, rely on this same network. Using computational fMRI, we show parallel encoding of effort and reward prediction errors (PEs) within distinct brain regions, with effort PEs expressed in dorsomedial prefrontal cortex and reward PEs in ventral striatum respectively. We show a common mesencephalic origin for these signals evident in overlapping, but spatially dissociable, dopaminergic midbrain regions expressing both types of PE. During action anticipation reward and effort expectations were integrated in ventral striatum consistent with a computation of an overall net-benefit of a stimulus. Thus, we show that motivationally relevant stimulus features are learned in parallel dopaminergic pathways, with formation of an integrated utility signal at choice.

Keywords

Effort prediction errors, reward prediction errors, computational fMRI, effort discounting, apathy, negative symptoms, substantia nigra / ventral tegmental area (SN/VTA), dorsomedial prefrontal cortex, ventral striatum

Significance Statement

Learning about multiple features of a choice option is crucial for optimal decision making. How such multi-attribute learning is realized remains unclear. Using functional MRI, we show the brain exploits separate mesolimbic and mesocortical networks to simultaneously learn about reward and effort attributes. We show a double-dissociation, evident in the expression of effort learning signals in dorsomedial prefrontal and reward learning in ventral

striatal areas with this dissociation being spatially mirrored in dopaminergic midbrain. At the time of choice these segregated signals are integrated in the ventral striatum. These findings highlight how the brain parses parallel learning demands.

Introduction

Organisms need to make energy-efficient decisions in order to maximize benefits and minimize costs, a trade-off exemplified in effort expenditure (1–3). A key example is during foraging, where an overestimation of effort can lead to inaction and starvation (4), whereas underestimation of effort can result in persistent failure as exemplified in the myth of Sisyphus (5).

In a naturalistic environment, we often simultaneously learn about success in expending sufficient effort into an action as well as the reward we obtain from this same action. The reward outcomes that signal success and failure of an action are usually clear, though the effort necessary to attain success is often less transparent. Only by repeatedly experiencing success and failure it is possible to acquire an estimate of an optimal level of effort needed to succeed, without unnecessary waste of energy. This type of learning is important in contexts as diverse as foraging, hunting and harvesting (6–8). Hull in his ‘law of less work’ proposed that organisms ‘gradually learn’ how to minimize effort expenditure (9). Surprisingly, we know little regarding the neurocognitive mechanisms that guide this form of simultaneous learning about reward and effort.

A mesolimbic dopamine system encodes a teaching signal tethered to prediction of reward outcomes (10, 11). These reward prediction errors (PEs) arise from dopaminergic neurons in substantia nigra and ventral tegmental area (SN/VTA) and are broadcast to ventral striatum (VS) to mediate reward-related adaptation and learning (12, 13). Dopamine is also thought to provide a motivational signal (14–18), while dopaminergic deficits in rodents impair how effort and reward are arbitrated (1, 4, 19). The dorsomedial prefrontal cortex (dmPFC, spanning pre-supplementary motor area [pre-SMA] and dorsal anterior cingulate cortex [dACC]) is a candidate substrate for effort learning. For example, selective lesioning of this region engender a preference for low effort choices (15, 20–23), while receiving effort

feedback elicit responses in this same region (24, 25). The importance of dopamine to effort learning is also hinted at in disorders with putative aberrant dopamine function, such as schizophrenia (26), where a symptom profile (‘negative symptoms’) often includes a lack of effort expenditure and apathy (27–30).

A dopaminergic involvement in effort arbitration (14, 15, 17, 19, 30) suggests that effort learning might proceed by exploiting the similar mesolimbic mechanisms as in reward learning, and this would predict effort PE signals in SN/VTA and VS. Alternatively, based on a possible role for dorsomedial prefrontal cortex, effort and reward learning signals might be encoded in two segregated (dopaminergic) systems, with reward learning relying on PEs within mesolimbic SN/VTA and VS and effort learning relying on PEs in mesocortical SN/VTA projecting to dmPFC. A final possibility is that during simultaneous learning, the brain might express a unified net benefit signal, integrated over reward and effort, and update this signal via a ‘utility’ PE alone.

To test these predictions we developed a paradigm wherein subjects learnt simultaneous reward and effort contingencies in an ecologically realistic manner, whilst also acquiring human functional magnetic resonance imaging (fMRI). We reveal a double-dissociation within mesolimbic and mesocortical networks in relation to reward and effort learning. These segregated teaching signals, with an origin in spatially dissociable regions of dopaminergic midbrain, were integrated in VS during action preparation consistent with a unitary net benefit signal.

Results

Effort and reward learning

Our behavioral task required 29 male subjects to learn simultaneously about, and adapt to, changing effort demands as well as changing reward magnitudes (Fig. 1A and Supplemental Information). On every trial, subjects saw one of two stimuli, where each stimulus was associated with a specific reward magnitude (1 to 7 points, 50% reward probability across entire task) and a required effort threshold (% of individual maximal force, determined during practice). These parameters were initially unknown to the subject and drifted over time, such that reward and effort magnitudes changed independently. After an effort execution phase the associated reward magnitude of the stimulus was shown together with categorical feedback as to whether the subject had exceeded a necessary effort threshold, where the latter was required to successfully reap the reward. Importantly, subjects were not informed explicitly about the height of the effort threshold, but only received feedback as to the success (or not) of their effort expenditure. On every trial, subjects received information about both effort and reward, and thus learned simultaneously about both reward magnitude and a required effort threshold, through a process of trial-and-error (Fig. 1B).

To assess learning we performed a multiple regression analysis (Fig. 1C) that predicted exerted effort on every trial. A significant effect of previous effort ($t(28)=15.96$, $p<.001$) indicated subjects were not exerting effort randomly but approximated the effort expended with previous experience of the same stimulus, as expected from a learning process. Subjects increased their effort for higher rewards ($t(28)=4.97$, $p<.001$), consistent with a motivation to expend greater energy on high value choices. Lastly, subjects exerted more effort following trials where they failed to exceed an effort threshold ($t(28)=10.75$, $p<.001$), consistent with adaptation of effort to a required threshold. Subsequent analysis showed that subjects not only increased effort expenditure after missing an effort threshold (Fig. 1D, $t(28)=17.08$, $p<.001$),

but also lessened their effort after successfully surpassing an effort threshold ($t(28)=-17.15$, $p<.001$), in accordance with the predictions of Hull's law (9). Thus, these analyses combined reveal subjects were able to simultaneously learn about both rewards and effort requirements.

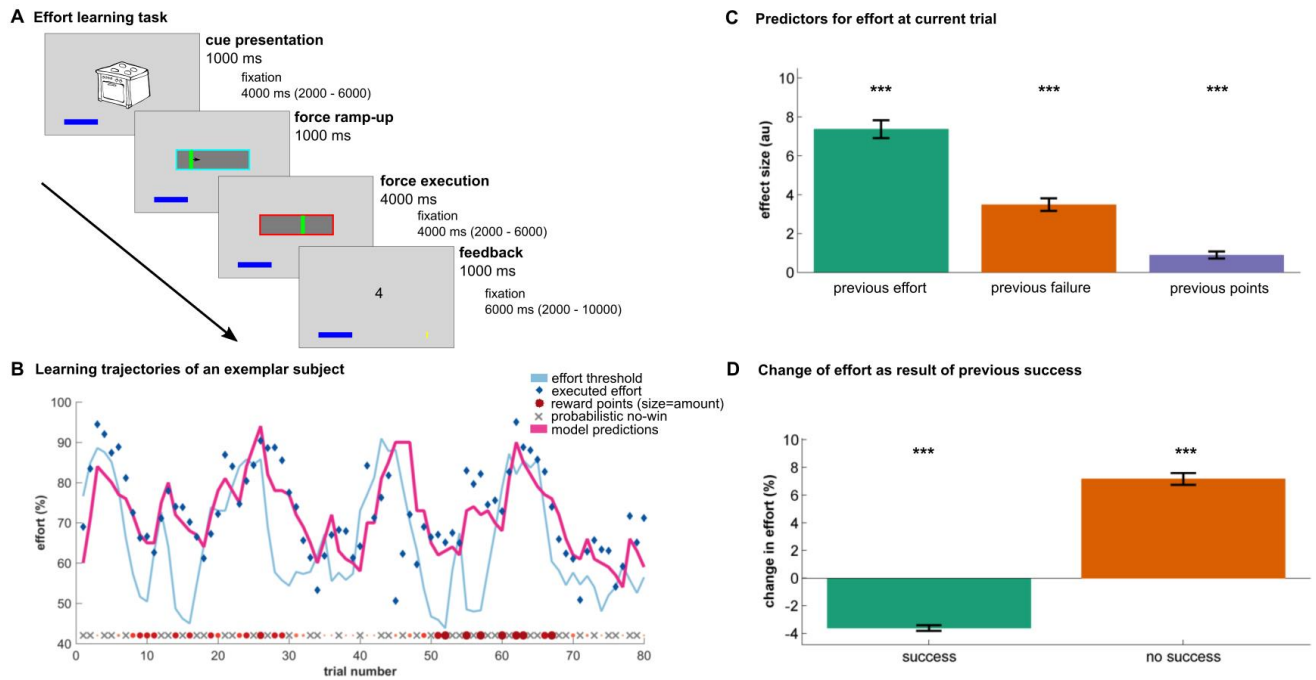


Figure 1. Effort learning task and behavior. (A) Each stimulus is associated with a changing reward magnitude and effort threshold. After seeing a stimulus, subjects exert effort using a custom-made, MR-compatible, pneumatic hand-gripper. Following a ramp-up phase (blue frame), subjects continuously exert a constant force to exceed an effort threshold (red frame phase). The duration of force exertion was kept constant across all trials to obviate temporal discounting that could confound effort execution (14). If successful, subjects received points that were revealed during feedback (here: 4 points). If a subject exerts too little effort (i.e. does not exceed the effort threshold), a cross is superimposed over number of (potential) points indicating they will receive no points on that trial, but still allows subjects to learn about the potential reward associated with a given stimulus. (B) Effort and reward trajectories and actual choice behavior of an exemplar subject. Both effort threshold (light blue line) and reward magnitude (red points) change across time. Rewards were delivered probabilistically, yielding no reward (bottom: grey crosses) on half the trials, independent of whether subjects exceeded a required effort threshold ('0' presented on screen). The exemplar subject can be seen to adapt their behavior (blue diamonds) to a changing effort threshold. Model predictions (pink, depicting maximal subjective net benefit) closely follow subject's behavior. As predicted by our computational model, this subject modulates their effort expenditure based on potential reward. In low effort trials, the subject exerts substantially more effort in high-reward trials (e.g. left side) compared to similar situations in low-reward trials (e.g. right side). (C) Group analysis of 29 male subjects shows that the exerted effort is predicted by factors of previous effort, failure to exceed the threshold on previous trial, and reward magnitude, demonstrating that subjects successfully learned about reward and effort requirements. (D) If subjects fail to exceed an effort threshold they, on average, exert more force in a subsequent trial (orange). Following successful trials, subjects reduce the exerted force and adapt their behavior by minimizing effort (green). (bar plots: mean \pm 1SEM.) *** $p<.001$; a.u. arbitrary units.

A computational framework for effort and reward learning

To probe deeper how precisely subjects learn about reward and effort, we developed a computational reinforcement learning model that predicts effort exerted at every trial, and compared this model with alternative formulations (see Fig. S1, Supplemental Information for detailed model descriptions). Our core model had three distinct components: reward learning, effort learning, and reward-effort arbitration (i.e. effort-discounting) that were used to predict effort execution at each trial. To capture reward learning we included a Rescorla-Wagner like model (31), where reward magnitude learning occurred via a reward prediction error (PE). Note our definition of reward PE deviates from standard notation (10, 11, 32), dictated in part by our design. First, our reward learning algorithm does not track actual rewarded points. Because subjects learned about reward magnitude even if they failed to surpass an effort threshold, and thus not harvest any points (as hypothetical rewards were visible behind a superimposed cross), the algorithm tracks the magnitude of potential reward. This implementation aligns with findings that dopamine encodes a prospective (hypothetical) prediction error signal (33, 34). Second, we employed a probabilistic reward schedule, similar to that used in previous studies of reward learning (35–37). Subjects received 0 points in 50% of the trials (fixed for entire experiment), which in turn did not influence a change in reward magnitude. Using model comparison (all model fits are shown in Fig. S2) we found that a model incorporating learning from these 0-outcome trials outperformed a more optimal model that exclusively learned from actual reward magnitudes. This is in line with classical reinforcement learning approaches (31, 32), wherein reward expectation manifests as a weighted history of all experienced rewards.

To learn about effort, we adapted an iterative logistic regression approach (38), where subjects are assumed to learn about effort threshold based upon a PE. We implemented this approach because subjects did not receive explicit information about the exact height of the effort threshold and instead had to infer it based on their success-history. Here, we define an

effort PE as a difference between a subjects' belief in succeeding, given the executed effort, and their actual success in surpassing an effort threshold. This effort PE updates a belief about the height of the effort threshold and thus the belief of succeeding given a certain effort. Note this does not describe a simple motor or force PE signal given that a force PE would be evident during force execution, in order to signal a deviation between a currently executed and an expected force. Moreover, in our task effort PEs are realized at outcome presentation in the absence of motor execution, signaling a deviation from a hidden effort threshold. Finally, as we are not interested in a subjective, embodied, experience of ongoing force execution we visualized the executed effort by means of a thermometer, an approach used in previous studies (39, 40).

The two independent learning modules, involving effort or reward, are combined at decision time to form an integrated net utility of the stimulus at hand. Previous studies indicate this reward-effort arbitration follows a quadratic or sigmoidal, rather than a hyperbolic, discount function (39–41). As in these prior studies, we also found a sigmoidal discounting function best described this arbitration (Fig. S2) (39, 40). Furthermore, it was better explained if reward magnitude modulated not only the height of this function, but also its indifference point. A sigmoidal form predicts that the utility of a choice will decrease as the associated effort increases. Our model predicts utility is little affected in low effort conditions (cf. Fig. 1B, 6). Moreover, the impact of effort is modulated by reward such that in high reward conditions subjects are more likely to exert greater effort to ensure they surpass an effort threshold (cf. Fig. S1).

To assess whether subjects learned using a PE-like teaching signal, we compared our PE-based learning model to alternative formulations (Fig. S2). A first comparison revealed the PE-learning model outperformed non-learning models where reward or effort was fixed rather than dynamically adapted, supporting the idea that subjects simultaneously learned about, and

adjusted, their behavior to both features. We also found that the effort PE model outperformed a heuristic model that only adjusted its expectation based on success, but did not scale the magnitude of adjustment using an effort PE, in line with a previous study showing a PE-like learning of effort (24). In addition, we compared the model to an optimal reward learning model, which tracks the previous reward magnitude and ignores the probabilistic null outcomes, revealing that a PE-based reward learning outperformed this optimal model.

Finally, because our model was optimized to predict executed effort, we examined whether model-driven PEs also predicted an observed trial-by-trial change in effort. Using multiple regression we found that model-derived PEs indeed have behavioral relevance, and both effort ($t(28)=13.50$, $p<.001$) and reward PEs ($t(28)=2.10$, $p=.045$) significantly predict effort adaptation. This provides model validity consistent with subjects learning about effort and reward using PE-like signals.

Distinct striatal and cortical representations of reward and effort prediction errors

Using fMRI we tested whether model-derived effort and reward PEs are subserved by similar or distinct neural circuitry. We analyzed effort and reward PEs during feedback presentation by entering both in the same regression model (non-orthogonalized; correlation between regressors: $r=.056\pm.074$; Fig. S4). Bilateral VS responded significantly to reward PEs ($p<.05$, whole-brain FWE correction; Table S1 for all activations), but not to effort PEs (Fig. 2A-C). In contrast, dmPFC (peaking in pre-SMA extending into dACC) responded to effort PEs ($p<.05$, whole-brain FWE correction, Fig. 2D-F, Table S1), but not to reward PEs (Fig. 2F). In relation to dmPFC, activity increased if an effort threshold was higher than expected and was attenuated if it was lower than expected, suggestive of an invigorating function for future action. This finding is also in keeping with previous work on effort outcome (24, 25), and a significant influence of dmPFC activity on subsequent change in effort execution (effect

size: 0.04 ± 0.07 ; $t(27) = 2.82$, $p = .009$) supports its behavioral relevance in this task, and is consistent with updating a subject's expectation about future effort requirements (33). Interestingly, the dmPFC area processing effort PE peaks anterior to pre-SMA, and lies anterior to where anticipatory effort signals are found in SMA (Fig. S7), suggesting a spatial distinction between effort anticipation and evaluation.

Neither VS nor dmPFC showed a significant interaction effect between effort and reward PEs (dmPFC: effect size = $-.32 \pm 3.77$; $t(27) = -.45$, $p = .657$; VS: effect size = $-.41 \pm 1.59$; $t(27) = -1.36$, $p = .185$). Post-hoc analysis confirmed that both components of a prediction error, expectation and outcome, were represented in these two regions (Fig. 2B&E), consistent with a full PE encoding rather than simply indexing an error signal (cf. 42). The absence of any effect of probabilistic 0 outcomes in dmPFC, further supports the idea that this region tracks an effort PE, rather than a general negative feedback signal (effect size: $.01 \pm .18$; $t(27) = .21$, $p = .832$). No effects were found for negative-going (inverse) PEs in either reward or effort conditions (e.g. increasing activation for decreasing reward PEs; Table S1). To examine the robustness of this double-dissociation, we sampled activity from independently derived regions-of-interest (ROIs; VS derived from www.neurosynth.org, dmPFC from 25) and, again, found a significant double-dissociation in both VS (Fig. S5A) and dmPFC (Fig. S5B). Moreover, this double-dissociation was also evident in a whole brain comparison between effort and reward PEs (Fig. S5C; dmPFC: MNI: -14 -9 69, $t = 6.05$, $p < .001$ cluster-extent FWE, height-threshold $p = .001$; VS: MNI: 15 9 -6, $t = 6.73$, $p < .001$ cluster-extent FWE).

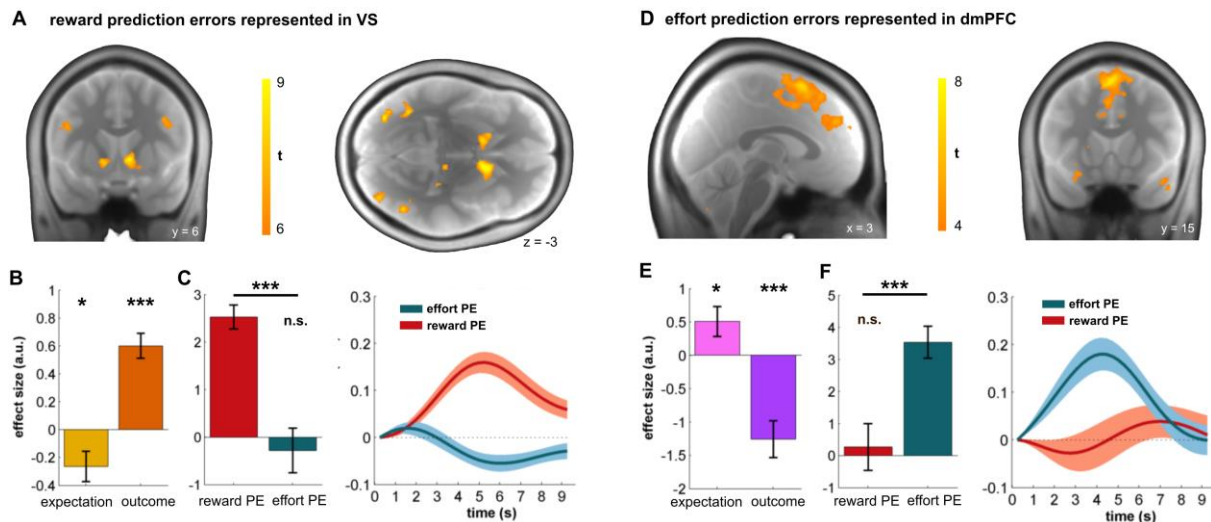


Figure 2. Separate reward and effort PEs in striatum and cortex. (A) Reward PEs encoded in the bilateral ventral striatum ($p < .05$, whole-brain height-FWE correction). (B) Analysis of VS shows an encoding of a full reward PE, reflecting both expectation ($t(27) = -2.44$, $p = .021$) and outcome components ($t(27) = 6.68$, $p < .001$). (C) VS response at outcome showed a significantly increased response to reward PEs relative to effort PEs ($t(27) = 5.80$, $p < .001$), with no evidence for an effect of effort PE ($t(27) = -.60$, $p = .554$). (D) Effort PEs encoding in dorsomedial prefrontal cortex ($p < .05$, whole-brain height-FWE correction). This PE includes components reflecting effort expectation ($t(27) = 2.28$, $p = .030$) and outcome ($t(27) = -4.59$, $p < .001$; i.e. whether or not threshold was surpassed) (E). Activity in dmPFC increased when an effort threshold is higher than expected and decreased when it is lower than expected. (F) dmPFC showed no encoding of a reward PE ($t(27) = -.37$, $p = .714$), and effort PEs were significantly greater than reward PEs in this region ($t(27) = 4.87$, $p < .001$). The findings are consistent with effort and reward PEs being processed in segregated brain regions. (bar and line plots: mean effect size for regressor ± 1 SEM.) * $p < .05$, *** $p < .001$, n.s. $p > .10$

Additionally, we controlled for unsigned (i.e., salience) effort and reward PE signals by including them as additional regressors in the same fMRI model (correlation matrix shown in Fig. S4), as these signals are suggested to be represented in the dmPFC (e.g., 43). Interestingly, when analyzing the unsigned salience PEs, we found that both effort and reward salience PEs elicit responses in regions typical for a salience network (44), and a conjunction analysis across the two salience PEs showed common activation in left anterior insula and intraparietal sulcus (Fig. S3).

Simultaneous representations of effort and reward PEs in the dopaminergic midbrain

We next asked whether an effort PE in dmPFC reflects an influence from a mesocortical input originating within SN/VTA. Dopaminergic cell populations occupy midbrain structures,

substantia nigra and ventral tegmental area (SN/VTA), and project to a range of cortical and subcortical brain regions (45–47). Dopamine neurons in SN/VTA encode reward PEs (11, 48) that are broadcast to VS (12). Similar neuronal populations have been found to encode information about effort and reward (49, 50). Using an anatomically defined mask of SN/VTA, we found that at the time of feedback this region encodes both reward and effort PEs (Fig. 3, $p < .05$, small-volume FWE correction for SN/VTA, Table S1), consistent with a common dopaminergic midbrain origin for striatal and cortical PE representations.

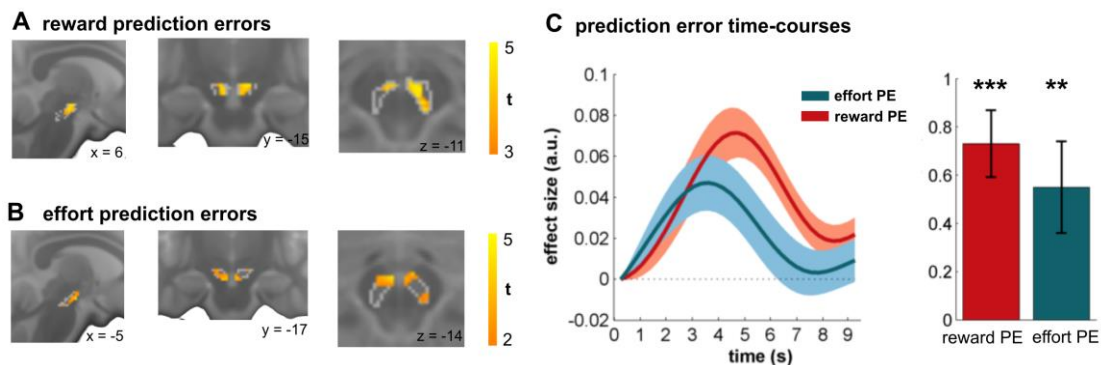


Figure 3. Dopaminergic midbrain encodes reward and effort PEs at outcome. Analysis of the SN/VTA revealed a significant reward (A) and effort (B) PE signal ($p < .05$, small-volume height-FWE correction for anatomically defined SN/VTA). Grey lines depict boundaries of anatomical SN/VTA mask. A simultaneous encoding of both PEs (C; mean activity in anatomical SN/VTA: reward PE: $t(27)=5.26$, $p < .001$, effort PE: $t(27)=2.90$, $p = .007$) suggests a common origin of the dissociable (sub-)cortical representations. Activation increase signals the outcome was better than expected for reward PEs, but indicates an increased effort threshold for effort PEs. (bar and line plots: mean effect size for regressor ± 1 SEM.) ** $p < .01$, *** $p < .001$.

Ascending mesolimbic and mesocortical connections encode PEs

PEs in both dopaminergic midbrain and (sub-)cortical regions suggest that SN/VTA express effort and reward learning signals which are then broadcast to VS and dmPFC. However, there are also important descending connections from dmPFC and VS to SN/VTA (51, 52), providing a potential source of top-down influence on midbrain. To resolve directionality of influence we used a directionally sensitive analysis of effective connectivity. This analysis compares different biophysically-plausible generative models and from this

determines the model with the best-fitting neural dynamics (dynamic causal modelling (DCM; 53); Materials and Methods). We found strong evidence in favor of a model where effort and reward PEs provide a driving influence on ascending compared to descending or mixed connections (Bayesian random-effects model selection: expected posterior probability=.55, exceedance probability=.976, Bayesian omnibus risk=2.83e-4), a finding consistent with PEs computed within SN/VTA being broadcast to their distinct striatal and cortical targets.

A spatial dissociation of effort and reward PEs in SN/VTA

A functional double-dissociation between VS and dmPFC, but a simultaneous representation of both PEs in dopaminergic midbrain, raises a question as to whether effort and reward PEs are spatially dissociable within the SN/VTA. Evidence in rodents and non-human primates point to SN/VTA containing dissociable dopaminergic populations projecting to distinct areas of cortex and striatum (45–47, 54, 55). Specifically, mesolimbic projections to striatum are located in medial parts of the SN/VTA, whereas mesocortical projections to prefrontal areas originate from more lateral subregions (46, 47, 56). However, there is considerable spatial overlap between these neural populations (46, 47) as well as striking topographic differences between species, which cloud a full understanding of human SN/VTA topographical organization (45, 46).

A recent human structural study (57) segregated SN/VTA into ventrolateral and dorsomedial SN/VTA subregions. This motivated us to examine whether SN/VTA effort and reward PEs are dissociable along these axes (Fig. 4A). Using unsmoothed data, we tested how well activity in each SN/VTA voxel is predicted by either effort or reward PEs (Materials and Methods). The location of each voxel along the ventral-dorsal and medial-lateral axis was then used to predict the t-value difference between effort and reward PEs. A significance related to both spatial gradients (Fig. 4B, S6; ventral-dorsal gradient: $\beta=-.151$, 95% confidence intervals

C.I.=-0.281--0.022, $p=.016$; medial-lateral gradient $\beta=.469$, 95% C.I.=.336-.602, $p<.001$) provided evidence that dorsomedial SN/VTA is more strongly affiliated to reward PE encoding, whereas the ventrolateral SN/VTA was more affiliated to effort PE encoding. We also examined whether this dissociation reflected different projection pathways using functional connectivity measures of SN/VTA with VS and dmPFC. Analyzing trial-by-trial BOLD-coupling (after regressing out the task-related effort and reward PE effects) we replicated these spatial gradients, with dorsomedial and ventrolateral SN/VTA more strongly coupled to VS and dmPFC respectively (Fig. 4B; ventral-dorsal: $\beta=-.220$, 95% C.I.=-.373--.067, $p=.002$; medial-lateral; $\beta=.466$, 95% C.I.=.310-.622, $p<.001$). Similar results were obtained when using effect sizes rather than t-values, and when computing gradients on a single-subject level in a summary statistics approach.

To explore further the spatial relationship between SN/VTA, dmPFC and VS, we investigated structural associations between these areas. We used structural co-variance analysis (58), which investigates how gray matter (GM) densities co-vary between brain regions, and has been shown sensitive to identify anatomically and functionally relevant networks (59). Specifically, we asked how well gray matter (GM) density in each SN/VTA voxel is predicted by dmPFC and VS GM (regions defined by their functional activations), and their spatial distribution between subjects. Importantly, this analysis is entirely independent from our previous analyses as it investigates individual GM differences, as opposed to trial-by-trial functional task associations. Note there was no association between BOLD response and GM (dmPFC: $r=.155$, $p=.430$; VS: $r=.100$, $p=.612$; SN/VTA effort PEs: $r=.067$, $p=.737$; reward PEs: $r=.079$, $p=.690$). We found both spatial gradients were significant (Fig. 4C, ventral-dorsal gradient $\beta=-.012$, 95% C.I.=-.018--.006, $p<.001$; medial-lateral; $\beta=.007$, 95% C.I.=.002-.013, $p=.007$), suggesting that SN/VTA GM was more strongly associated with dmPFC GM in

ventrolateral, and with VS GM in dorsomedial areas, thus confirming the findings of our previous analyses.

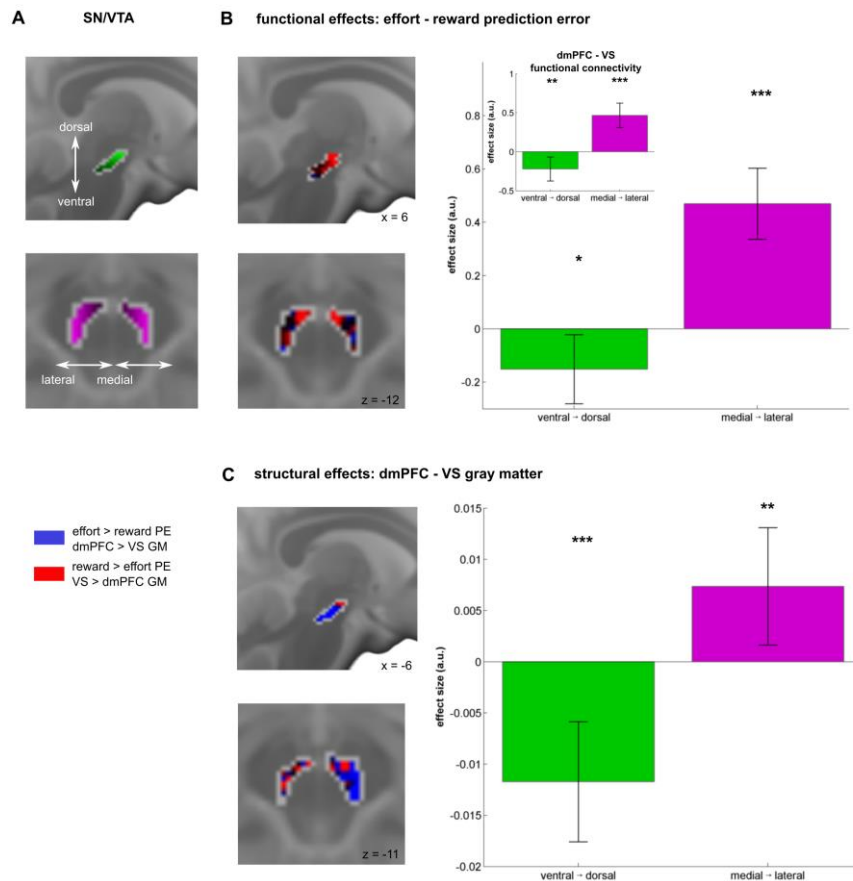


Figure 4. SN/VTA spatial expression mirrors cortical and striatal organization. Effort and reward PEs in the SN/VTA follow a spatial distribution along a ventral-dorsal (green color gradients) and a medial-lateral (violet color gradients) gradient respectively (A, B). Multiple regression analysis revealed that ventral (green bars) and lateral (violet bars) voxels of the SN/VTA are representing effort PE more strongly, relative to reward PEs. Effect maps (left) show that dorsomedial voxels preferentially encode reward PEs (red colors), whereas ventrolateral voxels more strongly encode effort prediction errors (blue colors; also see Fig. S6). A functional connectivity analysis (B, small bar plot) revealed SN/VTA expressed spatially distinct functional connectivity patterns: ventral and lateral voxels are more likely to co-activate with dmPFC, whereas dorsal and medial SN/VTA voxels are more likely to co-active with VS activity. (C) Gray matter (GM) analysis replicates functional findings in revealing that gray matter in ventro-lateral SN/VTA covaried with dmPFC GM density, whereas dorso-medial SN/VTA GM was associated with VS GM density. Our findings of consistent spatial gradients within the SN/VTA thus suggest distinct mesolimbic and mesocortical pathways that can be analyzed along ventral-dorsal and medial-lateral axes in humans. (bar graphs: effect size \pm 95% C.I.). * $p < .05$, ** $p < .01$, *** $p < .001$.

Apathy is predicted by prefrontal, but not striatal function

Several psychiatric disorders, including schizophrenia, express effort-related deficits (e.g., 28–30). A long-standing hypothesis assumes an imbalance between striatal and cortical dopamine (26, 60, 61), involving excess dopamine release in striatal (62, 63) but deficient dopamine release in cortical areas (64). While the former is linked to positive symptoms, such as hallucinations, the latter is considered relevant to negative symptoms, such as apathy (26, 28, 29). Given the striatal-cortical double-dissociation, we examined whether apathy scores in our subjects, as assessed using the apathy evaluation score (AES; 65), were better predicted by dmPFC or VS activation. We ran a 5-fold cross-validated prediction of AES total score using either mean functional responses in dmPFC or VS (using same functional ROIs as above), utilizing effort and reward PE responses in both regions (correlation between predictors: dmPFC: $r=.149$, $p=.458$; VS: $r=.144$, $p=.473$). We found that dmPFC activations were highly predictive of apathy scores (Fig. 5A, $p<.001$, using permutation tests, see Materials and Methods). Interestingly, the effect sizes for both, reward ($.573\pm.050$) and effort ($.351\pm.059$) prediction errors in the dmPFC showed a positive association with apathy, meaning that the bigger a prediction error in the dmPFC, the more apathetic a person was. Activity in the VS did not predict apathy (Fig. 5B, $p=.796$). This was also reflected by a finding that extending a dmPFC-prediction model with VS activation did not improve apathy prediction ($p=.394$). There was no association between dmPFC ($r=-.225$, $p=.258$) or VS ($r=.142$, $p=.481$) GM and apathy. Furthermore, we found no link between overt behavioral variables and apathy (money earned: $r=-.02$, $p=.927$; mean exerted effort: $r=.00$, $p=.99$; standard deviation exerted effort: $r=-.24$, $p=.235$; N trials not succeeding effort threshold: $r=-.13$, $p=.525$). These findings suggest self-reported apathy was specifically related to PE processing in dmPFC. Intriguingly, finding an effect of dmPFC reward PEs on apathy in the absence of a reward PE signal in this area at a group level (Fig. 2F) suggests an interpretation that apathy might be related to an

impoverished functional segregation between mesolimbic and mesocortical pathways. Indeed, we find a significant effect of reward PEs only in more apathetic subjects (median split analysis: low apathy group: effect size: -1.23 ± 3.43 ; $t(13) = -1.35$, $p = .201$; high apathy group: 2.37 ± 2.83 ; $t(12) = 3.02$, $p = .011$) supporting this notion.

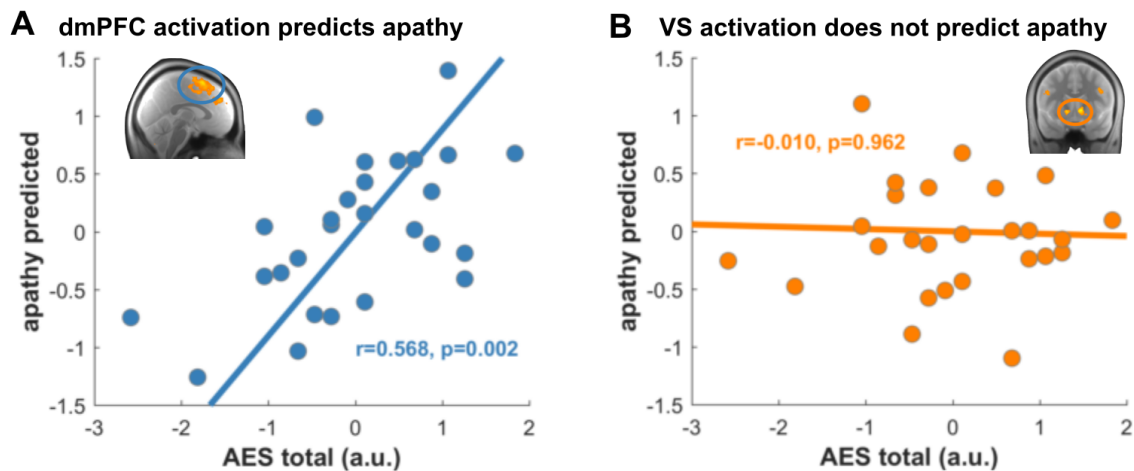


Figure 5. Apathy related to cortical, but not striatal function. (A) Prediction error signals in the dmPFC significantly predicted apathy scores as assessed using an apathy self-report questionnaire (AES total score). (B) PE signals in VS were not predictive of apathy.

VS encodes subjective net benefit

During anticipation it is suggested VS encodes an overall net benefit or integrated utility signal, incorporating both effort and reward (2, 66). We examined whether VS activity during cue presentation tracked both reward and effort expectation. Using a region-of-interest analysis (bilateral VS extracted from reward PE contrast), we found a significant reward expectation effect (Fig. 6, $t(27) = 2.23$, $p = .035$), but no effect of effort expectation ($t(27) = .72$, $p = .476$) at stimulus presentation. These findings accord with predictions of our model, where subjective value increases as a direct function of reward, but where subjective value does not decrease linearly with increasing effort (Fig. 6, left panel). Instead, the sigmoidal function of our reward-effort arbitration model suggests that effort influences subjective value through an interaction with reward. This predicts is that during low effort trials reward has little impact on subjective value, whereas for high effort the differences between low and high reward trials will engender

significant change in subjective value. We formally tested this by examining the interaction between effort and reward, and found a significant effect (Fig. 6, $t(27)=2.94$, $p=.007$). A post-hoc median-split analysis confirmed the model's prediction evident in a significant effect of reward in high ($t(27)=3.28$, $p=.003$), but not in low effort trials ($t(27)=.61$, $p=.547$).

ventral striatum encodes expected reward and its interaction with effort

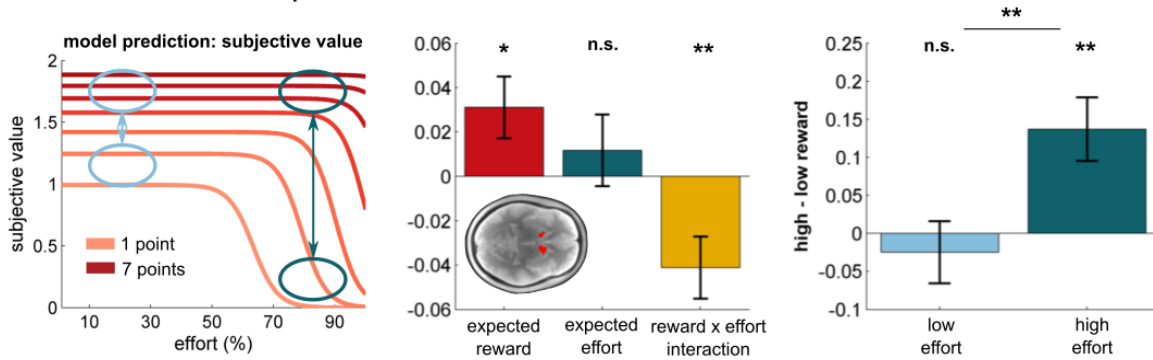


Figure 6. Unified and distinct representations of effort and reward during anticipation. Ventral striatum encoded a subjective value signal in accord with predictions of our computational model (left panel). A main effect of expected reward (middle panel) reflected an increase in subjective value with higher reward. The reward x effort interaction (middle panel) and the subsequent median-split analysis (right panel) shows a difference between high and low rewards is more pronounced during high-effort trials, as predicted by our model (blue arrows in left panel). A similar interaction effect was found when using a literature-based VS ROI (reward x effort expectation: $t(27)=-2.28$, $p=.016$; high - low reward in high effort: $t(27)=2.51$, $p=.018$; high - low reward in low effort: $t(27)=-.41$, $p=.684$). (bar and line plots: mean effect size for regressor ± 1 SEM.) * $p<.05$, ** $p<.01$, n.s. $p>.10$

Discussion

Tracking multiple aspects of a choice option, such as reward and effort, is critical for efficient decision making and demands simultaneous learning of these choice attributes. Here we show that the brain exploits distinct mesolimbic and mesocortical pathways to learn these choice features in parallel with a reward PE in VS and effort PE represented in dmPFC. Critically, we demonstrate that both types of PE at outcome satisfy requirements for a full PE, representing both an effect of expectation and outcome (cf. 42) and thus extend previous single-attribute learning studies for reward PE (e.g., 12, 36, 42) and effort outcomes (24, 25).

Our study is the first to show a functional double-dissociation between VS and dmPFC, highlighting their preferential processing of reward and effort PE respectively. This functional and anatomical segregation provides an architecture that can enable the brain to learn about multiple decision choice features simultaneously, specifically predictions of effort and reward. Although dopaminergic activity cannot be assessed directly using fMRI, both effort and reward PEs were evident in segregated regions of dopaminergic rich midbrain, and where an effective connectivity analysis indicated a directional influence from SN/VTA towards subcortical (reward PE) and cortical (effort PE) targets via ascending mesolimbic and mesocortical pathways respectively.

Dopaminergic midbrain is thought to comprise several distinct dopaminergic populations that have dissociable functions (54, 56, 67, 68). Here, we demonstrate a segregation between effort and reward learning within SN/VTA across the domains of task activation, functional connectivity and gray matter density. In SN/VTA a dorso-medial encoding of reward PEs, and a ventro-lateral encoding of effort PEs, extends on previous studies on SN/VTA subregions (56, 57, 67, 68) by demonstrating this segregation has functional implications that are exploited during multi-attribute learning. In contrast to previous studies on SN/VTA substructures (56, 67–69) we performed whole-brain imaging,

which allowed us to investigate the precise interactions between the dopaminergic midbrain and striatal/cortical areas. However, this required a slightly lower spatial SN/VTA resolution than previous studies (56, 67–69) restricting our analyses to spatial gradients across the entire SN/VTA rather than subregion analyses. We speculate the dorsomedial region showing reward PE activity is likely to correspond to a dorsal tier of dopamine neurons known to form mesolimbic connections projecting to VS regions (55) (see Fig S6). By contrast the ventrolateral region expressing effort PE activation is likely to be related to ventral tier dopamine neurons (46, 55) that form a mesocortical network targeting dmPFC and surrounding areas (46, 47).

Our computational model showed learning about choice features exploits PE-like learning signals, and in so doing extends on previous models by integrating reward and effort learning (cf. 24, 70) with effort-discounting (39–41). The effort PE encoded in dmPFC can be seen as representing an adjustment in belief about the height of a required effort threshold. It is interesting to speculate about the functional meaning of this PE signal, such as whether this is more likely to signal motivation or the costs of a stimulus. Our findings that effort PEs have an invigorating function favor the former notion, though we acknowledge we cannot say whether effort PEs would also promote avoidance if our task design included an explicit option to default. External support for an invigorating function comes from related work on dopamine showing that it broadcasts a motivational signal (16, 71) that in turn influences vigor (72–74). Interestingly, direct phenomenological support for such a motivational signal comes from observations in human subjects undergoing electrical stimulation of cingulate cortex, who report a motivational urge and a determination to overcome effort (75).

VS is suggested to encode net benefit or integrated utility of a choice option (1, 30), in simple terms the value of a choice integrated over benefits (rewards) and costs (effort). The interaction between effort and reward expectations we observe in VS during anticipation is

consistent with encoding of an integrated net benefit signal, but in our case this occurs exclusively during anticipation. In accordance with our model reward magnitude is less important in low effort trials but assumes particular importance during high effort trials. However, the absent reward effect in low effort trials contrasts with studies that find reward-related signals in the VS during cue presentation, but in the latter there is no effort requirement (e.g. 76). This deviation from previous findings might reflect that subjects in our task need to execute effort prior to obtaining a reward. Nevertheless, the convergence of our model predictions and the interaction effect seen in VS support the notion that a net benefit signal is formed at the time of action anticipation, when an overall stimulus value is important in preparing an appropriate motor output.

Effortful decision making assumes considerable interest in the context of pathologies such as schizophrenia, and may provide for a quantitative metric of negative symptoms (28). An association between impaired effort arbitration and negative symptoms in patients with schizophrenia (e.g. 29) supports this conjecture, though it is unknown whether such an impairment is tied to a prefrontal or striatal impairment. Within our volunteer sample apathy was related to aberrant expression of learning signals in dmPFC, but not in VS. This suggest negative symptoms may be linked to a breakdown in a functional segregation between mesolimbic and mesocortical pathways, and this idea accords with evidence of apathetic behavior following seen following ACC lesioning (23).

We used a naturalistic task reflecting the fact that in many environments effort requirements are rarely explicit and are usually acquired by trial and error, while reward magnitudes are often explicitly learned. Although this entails a difference in how feedback is presented, we consider it unlikely to influence neural processes, given that previous studies with balanced designs showed activations in remarkably similar regions to ours (e.g., 24, 25) and because prediction error signals are generally insensitive to outcome modality

(primary/secondary reinforcer, magnitude/probabilistic feedback, learning/no learning) (e.g., 12, 36, 70, 77). Moreover, the specificity of the signals in VS and dmPFC for either reward or effort, including a modulation by expectation, favors a view that the pattern of responses observed in these regions reflect specific prediction error signals as opposed to simple feedback signals.

It is interesting to conjecture whether a spatial dissociation that we observe for simultaneous learning of reward and effort also holds if subjects learn choice-features at different times, or learn about choice features other than reward and effort. Our finding of a mesolimbic network encoding reward PEs during simultaneous learning accords with results from simple reward-alone learning (e.g., 12). This, together with a known involvement of dmPFC in effort-related processes (24, 25, 40), renders it likely that the same pathways are exploited in unidimensional learning contexts. However, it remains unknown whether the same spatial segregation is needed to learn about multiple forms of reward that are associated with VS activity (e.g., monetary, social).

In summary, we show that simultaneous learning about effort and reward involves dissociable mesolimbic and mesocortical pathways, with VS encoding a reward learning signal and dmPFC encoding an effort learning signal. Our data indicate these PE signals arise within SN/VTA where an overlapping, but segregated, topological organization reflects distinct neural populations projecting to cortical and striatal regions respectively. An integration of these segregated signals occurs in VS in line with an overall net benefit signal of an anticipated action.

Materials and Methods

Subjects

Twenty-nine healthy, right-handed, male, human volunteers (age: $24.1y \pm 4.5$, range: 18-35y) were recruited from local volunteer pools to take part in this experiment. All subjects were familiarized with the hand grippers and the task before entering the scanner (Supplementary Information). Subjects were paid on an hourly basis plus a performance-dependent reimbursement. We focused on male subjects because we wanted to minimize potential confounds which we observed in a pilot study, for example fatigue in high force exertion trials. One subject was excluded from fMRI analysis due to equipment failure during scanning. The study was approved by the UCL research ethics committee and all subjects gave written informed consent.

Task

The goal of this study was to investigate how humans simultaneously learn about reward and effort in an ecologically realistic manner. In the task (Fig. 1A), subjects were presented with one of two stimuli (duration 1000ms). The stimuli (pseudo-randomized, no more than 3 presentations of one stimulus in a row) were indicative of a potential reward (1 to 7 points, 50% reward probability) and an effort threshold that needed to be surpassed (range of effort threshold between 40% and 90% maximal force) in order to harvest a reward. Both points and effort thresholds slowly changed over time in a Gaussian random-walk-like manner (Fig. 1B), whereas the reward outcome probability remained stationary across the entire experiment, and subjects were informed about this beforehand. These trajectories were constructed so that reward and effort were de-correlated, and indeed the realized correlation between effort and reward prediction errors were minimal. Moreover, independent trajectories for effort and reward allowed us to cover a wide range of effort and reward expectation combinations

enabling us to comprehensibly assess a reward-effort arbitration function. Thus, to master the task, subjects had to simultaneously learn both reward and effort thresholds. After a jittered fixation cross (mean 4000ms, uniformly distributed between 2000 and 6000ms), subjects had to squeeze a force-gripper with their right hand for 5000ms. During the first 1000ms, the subjects increased their force to the desired level (as indicated by a horizontal thermometer; blue frame phase). During the last 4000ms, subjects maintained a constant force (red frame phase) and released as soon as the thermometer disappeared from the screen. After another jittered fixation cross (mean 4000ms, 2000-6000ms), subjects received feedback whether and how many points they received for this trial (duration 1000ms). If the exerted effort was above the effort threshold, subjects received the points that were on display. If the subjects' effort did not exceed the threshold, a cross appeared above the number on display, which indicated that the subject did not receive any points for that trial. More details about the task are provided in the Supplementary Information.

Behavioral analysis

To assess the factors that influence effort execution (Fig. 1C), we used multiple regression to predict the exerted effort at each trial. As predictors, we entered the exerted effort as well as the number of points (displayed during feedback) on the previous trial, and whether the force threshold was successfully surpassed on the previous trial. Please note that the previous trial was determined as the last trial that the same stimulus was presented. The regression weights of the normalized predictors were obtained for each individual and then tested for consistency across subjects using t-tests. To test how subjects changed their effort level based on whether they surpassed the threshold or not ('success'), we analyzed the change in effort conditioned on their success. For each subject, we calculated the average change in

effort for success and non-success trials and then tested consistency using t-tests across all subjects (Fig. 1D).

Computational Modelling

We developed novel computational reinforcement learning models (32) to formalize the processes underlying effort and reward learning in this task. All models were fitted to individual subject's behavior (executed effort at each trial) and a model comparison using BIC was performed to select the best-fitting model (Fig. S2). The preferred model was then used for the fMRI analysis. The models and model comparison are detailed in the Supplemental Information.

fMRI data acquisition and preprocessing

MRI was acquired using a Siemens Trio 3T scanner, equipped with a 32-channel head-coil. We used an EPI-sequence that was optimized for minimal signal-dropout in striatal, medial prefrontal and brain-stem regions (78). Each volume was formed of 40 slices with 3mm isotropic voxels (TR=2.8s, TE=30ms, slice tilt: -30° (T>C)). A total of 1252 scans were acquired across all four sessions. The first 6 scans of each session were discarded to account for T1 saturation effects. Additionally, field maps (3mm isotropic, whole-brain) were acquired to correct the EPIs for field strength inhomogeneity.

All functional and structural MRI analyses were performed using SPM12 (<http://www.fil.ion.ucl.ac.uk>). The EPIs were first realigned and unwarped using the field maps. EPIs were then co-registered to the subject-specific anatomical images and normalized using DARTEL-generated (79) flow-fields, which resulted in a final voxel resolution of 1.5mm (standard size for DARTEL normalization). For the main analysis, the normalized EPIs were smoothed with a 6mm FWHM-kernel to satisfy the smoothness assumptions of the statistical

correction algorithms. For the gradient-analysis of the SN/VTA ('unsmoothed analysis'), we used a small smoothing kernel of 1mm to preserve more of the voxel-specific signals. We applied this very small smoothing kernel rather than no kernel to prevent aliasing artefacts that naturally arise from the DARTEL-normalization procedure.

fMRI data analysis

The main goal of the fMRI analysis was to determine the brain regions that track reward and effort prediction errors (PEs). To this end, we used the winning computational model and extracted the model predictions for each trial. To derive the model predictions, we used the average parameter estimates across all subjects, similar to previous studies (43, 80–84). This ensures more regularized predictions and does not introduce subject-specific biases. At the time of feedback, we entered four parametric modulators: effort PEs, reward PEs, absolute effort PEs, absolute reward PEs. For all analyses, we normalized the parametric modulators beforehand and disabled the orthogonalization procedure in SPM (correlation between regressors shown in Fig. S4). This means that all parametric modulators compete for variance and we thus only report effects that are uniquely attributable to the given regressor. The sign of the PE regressors was set so that positive reward PEs means that a reward is better than expected, and for effort PEs a positive PE means that the threshold is higher than expected (more effort is needed). The task sequences were designed so as to minimize a correlation between effort and reward PEs, as well as between expectation and outcome signals within a PE (effort PE: $r=.087\pm.105$; reward PE: $r=-.002\pm 0.109$), and thus to maximize sensitivity or our analyses. To control for other events of the task, we added the following regressors as nuisance covariates: stimulus presentation with parametric modulators for expected reward, expected effort, expected reward*effort interaction, stimulus identifier. To control for any

movement-related artefacts, we also modelled the force execution period (block duration: 5000ms) with executed effort as parametric modulator. Moreover, we regressed out movements using the realignment parameters, as well as pulsatile and breathing artefacts (85–88). Each run was modelled as a separate session to account for offset differences in signal intensity.

On the second level, we used the standard summary statistics approach in SPM (89) and computed the consistency across all subjects. We used whole-brain family-wise error correction $p < .05$ to correct for multiple comparisons (if not stated otherwise) using settings that do not show any biases in discovering false positives (90, 91). We examined the effect of each regressor-of-interest (effort, reward PE) using a one-sample t-test to assess the regions in which there was a representation of the regressor. Subsequent analyses (Fig 2B-C, E-F) were performed on the peak voxel in the given area. Prediction errors were compared using paired t-tests. To assess the effect of effort and reward PEs on the SN/VTA, we used the same GLM applying small-volume FWE-correction (uncorrected threshold $p < .001$) based on our anatomical SN/VTA mask (see below), similar to previous studies (e.g. 92, 93).

For the analysis of the cue phase, we extracted responses of a VS-ROI and then assessed the impact of our model-derived predictors expected effort, expected reward, and their interaction.

Effective connectivity analysis

To assess whether PE signals are more likely to be projected from SN/VTA to (sub-)cortical areas or vice versa, we ran an effective connectivity analysis using dynamical causal modelling (DCM; 53). DCM allows the experimenter to specify, estimate and compare biophysically plausible models of spatiotemporally distributed brain networks. In the case of fMRI, generative models are specified which describe how neuronal circuitry causes the BOLD response which in turn elicits the measured fMRI time series. Bayesian model selection (94) is

used to determine which of the competing models best explains the data (in terms of balancing accuracy and complexity), drawing upon the slow emergent dynamics which result from the interaction of fast neuronal interactions (referred to as the slaving principle; 89).

We compared several models, all consisting of three regions: SN/VTA (using the anatomical ROI), VS, and dmPFC (using the functional contrasts for ROI definition). As fixed inputs, we used the onset of feedback as a stimulating effect on all three nodes. We assumed bidirectional connections between SN/VTA and dmPFC/VS regions, reflecting the well-known bidirectional communication. The models differed in how PEs influenced these connections. Based on the assumption that PEs are computed in the originating brain structure and influence the downstream brain region, we tested whether PEs modulated the connections originating from SN/VTA, or targeting it. This same approach (PEs affecting modulation of intrinsic connections) was used in previous studies investigating the effects of PEs on effective connectivity (e.g., 96, 97). We compared 6 models in total. In the winning ascending model, reward and effort PEs (each only modulating the connection to its (sub-) cortical target region) modulated the ascending connections (e.g. reward PEs modulated connectivity from SN/VTA to VS). In the descending model, PEs modulated the descending connections from VS and dmPFC to SN/VTA. Additional models tested whether only having one ascending modulation (either effort or reward PEs), or having one ascending and one descending modulation, fitted the data better. DCMs were fitted for each subject and run separately and Bayesian random-effects comparison (94) was used for model comparison.

fMRI analysis of SN/VTA gradients

For the analysis of the SN/VTA gradients with the unsmoothed data, the model was identical to the one above, with the exception that the feedback on each trial was modelled as a separate regressor. This allowed us to obtain an estimate of the BOLD response, separately

for each trial (necessary for functional connectivity analysis) in keeping with the same main effects as in normal mass-univariate analyses (cf. 37). These responses were then used to perform our gradient analyses.

We performed two SN/VTA gradient analyses with the functional data. For the PE analysis, we used the model-derived PEs (as described above) to predict the effects of effort and reward PEs on each voxel of our anatomically-defined SN/VTA mask. We then calculated t-tests for each voxel on the second level, using the beta coefficients of all subjects. As we were interested whether there is a spatial dissociation/gradient between the two PE types, we then calculated the difference of the absolute t-values between the two prediction errors, for each voxel separately. This metric allows us to measure whether a voxel was more predictive of effort or reward PEs. To ensure that we only use voxels that have some response to the PEs, we discarded voxel that has an absolute t-value <1 for both prediction errors. We used the absolute of the t-values for our contrast to account for potential negative encoding.

To calculate the gradients, we used a multiple regression approach to predict the t-value differences (e.g., effort - reward PE). As predictors, we used the voxel location in a ventral-dorsal gradient and a voxel location in a medial-lateral gradient. Both gradients entered the regression, together with a nuisance intercept. This analysis resulted in a beta weight for each of the gradients which indicates whether the effect of the prediction errors follows a spatial gradient or not. We obtained the 95%-confidence intervals (95% C.I.) of the beta weights and calculated the statistical significance using permutation tests (10 000 iterations; randomly permuting the spatial coordinates of each voxel).

For the second, functional connectivity analysis, we used the very same pipeline. However, instead of using model-derived prediction errors, we now used the BOLD-response for every trial from the dmPFC and bilateral VS (mean activation across entire ROI). The ROIs were determined based on task main effect (dmPFC based on effort PEs, VS based on reward

PEs, both thresholded at $p_{FWE} < .05$). To ensure this analysis did not reflect the task effects, we regressed out the task effect (reward / effort PEs) prior to the main analysis.

We found similar effect when using beta weights (which do not take measurement uncertainty into account) instead of t-values, and also if we include all voxels, irrespective of whether they respond to any of the PEs. Similar results were also obtained when using a summary statistics approach, in which spatial gradients were obtained for each single subject.

Predicting apathy through BOLD responses

To assess whether self-reported apathy – a potential reflection of non-clinical negative symptoms (27) – was related to neural responses in our task, we tested whether we can predict apathy by using task-related activation. Apathy was assessed using a self-report version of the apathy evaluation scale (AES; 65) (missing data from one subject). We used the total scores as a dependent variable in a 5-fold cross-validated regression (cf. 84, 98). To assess whether apathy was more closely linked to dmPFC or VS activation, we used the activation in the given ROI (using mean activation at $p < .05$ FWE-ROI, same as in previous analyses), including both effort and reward prediction error signals at the time of outcome. To assess prediction accuracy, we then calculated the L2-norm between predicted and true apathy scores across all subjects (cf. 98). To establish a statistical null-distribution, we ran permutation tests by randomly shuffling the PE responses. To assess whether the VS predictors improved a dmPFC prediction, we compared the predictive performance of the dmPFC model to an extended model with VS activations as additional predictors. Permutation tests (by permuting the additional regressors) were again used to assess significance between the dmPFC and the full model.

sMRI data acquisition and analysis

Structural images were acquired using quantitative multi-parameter maps (MPMs) in an 3D multi-echo fast low angle shot (FLASH) sequence with a resolution of 1mm isotropic voxels (99). Magnetic transfer (MT) images were used for gray matter quantification, as they are particularly sensitive for subcortical regions (100). In total, 3 different FLASH sequences were acquired with different weightings: predominantly MT ($TR/\alpha = 23.7 \text{ ms}/6^\circ$; off-resonance Gaussian MT pulse of 4 ms duration preceded excitation, 2 kHz frequency offset, 220° nominal flip angle), proton density (PD; $23.7 \text{ ms}/6^\circ$), and T1-weighting ($18.7 \text{ ms}/20^\circ$) (101). To increase signal-to-noise ratio, we averaged signals of six equidistant bipolar gradient echoes (TE: 2.2-14.7 ms). To calculate the semi-quantitative MT-maps, we used mean signal amplitudes and additional T1 maps (102), and additionally eliminated influences of B1 inhomogeneity and relaxation effects (103).

To normalize functional and structural maps, we segmented the MTs maps (using heavy bias regularization to account for the quantitative nature of MPMs), and generated flowfields using DARTEL (79) with the standard settings for SPM12. The flowfields were then used for normalizing functional as well as structural images. For the normalization of the structural images (MT), we used the VBQ toolbox in SPM12 with an isotropic Gaussian smoothing kernel of 3mm.

To investigate anatomical links between SN/VTA and VS and dmPFC, we performed a (voxel-based morphometry [VBM]-based) structural co-variance analysis (58). The approach assumes that brain regions that are anatomically and functionally related (e.g. form a common network) should co-vary in gray matter density between subjects. This means that subjects with a strong expression of a dmPFC gray matter density should also express greater gray matter density in ventrolateral SN/VTA, possibly reflecting genetic, developmental or environmental influences (59). We used the segmented, normalized gray matter MT maps and applied the Jacobian of the normalization step to preserve total tissue volume (104), as reported in a

previous study (84). To account for differences in global brain volume, we calculated the total intracranial volume and used it as a nuisance regressor in the analysis. For each subject, we extracted the mean gray matter density in dmPFC and bilateral VS (mask derived from functional contrasts, thresholded at $p_{FWE} < .05$, see above). Additionally, we extracted the gray matter density of each voxel in our SN/VTA mask. We then calculated the effect of dmPFC and VS gray matter in a linear regression model predicting the gray matter density in every voxel in the SN/VTA. Similar to our functional analysis, we calculated the difference of the t-values for dmPFC and VS for each voxel (dmPFC-VS). These were then used for the same spatial gradient analysis as in the analysis described above.

For all our SN/VTA analyses, we used a manually drawn anatomical region-of-interest in MRICron (105). We used the mean structural MT image where SN/VTA can be easily distinguished from surrounding areas as a bright white stripe (45), similar to previous studies (92, 93).

Data availability

Imaging results are available online on <http://neurovault.org/collections/IAYMWZIY/>.

Acknowledgements

We thank Francesco Rigoli, Peter Zeidman, Philipp Schwartenbeck, and Gabriel Ziegler for helpful discussions. We also thank Peter Dayan and Laurence Hunt for comments on an earlier version of the manuscript. We thank Francesca Hauser-Piatti for creating graphical stimuli and illustrations. Finally, we thank Al Reid for providing the force grippers. A Wellcome Trust Cambridge-UCL Mental Health and Neurosciences Network grant (095844/Z/11/Z) supported all authors. RJD holds a Wellcome Trust Senior Investigator Award (098362/Z/12/Z). The UCL-Max Planck Centre is a joint initiative supported by UCL

and the Max Planck Society. The Wellcome Trust Centre for Neuroimaging is supported by core funding from the Wellcome Trust (091593/Z/10/Z). The authors declare no competing financial interests.

Author contributions

TUH, EE & RJD designed research; TUH & EE performed research; TUH analyzed data; TUH, EE & RJD wrote the paper.

Conflict of Interest

The authors declare no conflict of interest.

References

1. Phillips PEM, Walton ME, Jhou TC (2007) Calculating utility: preclinical evidence for cost-benefit analysis by mesolimbic dopamine. *Psychopharmacology (Berl)* 191(3):483–495.
2. Crosson PL, Walton ME, O'Reilly JX, Behrens TEJ, Rushworth MFS (2009) Effort-based cost-benefit valuation and the human brain. *J Neurosci Off J Soc Neurosci* 29(14):4531–4541.
3. Walton ME, Kennerley SW, Bannerman DM, Phillips PEM, Rushworth MFS (2006) Weighing up the benefits of work: behavioral and neural analyses of effort-related decision making. *Neural Netw Off J Int Neural Netw Soc* 19(8):1302–1314.
4. Zhou QY, Palmiter RD (1995) Dopamine-deficient mice are severely hypoactive, adipsic, and aphagic. *Cell* 83(7):1197–1209.
5. Homer (1996) *The Odyssey* (Penguin Classics). Reprint edition.
6. Linares OF (1976) “Garden hunting” in the American tropics. *Hum Ecol* 4(4):331–349.
7. Rosenberg DK, McKelvey KS (1999) Estimation of Habitat Selection for Central-Place Foraging Animals. *J Wildl Manag* 63(3):1028–1038.
8. Kolling N, Behrens TEJ, Mars RB, Rushworth MFS (2012) Neural mechanisms of foraging. *Science* 336(6077):95–98.
9. Hull C (1943) *Principles of behavior: an introduction to behavior theory* (Appleton-Century, Oxford, England).
10. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J Neurosci* 16(5):1936–1947.
11. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593–1599.
12. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442(7106):1042–1045.
13. Ferenczi EA, et al. (2016) Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science* 351(6268):aac9698.
14. Floresco SB, Tse MTL, Ghods-Sharifi S (2008) Dopaminergic and glutamatergic regulation of effort- and delay-based decision making. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol* 33(8):1966–1979.
15. Schweimer J, Hauber W (2006) Dopamine D1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learn Mem Cold Spring Harb N* 13(6):777–782.
16. Salamone JD, Correa M (2012) The mysterious motivational functions of mesolimbic dopamine. *Neuron* 76(3):470–485.

17. Denk F, et al. (2005) Differential involvement of serotonin and dopamine systems in cost-benefit decisions about delay or effort. *Psychopharmacology (Berl)* 179(3):587–596.
18. Salamone JD, Cousins MS, Bucher S (1994) Anhedonia or anergia? Effects of haloperidol and nucleus accumbens dopamine depletion on instrumental response selection in a T-maze cost/benefit procedure. *Behav Brain Res* 65(2):221–229.
19. Walton ME, et al. (2009) Comparing the role of the anterior cingulate cortex and 6-hydroxydopamine nucleus accumbens lesions on operant effort-based decision making. *Eur J Neurosci* 29(8):1678–1691.
20. Schweimer J, Saft S, Hauber W (2005) Involvement of catecholamine neurotransmission in the rat anterior cingulate in effort-related decision making. *Behav Neurosci* 119(6):1687–1692.
21. Walton ME, Bannerman DM, Rushworth MFS (2002) The role of rat medial frontal cortex in effort-based decision making. *J Neurosci Off J Soc Neurosci* 22(24):10996–11003.
22. Walton ME, Bannerman DM, Alterescu K, Rushworth MFS (2003) Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effort-related decisions. *J Neurosci Off J Soc Neurosci* 23(16):6475–6479.
23. Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MFS (2006) Separate neural pathways process different decision costs. *Nat Neurosci* 9(9):1161–1168.
24. Skvortsova V, Palminteri S, Pessiglione M (2014) Learning to minimize efforts versus maximizing rewards: computational principles and neural correlates. *J Neurosci Off J Soc Neurosci* 34(47):15621–15630.
25. Scholl J, et al. (2015) The Good, the Bad, and the Irrelevant: Neural Mechanisms of Learning Real and Hypothetical Rewards and Effort. *J Neurosci Off J Soc Neurosci* 35(32):11233–11251.
26. Elert E (2014) Aetiology: Searching for schizophrenia’s roots. *Nature* 508(7494):S2-3.
27. Marin RS (1991) Apathy: a neuropsychiatric syndrome. *J Neuropsychiatry Clin Neurosci* 3(3):243–254.
28. Green MF, Horan WP, Barch DM, Gold JM (2015) Effort-Based Decision Making: A Novel Approach for Assessing Motivation in Schizophrenia. *Schizophr Bull*:sbv071.
29. Hartmann MN, et al. (2015) Apathy but not diminished expression in schizophrenia is associated with discounting of monetary rewards by physical effort. *Schizophr Bull* 41(2):503–512.
30. Salamone JD, et al. (2016) The pharmacology of effort-related choice behavior: Dopamine, depression, and individual differences. *Behav Processes*. doi:10.1016/j.beproc.2016.02.008.

31. Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, eds Black AH, Prokasy WF (Appleton-Century-Crofts, New York), pp 64–99.
32. Sutton RS, Barto AG (1998) *Reinforcement learning: An introduction* (MIT Press).
33. Enomoto K, et al. (2011) Dopamine neurons learn to encode the long-term value of multiple future rewards. *Proc Natl Acad Sci U S A* 108(37):15462–15467.
34. Kishida KT, et al. (2016) Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. *Proc Natl Acad Sci U S A* 113(1):200–205.
35. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69(6):1204–1215.
36. Chowdhury R, et al. (2013) Dopamine restores reward prediction errors in old age. *Nat Neurosci* 16(5):648–653.
37. Hauser TU, et al. (2015) Temporally Dissociable Contributions of Human Medial Prefrontal Subregions to Reward-Guided Learning. *J Neurosci Off J Soc Neurosci* 35(32):11209–11220.
38. Bishop C (2007) *Pattern Recognition and Machine Learning* (Springer, New York).
39. Klein-Flügge MC, Kennerley SW, Saraiva AC, Penny WD, Bestmann S (2015) Behavioral modeling of human choices reveals dissociable effects of physical effort and temporal delay on reward devaluation. *PLoS Comput Biol* 11(3):e1004116.
40. Klein-Flügge MC, Kennerley SW, Friston K, Bestmann S (2016) Neural Signatures of Value Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward Trade-off. *J Neurosci Off J Soc Neurosci* 36(39):10002–10015.
41. Hartmann MN, Hager OM, Tobler PN, Kaiser S (2013) Parabolic discounting of monetary rewards by physical effort. *Behav Processes* 100:192–196.
42. Rutledge RB, Dean M, Caplin A, Glimcher PW (2010) Testing the reward prediction error hypothesis with an axiomatic model. *J Neurosci Off J Soc Neurosci* 30(40):13525–13536.
43. Hauser TU, et al. (2014) The Feedback-Related Negativity (FRN) revisited: New insights into the localization, meaning and network organization. *NeuroImage* 84:159–168.
44. Seeley WW, et al. (2007) Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci Off J Soc Neurosci* 27(9):2349–2356.
45. Düzel E, et al. (2009) Functional imaging of the human dopaminergic midbrain. *Trends Neurosci* 32(6):321–328.

46. Björklund A, Dunnett SB (2007) Dopamine neuron systems in the brain: an update. *Trends Neurosci* 30(5):194–202.
47. Williams SM, Goldman-Rakic PS (1998) Widespread origin of the primate mesofrontal dopamine system. *Cereb Cortex N Y N 1991* 8(4):321–345.
48. Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307(5715):1642–1645.
49. Varazzani C, San-Galli A, Gilardeau S, Bouret S (2015) Noradrenaline and dopamine neurons in the reward/effort trade-off: a direct electrophysiological comparison in behaving monkeys. *J Neurosci Off J Soc Neurosci* 35(20):7866–7877.
50. Pasquereau B, Turner RS (2013) Limited encoding of effort by dopamine neurons in a cost-benefit trade-off task. *J Neurosci Off J Soc Neurosci* 33(19):8288–8300.
51. Carr DB, Sesack SR (2000) Projections from the Rat Prefrontal Cortex to the Ventral Tegmental Area: Target Specificity in the Synaptic Associations with Mesoaccumbens and Mesocortical Neurons. *J Neurosci* 20(10):3864–3873.
52. Haber SN, Behrens TEJ (2014) The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders. *Neuron* 83(5):1019–1039.
53. Friston KJ, Harrison L, Penny W (2003) Dynamic causal modelling. *NeuroImage* 19(4):1273–1302.
54. Parker NF, et al. (2016) Reward and choice encoding in terminals of midbrain dopamine neurons depends on striatal target. *Nat Neurosci* advance online publication. doi:10.1038/nn.4287.
55. Haber SN, Knutson B (2010) The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacol Off Publ Am Coll Neuropsychopharmacol* 35(1):4–26.
56. Krebs RM, Heipertz D, Schuetze H, Duzel E (2011) Novelty increases the mesolimbic functional connectivity of the substantia nigra/ventral tegmental area (SN/VTA) during reward anticipation: Evidence from high-resolution fMRI. *NeuroImage* 58(2):647–655.
57. Chowdhury R, Lambert C, Dolan RJ, Düzel E (2013) Parcellation of the human substantia nigra based on anatomical connectivity to the striatum. *NeuroImage* 81:191–198.
58. Mechelli A, Friston KJ, Frackowiak RS, Price CJ (2005) Structural covariance in the human cortex. *J Neurosci Off J Soc Neurosci* 25(36):8303–8310.
59. Alexander-Bloch A, Giedd JN, Bullmore E (2013) Imaging structural co-variance between human brain regions. *Nat Rev Neurosci* 14(5):322–336.
60. Laruelle M (2013) The second revision of the dopamine theory of schizophrenia: implications for treatment and drug development. *Biol Psychiatry* 74(2):80–81.

61. Laruelle M (2014) Schizophrenia: from dopaminergic to glutamatergic interventions. *Curr Opin Pharmacol* 14:97–102.
62. Howes OD, et al. (2011) Dopamine synthesis capacity before onset of psychosis: a prospective [18F]-DOPA PET imaging study. *Am J Psychiatry* 168(12):1311–1317.
63. Howes OD, et al. (2012) The nature of dopamine dysfunction in schizophrenia and what this means for treatment. *Arch Gen Psychiatry* 69(8):776–786.
64. Slifstein M, et al. (2015) Deficits in Prefrontal Cortical and Extrastriatal Dopamine Release in Schizophrenia: A Positron Emission Tomographic Functional Magnetic Resonance Imaging Study. *JAMA Psychiatry* 72(4):316–324.
65. Marin RS, Biedrzycki RC, Firinciogullari S (1991) Reliability and validity of the Apathy Evaluation Scale. *Psychiatry Res* 38(2):143–162.
66. Phillips MR (2009) Is distress a symptom of mental disorders, a marker of impairment, both or neither? *World Psychiatry Off J World Psychiatr Assoc WPA* 8(2):91–92.
67. D’Ardenne K, Lohrenz T, Bartley KA, Montague PR (2013) Computational heterogeneity in the human mesencephalic dopamine system. *Cogn Affect Behav Neurosci* 13(4):747–756.
68. Pauli WM, et al. (2015) Distinct Contributions of Ventromedial and Dorsolateral Subregions of the Human Substantia Nigra to Appetitive and Aversive Learning. *J Neurosci Off J Soc Neurosci* 35(42):14220–14233.
69. D’Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* 319(5867):1264–1267.
70. O’Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38(2):329–337.
71. Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* 68(5):815–834.
72. Hamid AA, et al. (2016) Mesolimbic dopamine signals the value of work. *Nat Neurosci* 19(1):117–126.
73. Rigoli F, Chew B, Dayan P, Dolan RJ (2016) The Dopaminergic Midbrain Mediates an Effect of Average Reward on Pavlovian Vigor. *J Cogn Neurosci*:1–15.
74. Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology (Berl)* 191(3):507–520.
75. Parvizi J, Rangarajan V, Shirer WR, Desai N, Greicius MD (2013) The will to persevere induced by electrical stimulation of the human cingulate gyrus. *Neuron* 80(6):1359–1367.
76. Rigoli F, Friston KJ, Dolan RJ (2016) Neural processes mediating contextual influences on human choice behaviour. *Nat Commun* 7:12416.

77. Rutledge RB, Skandali N, Dayan P, Dolan RJ (2014) A computational and neural model of momentary subjective well-being. *Proc Natl Acad Sci* 111(33):12252–12257.
78. Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. *NeuroImage* 33(2):493–504.
79. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *NeuroImage* 38(1):95–113.
80. Pine A, Shiner T, Seymour B, Dolan RJ (2010) Dopamine, time, and impulsivity in humans. *J Neurosci Off J Soc Neurosci* 30(26):8888–8896.
81. Voon V, et al. (2010) Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* 65(1):135–142.
82. Seymour B, Daw ND, Roiser JP, Dayan P, Dolan RJ (2012) Serotonin Selectively Modulates Reward Value in Human Decision-Making. *J Neurosci* 32(17):5833–5842.
83. Hauser TU, Iannaccone R, Walitza S, Brandeis D, Brem S (2015) Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *NeuroImage* 104:347–354.
84. Eldar E, Hauser TU, Dayan P, Dolan RJ (2016) Striatal structure and function predict individual biases in learning to avoid pain. *Proc Natl Acad Sci U S A* 113(17):4812–4817.
85. Hutton C, et al. (2011) The impact of physiological noise correction on fMRI at 7 T. *NeuroImage* 57(1):101–112.
86. Glover GH, Li T-Q, Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med* 44(1):162–167.
87. Kasper L, et al. (2009) Cardiac Artefact Correction for Human Brainstem fMRI at 7 Tesla. *Proc Org Hum Brain Mapp*.
88. Kasper L, et al. (2017) The PhysIO Toolbox for Modeling Physiological Noise in fMRI Data. *J Neurosci Methods*. doi:10.1016/j.jneumeth.2016.10.019.
89. Holmes AP, Friston K (1998) Generalisability, random effects and population inference. *NeuroImage* 7:S754.
90. Eklund A, Nichols TE, Knutsson H (2016) Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc Natl Acad Sci* 113(28):7900–7905.
91. Flandin G, Friston KJ (2016) Analysis of family-wise error rates in statistical parametric mapping using random field theory. *ArXiv160608199 Stat*. Available at: <http://arxiv.org/abs/1606.08199> [Accessed May 31, 2017].
92. Schwartenbeck P, FitzGerald THB, Mathys C, Dolan R, Friston K (2014) The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. *Cereb Cortex N Y N 1991*. doi:10.1093/cercor/bhu159.

93. Koster R, Guitart-Masip M, Dolan RJ, Düzel E (2015) Basal Ganglia Activity Mirrors a Benefit of Action and Reward on Long-Lasting Event Memory. *Cereb Cortex N Y N 1991* 25(12):4908–4917.
94. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *NeuroImage* 46(4):1004–1017.
95. Friston KJ, Dolan RJ (2010) Computational and dynamic models in neuroimaging. *NeuroImage* 52(3):752–765.
96. den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. *Cereb Cortex N Y N 1991* 19(5):1175–1185.
97. den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. *J Neurosci Off J Soc Neurosci* 30(9):3210–3219.
98. Hauser TU, et al. (2017) Increased decision thresholds enhance information gathering performance in juvenile Obsessive-Compulsive Disorder (OCD). *PLOS Comput Biol* 13(4):e1005440.
99. Weiskopf N, et al. (2013) Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. *Front Neurosci* 7. doi:10.3389/fnins.2013.00095.
100. Helms G, Draganski B, Frackowiak R, Ashburner J, Weiskopf N (2009) Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. *NeuroImage* 47(1):194–198.
101. Callaghan MF, et al. (2014) Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. *Neurobiol Aging* 35(8):1862–1872.
102. Helms G, Dathe H, Dechent P (2008) Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. *Magn Reson Med* 59(3):667–672.
103. Helms G, Dathe H, Kallenberg K, Dechent P (2008) High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. *Magn Reson Med* 60(6):1396–1407.
104. Ashburner J, Friston KJ (2000) Voxel-based morphometry--the methods. *NeuroImage* 11(6 Pt 1):805–821.
105. Rorden C, Brett M (2000) Stereotaxic display of brain lesions. *Behav Neurol* 12(4):191–200.
106. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS (2008) Associative learning of social value. *Nature* 456(7219):245–249.
107. Chater N, Brown GD (1999) Scale-invariance as a unifying psychological principle. *Cognition* 69(3):B17-24.

108. Sutton RS (1988) Learning to predict by the methods of temporal differences. *Mach Learn* 3(1):9–44.
109. Marsh B, Kacelnik A (2002) Framing effects and risky decisions in starlings. *Proc Natl Acad Sci U S A* 99(5):3352–3355.
110. Pompilio L, Kacelnik A, Behmer ST (2006) State-dependent learned valuation drives choice in an invertebrate. *Science* 311(5767):1613–1615.
111. Goldberg DE (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley Professional, Reading, Mass). 1 edition.
112. Schwarz G (1978) Estimating the Dimension of a Model. *Ann Stat* 6(2):461–464.
113. Matsumoto M, Matsumoto K, Abe H, Tanaka K (2007) Medial prefrontal cell activity signaling prediction errors of action values. *Nat Neurosci* 10(5):647–656.
114. Kennerley SW, Behrens TEJ, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat Neurosci* 14(12):1581–1589.
115. Hauser TU, et al. (2014) Role of the Medial Prefrontal Cortex in Impaired Decision Making in Juvenile Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry*.
116. Iannaccone R, et al. (2015) Conflict monitoring and error processing: New insights from simultaneous EEG–fMRI. *NeuroImage*. doi:10.1016/j.neuroimage.2014.10.028.

Supplementary Information

Supplementary Task Details

We calculated the median effort employed during the red frame phase as the effort exerted at that trial. The individual effort trajectories were monitored online to ensure that the subjects kept the force approximately constant during the whole effort execution phase (feedback was given after each block if necessary). To maintain the subjects' motivation and engagement, the points gained were converted and added to a blue bar shown at the bottom of the screen. Every time the bar reached the yellow target line, subjects received £1.00, and the bar started over from the left side of the screen, similar as implemented in previous studies (e.g., 106). Each point that the subjects won translated into an approximate 2% increase in the bar. The subjects earned $£4.59 \pm 0.82$ on average. To de-correlate outcome success (effort above/below threshold) from actual amount won, subjects received 0 points in 50% of all trials (probabilistically determined, cf. Fig. 1B). A jittered fixation cross (mean 6000ms, range 2000-10000ms, uniform distribution) was shown between two trials. Each of the two stimuli was presented 80 times, equally distributed across the 4 sessions of approximately 15 minutes each. Timing of task events were determined by maximizing the design efficacy for effort and reward PEs.

To measure effort, we used a bespoke, MR-compatible, pneumatic force gripper for the right hand. Air pressure was converted to digital signals using a National Instruments data converter (NI-6009, National Instruments Corporation Ltd., Newbury, UK) using a sampling rate of 1000Hz. During effort execution, the position of the thermometer was updated every 20ms. The effort feedback, as well as the threshold, was shown relative to the maximal force that the subject was able to execute. The calibration of maximum force was done at the beginning of the experiment. To ensure subjects were not deliberately squeezing below their maximum force, we determined the maximum executed force during gripper practice and replaced the maximal force if necessary. To account for slow drifts in baseline pressure due of a change in temperature that affects air expansion in the pipes, we adjusted the baseline pressure at the beginning of every scanning session. Pilot studies showed this is sufficient to account for these slow temperature drifts. It is well known that many modalities are perceived in a logarithmic rather than in a linear fashion (107). We thus used log-converted force measures for display and the determination of the force thresholds. Pilot investigations confirmed that log-transformed force feedback felt more natural and was easier to handle than non-converted feedback. Pilot studies revealed that an effort execution phase of 5000ms was feasible for males who, unlike females, were able maintain force levels across the whole experiment. Because pilot studies showed a high variability among female subjects, we decided to only test males in this study. We decided to modulate effort threshold based on the force and to keep time of execution constant, because effort is otherwise confounded by the time that one spends on the task, and thus confounds effort execution with temporal discounting (14).

Before entering the MRI, subjects were trained on the task. In a first phase, subjects were familiarized with the force grippers and learned how to control the 'thermometer'. This task was repeated in the scanner during the survey scans so that the subjects could adjust to the new environment. During the task instructions, subjects were told about the changing force thresholds and reward magnitudes. During two practice runs, subjects learned how thresholds and points change and how they can adjust their force accordingly.

The task was delivered using the Cogent toolbox for Matlab (R2010, MathWorks, Natick MA, USA). Pulse, breathing and eye-tracking was recorded to monitor the subjects' states and was used for artefact correction.

Computational Modelling

We developed a novel computational reinforcement learning model (32) predicting executed effort at each trial to understand the processes underlying the effort and reward learning in this task. The model consists of three distinct parts (Fig. S1): effort learning module, reward learning module, and a reward-effort discounting and subjective utility module.

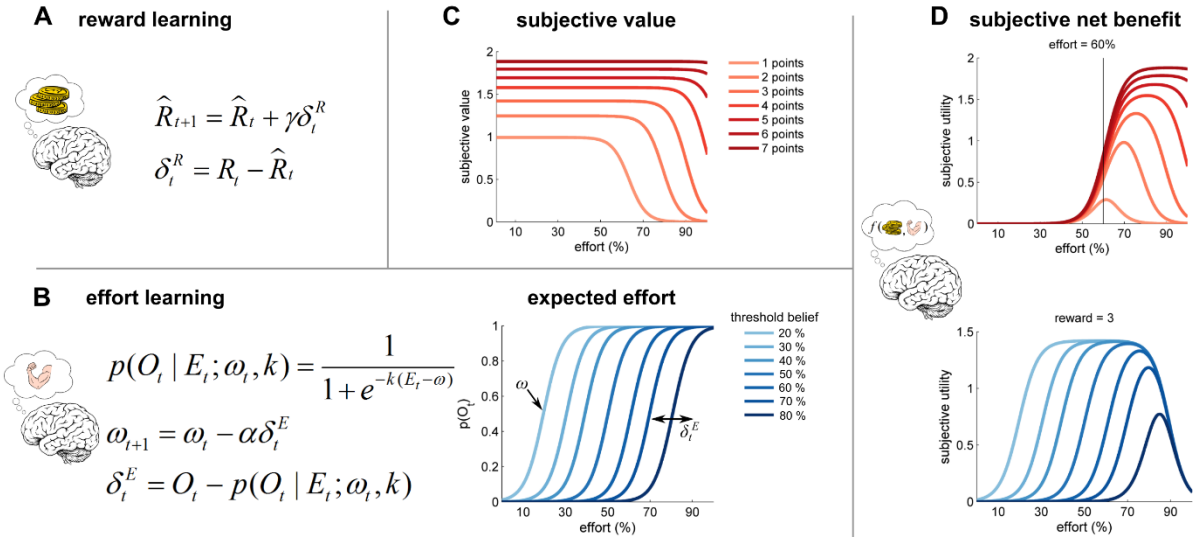


Figure S1. Computational framework for effort and reward learning. (A) Expected rewards R_t are learned using a reward PE δ_t^R and a learning rate γ . (B) Learning about effort resembles an iterative logistic regression of beliefs about being successful given the exerted effort $p(O_t | E_t)$ (blue lines depict different beliefs about the effort threshold). This belief is then updated on each trial using an effort PE δ_t^E that adjusts the indifference point of the belief ω_t . The free parameters α and k depict a learning rate and the precision of beliefs, respectively. (C) Effort discounting of reward magnitude follows a sigmoidal function so that rewards are more strongly discounted with higher effort. Model comparison reveals that reward magnitude not only affects the height of the sigmoid but also its indifference point, which means that lower reward are discounted already at lower effort levels, whereas high rewards are discounted only with the highest effort (red lines depict how different reward magnitudes are discounted as a function of effort). (D) Based on effort and reward expectation as well as reward discounting, we can compute the subjective net benefit for each trial. The upper panel depicts how subjective net benefit changes as a function of expected reward, given a fixed belief that the effort requires is at 60%. This shows how for 1 point, the maximal benefit is just above 60%, whereas for high rewards, it is well above 60% to ensure that the points are won, accounting for the uncertainty of the effort belief. The lower panel shows how subjective net benefit changes for a reward magnitude of 3 as a function of different beliefs about the height of the effort threshold (blue lines). An example trajectory of maximal subjective net benefit can be found in Fig. 1B (pink line).

Effort learning

In the task subjects are not explicitly told how much effort is needed to succeed in a given trial, but they had to learn it based on their prior experience with a given stimulus. Because the subjects do not receive explicit feedback about the height of the effort threshold, it is not possible to calculate the effort prediction error in a simple Rescorla-Wagner-like (31) fashion, i.e. as the difference between expected and received effort (cf. eq. (1.5)). Rather, the subjects will update their belief about the probability of succeeding $p(O_t | E_t)$ on every trial t . To do so, we used a modified version of an iteratively reweighted least squares logistic regression (IRLS, Fig. S1B; , 38). Here, the belief of succeeding is computed using a sigmoidal transformation of the executed effort at trial t E_t

$$p(O_t | E_t; \omega_t, k) = \frac{1}{1 + e^{-k(E_t - \omega)}} , \quad (1.1)$$

where ω_t describes the indifference point of the sigmoid, which is equivalent to the current belief about the height of the effort threshold. Parameter k describes the uncertainty about the height and is used as a free parameter. E_t is the median log-force that was exerted between 0 (no force exerted) and 100 (individual maximum force, determined during practice). To account for additional perceptual uncertainty about the effort actually exerted (because the median over 4 seconds was used, we do not assume perfect knowledge), we converted E_t into a Gaussian distribution with a standard deviation of 5 (arbitrarily chosen) around the mean of the actually exerted force using sampling from a normal distribution (100 samples). For outcome, the actual outcome was used assuming no uncertainty about the visual feedback).

On every trial, ω_t is updated using an effort prediction error δ_t^E and a learning rate α (free parameter):

$$\omega_{t+1} = \omega_t - \alpha \delta_t^E \quad (1.2)$$

The effort prediction error δ_t^E is the difference between the success at trial t , O_t , and the prior belief given the exerted force $p(O_t | E_t; \omega_t, k)$:

$$\delta_t^E = O_t - p(O_t | E_t; \omega_t, k) \quad (1.3)$$

This learning rule can be seen as a simplified version on a temporal-difference based predictive learner (108), only that we did not incorporate the gradient of our prediction (cf eq. 2 in (108)) because our trajectories were relatively smooth and thus a gradient would have little effect (i.e. act as a scaling factor, which is now absorbed in learning rate α). Moreover, it is difficult to imagine how a gradient was effectively implemented in a biological system such as the human brain.

Reward learning

To learn about the number of points that are associated with a stimulus, we used a standard Rescorla-Wagner learning model (31) (Fig. S1A), where the expected reward magnitude R_t is being updated using a reward prediction error δ_t^R and a learning rate γ (free parameter):

$$R_{t+1} = R_t + \gamma \delta_t^R \quad (1.4)$$

The reward prediction error is the difference between the expected R_t and number of points that was presented R_t :

$$\delta_t^R = R_t - R_t \quad (1.5)$$

In this study, we used the current number shown on the screen as R_t , irrespective of whether the subject exerted enough effort to surpass the effort threshold. We decided to do so, because subjects learn about the reward magnitude irrespective of their effort success, and thus update their reward expectations. Moreover, model comparison (see main text) revealed that subjects also learned about probabilistic (50%) 0-outcomes, rather than ignoring the magnitude at that trial.

Reward-effort discounting and subjective net benefit

It is well known that effort and other costs discount rewards (9) and recent studies investigated the effort-discounting function in great detail (29, 39, 41). Here, we compared previously suggested functions (hyperbolic, quadratic, sigmoidal) and extended these. For all models, we introduced a utility parameter τ that accounts for non-linearities in the subjective representation of reward (1, 39, 109, 110) and this improves model fit (Fig. S2).

The quadratic discounting calculated the subjective value of a reward $v(R|e)$ given an effort e , is based on Hartman et al.'s studies (29, 41), and has a free decay kernel parameter κ that depicts the discounting steepness:

$$v(R|e) = R^\tau - \kappa e^2 \quad (1.6)$$

The hyperbolic discounting was originally introduced as mirroring hyperbolic temporal discounting and is formalized with a discounting kernel κ :

$$v(R|e) = \frac{R^\tau}{1 + \kappa e} \quad (1.7)$$

The sigmoidal effort-discounting function is more flexible than the quadratic and hyperbolic functions and has two free parameters: indifference point p and the slope κ . Similar to Klein-Függe et al. (39), we extended the sigmoid by two terms: subtracting $\frac{1}{1 + e^{\kappa p}}$ to ensure that $v(R) = R^\tau$ when effort e equals 0 (i.e., no

discounting). By multiplying $\left(1 + \frac{1}{e^{\kappa p}}\right)$, it is ensured that the subjective values $v(R)$ will not become negative for high effort, in keeping with previous modelling of effort discounting (39).

$$v(R | e) = R^\tau \left(1 - \left(\frac{1}{1 + e^{-\kappa(e-p)}} - \frac{1}{1 + e^{\kappa p}} \right) \left(1 + \frac{1}{e^{\kappa p}} \right) \right) \quad (1.8)$$

In this sigmoidal discounting, the reward only affects the height of the sigmoid, but the indifference point p is unaffected by the reward. This means that the acceleration of discounting always occurs at the same effort level. However, it might be that with high rewards, discounting only emerges at very high effort, whereas low rewards discount at low effort (cf. horizontal shifts in Fig. S1). We thus implemented an additional sigmoidal function where height and indifference point are modulated by expected rewards:

$$v(R | e) = R^\tau \left(1 - \left(\frac{1}{1 + e^{-\kappa(e-pR^\tau)}} - \frac{1}{1 + e^{\kappa pR^\tau}} \right) \left(1 + \frac{1}{e^{\kappa pR^\tau}} \right) \right) \quad (1.9)$$

On each trial, we assume that subjects exert the effort that has the highest subjective net benefit for them by combining the beliefs about the effort threshold and the belief about the current reward. The net benefit for every effort e (0-100%, Fig. S1) is calculated as the product of the belief about the effort threshold and the subjective value of that effort, and the normalized probability density at the chosen effort was the used for calculating the model fit (log likelihood):

$$B(e) = p(O_t | e; \omega, k) v(R_t | e) \quad (1.10)$$

Alternatively, one could additionally account for the utility of not succeeding (i.e. expending effort without receiving a reward). Because the probability of not succeeding in this task is the inverse of the success-probability, such a model results in almost identical results as simulations revealed. We thus decided to compute utility as in eq. (1.10), similar to previous studies (39, 41).

Model parameters were estimated by maximizing the probabilistic benefit function, independently for each subject, using a genetic algorithm (111), and model comparison was done using fixed- and random-effects analysis for BIC (112).

Please note that the goal of our modelling was to derive a model that most adequately reflects the learning signals, and we did not seek to design it in order to maximize orthogonality between free parameters. It is thus possible that some of the parameters express a considerable covariance, and we thus decided to not analyze the parameters in detail.

Model comparison and selection

We compared several potential models and the best fitting model was selected for further analysis (Fig. S2). First, we compared the model fit using the four different effort-discounting functions. The sigmoidal effort-discounting, where height and indifference point are modulated by reward outperformed the other models clearly (eq. (1.9)).

Further model comparisons revealed that by imposing no learning for either rewards or effort (by setting the expected reward/effort to the mean expectation across the task), the model fits were clearly worse. This confirms that subjects learned about rewards and effort, and that these learning processes are necessary for the model to explain the behavior.

In addition, we compared the effort PE model to a heuristic effort learner. This model adjusted its effort expectations based on the outcome (success/no success). However, in contrast to the effort PE model, it did so by merely adjusting its expectation by the same amount, stable across trials. This means that such a model ignores the size of a prediction error (eq. 1.2) and only uses its valence, and consequently always changes the effort expectation by the same amount.

We also compared our model to a model in which rewards are learned using (near) optimal inference. In this task, the best approximation of the current reward magnitude is achieved by taking the previously displayed

reward magnitude while ignoring the probabilistic 0 rewards. This optimal model performed better than the no-learning model, but worse than the PE-based learning model.

Moreover, a model ignoring stimulus identity performed worse than the winning model, thus supporting the notion that subjects take stimulus-identity into account ($\log L = -17320$, $BIC = 35523$).

Lastly, we found that a model without utility parameter τ performs worse and a reward learning algorithm that ignores the probabilistic 0 reward magnitudes also has slightly worse model fits.

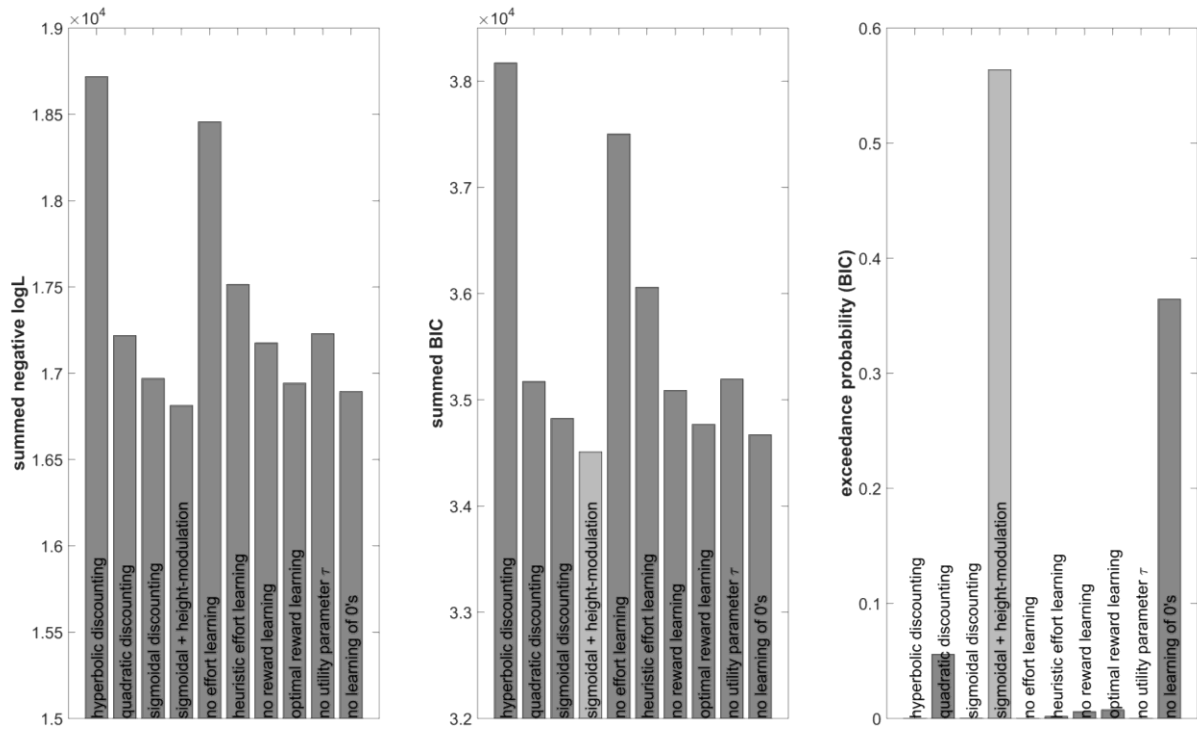


Figure S2. Model comparison. Model comparison revealed that a sigmoidal effort-discounting function, where height and indifference point are modulated by reward ('sigmoidal + height-modulation') performs best. Models that do not employ PE-based learning, such as a no-learning model, use heuristics or optimal inference model perform worse. Both, a utility parameter τ and the learning of 0-reward trials turned out to improve model fits. The model winning model comparison (using BIC) in a fixed-effects (middle panel) and a random effects analysis (right panel; , 94) is highlighted in bright gray.

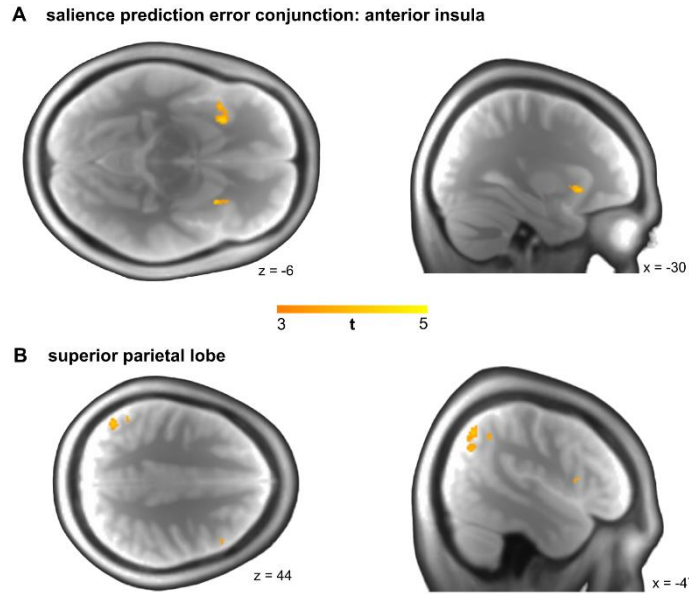


Figure S3. Salience PEs across domains. Whole-brain conjunction analysis ($p < .05$, cluster-extent FDR correction, height threshold $p < .001$) reveals that unsigned, salience PEs of both domains activate a common network including the left anterior insula (A; MNI: -29, 20 -6, cluster 106 voxels, peak $t=4.48$) and intraparietal sulcus (B; MNI: -47 -62 44, cluster 179 voxel, peak $t=4.35$).

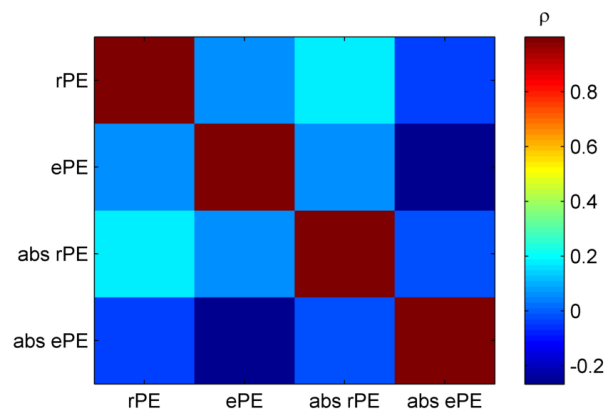


Figure S4. Correlation between fMRI regressors. To control for potential collinearities between our regressors of interest, we disabled the orthogonalization in our analysis, letting the regressors compete for variance (37). Because we independently varied effort and reward trajectories, we were able to achieve very low correlations between our regressors. Color bar: average Pearson correlation coefficient; rPE: reward prediction error (PE); ePE: effort PE; abs rPE: absolute, salience rPE; abs ePE: salience ePE.

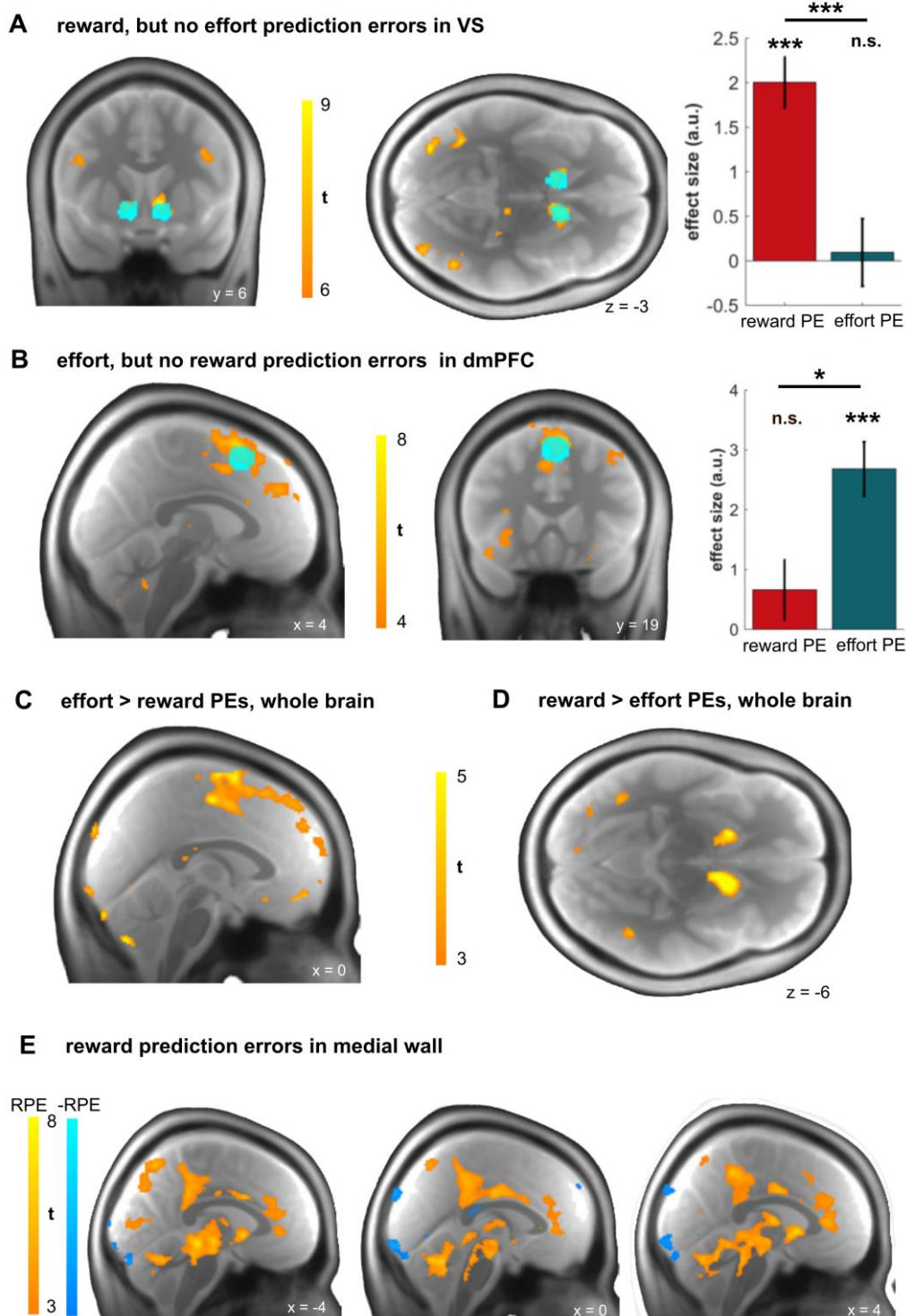


Figure S5. Double dissociation of reward and effort predictions error in cortex and striatum. The double-dissociation between effort and reward PEs was also evident in literature-based regions-of-interest of (A) the ventral striatum (VS-ROI derived from <http://www.neurosynth.org/>; reward PE: $t(27)=6.93$, $p<.001$; effort PE: $t(27)=.25$, $p=.807$; comparison: $t(27)=3.74$, $p<.001$), and (B) the dmPFC (ROI derived from (25): MNI: 2 18 54, 10mm sphere; effort PE: $t(27)=5.92$, $p<.001$; reward PE: $t(27)=1.31$, $p=.202$; comparison: $t(27)=-2.70$, $p=.012$). Light blue indicates ROIs overlaid over main effects as shown in Fig. 2. Whole-brain comparison between reward and effort prediction errors confirmed a double-dissociation in dmPFC (C) and VS (D) on a whole-brain corrected level. (E) Our findings do not contradict previous findings showing reward PEs in the medial wall (37, 113–115), as we also found reward PEs, but in distinct, more ventral areas of the medial wall (RPE: reward prediction error).

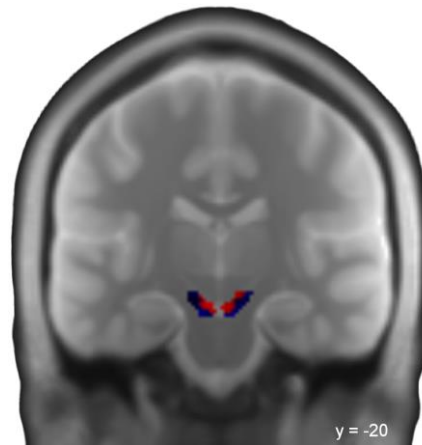


Figure S6. Coronal view on reward and effort PEs in SN/VTA. The spatial gradients for effort and reward PE distributions confirms that reward PEs (red) are primarily processed in dorsomedial regions, whereas effort PE (blue) are represented in ventrolateral areas. Also see Figures 3 & 4.

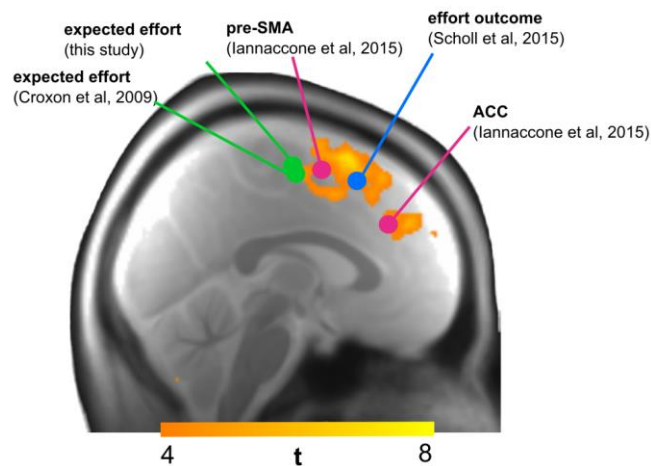


Figure S7. Effort representation along the medial wall. Effort PE activation (warm colors, as in Fig. 2D) spans dmPFC between pre-SMA and ACC (anatomical regions in pink derived from Iannaccone et al., 2015, (116)). The effort PE activation lies in close proximity with a previous finding (25) of effort outcomes (blue). Interestingly, the effort PEs lie anterior to activations for effort expectation (green). Both, a previous (2) and our study (effort expectation during anticipation: $p < .05$ cluster-extend FWE, MNI: [-15 -12 62], $t = 4.88$, cluster size=494) find effort expectation signals in SMA. This suggests that effort evaluation is processed more anteriorly to effort expectation. All activations are projected to the same sagittal plane (x-axis) for visualization purposes.

Table S1. Effort and reward prediction error activation. Main effect of effort and reward prediction error. Areas shown that are significant at $p < .05$, height-FWE corrected for multiple comparisons, $k > 30$, unless stated otherwise. ACC: anterior cingulate cortex; dmPFC: dorsomedial prefrontal cortex; IPS: intraparietal sulcus; ITL: inferior temporal lobe; n.s.: no significance; pre-SMA: pre-supplementary motor area; SN/VTA: substantia nigra/ventral tegmental area; VS: ventral striatum

Contrast	Region	Hemisphere	Cluster size (voxels)	x	y	z	T Score		
effort prediction error	temporal pole	right	39	48	8	-18	8.39		
	dmPFC (pre-SMA/ACC)	bilateral	47	-2	15	62	7.45		
	SN/VTA	bilateral	6	-3	-14	-15	4.34*		
-effort prediction error	n.s.								
reward prediction error	visual cortex	left	228	-38	-75	-8	11.83		
			818	-29	-84	17	10.51		
			61	-39	-72	14	7.91		
		right	73	38	-83	11	8.28		
			56	38	-84	-2	8.06		
			IPS	right	1764	30	-48	48	11.14
	VS	right	32	51	-33	53	7.58		
			323	-23	-69	45	8.50		
			319	11	11	0	10.30		
		left	81	-14	8	-3	8.29		
			ITL	right	336	41	-59	-11	9.81
			left	82	-38	-57	-5	9.49	
SN/VTA	left	109	-9	-26	-12	5.63**			
	right	104	12	-26	-11	5.07***			
-reward prediction error	n.s.								

* $p = .011$, small-volume FWE-corrected for anatomical SN/VTA mask
** $p = .001$, small-volume FWE-corrected for anatomical SN/VTA mask
*** $p = .002$, small-volume FWE-corrected for anatomical SN/VTA mask