Separate mesocortical and mesolimbic pathways encode effort and reward learning signals

Tobias U. Hauser\textsuperscript{a,b,1}, Eran Eldar\textsuperscript{a,b}, and Raymond J. Dolan\textsuperscript{a,b}

\textsuperscript{a}Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom; and \textsuperscript{b}Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London WC1B SEH, United Kingdom

Edited by Read Montague, Virginia Tech Carilion Research Institute, Department of Physics, Virginia Tech, Roanoke, VA, and accepted by Editorial Board Member Michael S. Gazzaniga July 17, 2017 (received for review April 5, 2017)

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. 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.

Organisms need to make energy-efficient decisions to maximize benefits and minimize costs, a tradeoff exemplified in effort expenditure (1–3). A key example occurs 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 clearly evident, although the effort necessary to attain success is often less transparent. Only by repeatedly experiencing success and failure is it 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 (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 engenders a preference for lower-effort choices (15, 20–23), while receiving effort feedback elicits 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 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 learned simultaneous reward and effort contingencies in an ecologically realistic manner, while 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. L4 and SI Appendix). 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 the entire task) and a required effort threshold (% individual maximal force, determined during practice). These parameters were initially unknown to the subjects 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

Significance

Learning about multiple features of a choice option is crucial for optimal decision making. How such multattribute learning is realized remains unclear. Using functional MRI, we show that 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 ventral striatum. These findings highlight how the brain parses parallel learning demands.

Author contributions: T.U.H., E.E., and R.J.D. designed research; T.U.H. and E.E. performed research; T.U.H. analyzed data; and T.U.H., E.E., and R.J.D. wrote the paper.

The editor also holds a joint appointment with University College London.

This article is a PNAS Direct Submission. R.M. is a guest editor invited by the Editorial Board.

1To whom correspondence should be addressed. Email: t.hauser@ucl.ac.uk.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1705643114/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1705643114

PNAS | Published online August 14, 2017 | E7395–E7404
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 \( t(28) = 15.96, P < 0.001 \) 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 \( t(28) = 4.97, P < 0.001 \), consistent with a motivation to expend greater energy on high-value choices. Last, subjects exerted more effort following trials where they failed to exceed an effort threshold \( t(28) = 10.75, P < 0.001 \), 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; \( t(28) = 17.08, P < 0.001 \)] but also lessened their effort after successfully surpassing an effort threshold \( t(28) = -17.15, P < 0.001 \), 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.

A Computational Framework for Effort and Reward Learning. To probe deeper into 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 SI Appendix, Fig. S1 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. Note that 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 used 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 the entire experiment), which in turn did not influence a change in reward magnitude. Using model comparison (all model fits are shown in SI Appendix, 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 subject’s 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 that this does not describe a simple motor or force PE signal, given that a force PE would be evident during force execution 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 that 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 that a sigmoidal discounting function best described this arbitration (SI Appendix, 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 high-reward conditions, subjects are more likely to exert greater effort to ensure they surpass an effort threshold (compare SI Appendix, Fig. S1).

To assess whether subjects learned using a PE-like teaching signal, we compared our PE-based learning model with alternative formulations (SI Appendix, Fig. S2). A first comparison revealed that the PE learning model outperformed nonlearning 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 with an optimal reward learning model, which tracks the previous reward magnitude and ignores the probabilistic null outcomes, revealing that 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 \( \left( t_{28} = 13.50, P < 0.001 \right) \) and reward PEs \( \left( t_{28} = 2.10, P = 0.045 \right) \) 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: \( r = 0.056 \pm 0.074; \) SI Appendix, Fig. S4). Bilateral VS responded significantly to reward PEs \( [P < 0.05, \text{whole-brain family-wise error (FWE)} \text{ correction}; \text{see SI Appendix, Table S1 for all activations}] \) but not to effort PEs (Fig. 2 A–C). In contrast, dmPFC (peaking in pre-SMA).
extending into dACC) responded to effort PEs (P < 0.05, whole-brain FWE correction; Fig. 2 D–F and SI Appendix, Table S1) but not to reward PEs (Fig. 2 F). 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(28) = 2.82, P = 0.009] supports its behavioral relevance in this task and is consistent with updating a subject’s expectation about future effort requirements (33). Interestingly, dmPFC area processing effort PE peaks anterior to pre-SMA and lies anterior to where anticipatory effort signals are found in SMA (SI Appendix, 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: −0.32 ± 3.77; (27) ± 0.45; P = 0.657; VS: effect size: −0.41 ± 1.59; (27) ± 1.36; P = 0.185]. Post hoc analysis confirmed that both components of a prediction error, expectation and outcome, were represented in these two regions (Fig. 2 B and 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: 0.01 ± 0.18; t(27) = 0.21, P = 0.832]. No effects were found for negative-heading (inverse) PEs in either reward or effort conditions (e.g., increasing activation for decreasing reward PEs; SI Appendix, 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 derived from ref. 25), and again found a significant double dissociation in both VS (SI Appendix, Fig. S5A) and dmPFC (SI Appendix, Fig. S5B). Moreover, this double dissociation was also evident in a whole-brain comparison between effort and reward PEs [SI Appendix, Fig. S5C; dmPFC: Montreal Neurological Institute (MNI) coordinates: −14 −9 69, t = 6.05, P < 0.001 cluster-extent FWE, height-threshold P = 0.001; VS: MNI: 15 9 −6, t = 6.73, P < 0.001 cluster-extent FWE].

Additionally, we controlled for unsignaled (i.e., salience) effort and reward PE signals by including them as additional regressors in the same fMRI model (correlation matrix shown in SI Appendix, Fig. S4), as these signals are suggested to be represented in dmPFC (e.g., 43). Interestingly, when analyzing the unsignaled 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 (SI Appendix, Fig. S3).

**Simultaneous Representations of Effort and Reward PEs in 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 the midbrain structures substantia nigra and ventral tegmental area, 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 < 0.05, small-volume FWE correction for SN/VTA; SI Appendix, Table S1), consistent with a common dopaminergic midbrain origin for striatal and cortical PE representations.

**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 modeling; 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 with descending or mixed connections (Bayesian random-effects model selection: expected posterior probability 0.55, exceedance probability 0.976, Bayesian omnibus risk 2.834; 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 nonhuman primates points 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 and SI Appendix, Fig. S6; ventral–dorsal gradient: \( \beta = 0.151 \), 95% confidence interval (C.I.) = 0.281 to 0.022, \( P = 0.016 \); medial–lateral gradient: \( \beta = 0.469 \), 95% C.I. = 0.336 to 0.062, \( P < 0.001 \)) provided evidence that dorsomedial SN/VTA is more strongly affiliated with reward PE encoding, whereas the ventrolateral SN/VTA was more affiliated with 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 blood oxygen level-dependent (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: \( \beta = 0.220 \), 95% C.I. = 0.373 to 0.067, \( P = 0.002 \); medial–lateral: \( \beta = 0.466 \), 95% C.I. = 0.310 to 0.622, \( P < 0.001 \)). 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 the spatial relationship between SN/VTA, dmPFC, and VS, we investigated structural associations between these areas. We used structural covariance analysis (58), which investigates how gray matter (GM) densities covary between brain regions, and has been shown sensitive for identifying anatomically and functionally relevant networks (59). Specifically, we asked how well 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: \( r = 0.155 \), \( P = 0.430 \); VS: \( r = 0.100 \), \( P = 0.612 \); SN/ VTA: effort PEs: \( r = 0.067 \), \( P = 0.737 \); reward PEs: \( r = 0.079 \), \( P = 0.690 \)). We found both spatial gradients were significant (Fig. 4C; ventral–dorsal gradient: \( \beta = 0.012 \), 95% C.I. = 0.018 to 0.006, \( P < 0.001 \); medial–lateral: \( \beta = 0.007 \), 95% C.I. = 0.002 to 0.013, \( P = 0.007 \)), suggesting that SN/VTA GM was more strongly associated with dmPFC GM in ventrolateral and VS GM in dorsomedial areas, thus confirming the findings of our previous analyses.

**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 Scale (AES (65)), were better predicted by dmPFC or VS activation. We ran a fivefold cross-validated prediction of AES total score using mean functional responses in either dmPFC or VS (using the same functional ROIs as above), using effort and reward PE responses in both regions (correlation between predictors: dmPFC: \( r = 0.149 \), \( P = 0.458 \); VS: \( r = 0.144 \), \( P = 0.473 \)). We found that dmPFC activations were highly predictive of apathy scores (Fig. 5A; \( P < 0.001 \), using permutation tests; Materials and Methods). Interestingly, the effect sizes for both reward (0.573 ± 0.050) and effort (0.351 ± 0.059) prediction errors in dmPFC showed a positive association with apathy, meaning that the bigger a prediction error in dmPFC, the more apathetic a person was. Activity in VS did not predict apathy (Fig. 5B; \( P = 0.796 \)). This was also reflected by a finding that extending a dmPFC-prediction model with VS activation did not improve apathy prediction (\( P = 0.394 \)). There was no association between dmPFC (\( r = 0.225 \), \( P = 0.258 \)) or VS (\( r = 0.142 \), \( P = 0.481 \)) GM and apathy. Furthermore, we found no link between overt behavioral variables and apathy (money earned: \( r = 0.02 \), \( P = 0.927 \); mean exerted effort: \( r = 0.00 \), \( P = 0.99 \); SD exerted effort: \( r = 0.24 \), \( P = 0.235 \); N trials not succeeding effort threshold: \( r = 0.13 \), \( P = 0.525 \)). 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: \( 1.23 ± 3.43 \);
incorporating both effort and reward (2, 66). We examined whether VS encodes an overall net benefit or integrated utility signal by examining the interaction between effort and reward, and thus extend previous single-attribute learning studies for reward PE (13). Instead, the sigmoidal function of our reward–effort arbitration model suggests that effort influences subjective value through an interaction with reward. This predicts that during low-effort trials, reward has little impact on subjective value, whereas for high-effort trials 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 = 0.007]. 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 = 0.003] but not in low-effort trials [t(27) = 0.61, P = 0.547].

Discussed

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 extend previous single-attribute learning studies for reward PE (e.g., 12, 26, 42) and effort outcomes (24, 25).

Our study shows 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 toward 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 dorsomedial encoding of reward PEs, and a ventrolateral encoding of effort PEs, extends previous studies on SN/VTA subregions (56, 57, 67, 68) by demonstrating that this segregation has functional implications that are exploited during multivariate learning. In contrast to previous studies on SN/VTA structures (56, 67–69), we performed whole-brain imaging, which allowed us to investigate the precise interactions between 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 that 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). By contrast, the ventrolateral region expressing effort PE activity 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 that 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, although 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 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 before obtaining a reward. Nevertheless, the convergence of our model predictions and the interaction effect seen in VS supports 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, although 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 suggests that 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 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 reflects 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 that 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 volunteers (age, 24.1 ± 4.5 y; range, 18 to 35 y) 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 (SI Appendix). 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 University College London (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 1,000 ms). The stimuli (pseudorandomized, no more than three 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 decorrelated, and indeed the realized correlation between effort and reward prediction errors was minimal. Moreover, independent trajectories for effort and reward allowed us to cover a wide range of effort and reward expectation combinations, enabling us to comprehensively 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, 4,000 ms; uniformly distributed between 2,000 and 6,000 ms), subjects had to squeeze a force gripper with their right hand for 5,000 ms. During the first 1,000 ms, the subjects increased their force to the desired level (as indicated by a horizontal thermometer; blue frame phase). During the last 4,000 ms, subjects maintained a constant force (red frame phase) and released as soon as the thermometer disappeared from the screen.

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

Hauser et al.
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 nonsuccess trials and then tested consistency using t tests across all subjects (Fig. 12).

Computational Modeling. 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 subjects’ behavior (executed effort at each trial), and a model comparison using the Bayesian information criterion was performed to select the best-fitting model (31 Appendix, Fig. S2). The preferred model was then used for the fMRI analysis. The models fitted model comparison are detailed in SI Appendix.

fMRI Data Acquisition and Preprocessing. MRI was acquired using a Siemens Trio 3Tesla 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 limbic brain regions that track reward and effort prediction errors. To this end, we used the BOLD response for every trial from the dmPFC and bilateral VS (mean parameter estimate for each voxel). The 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 modeled 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 ROI. We then compared the effects for each voxel for the prediction errors that follow a spatial gradient or not. 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 statistical significance using permutation tests (10,000 iterations; randomly permuting the spatial coordinates of each voxel). This metric allows us to measure whether a voxel was more predictive of effort or reward PEs. To ensure this analysis did not reflect the task effects, we regressed out the task effect (reward PE) before the main analysis. For the calculation, 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 or not. We obtained the 95% confidence intervals 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 parameter estimate for each voxel). The PEs were then used to perform our gradient analyses.

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 modeling (53). DCM allows the experimenter to specify, estimate, and compare biophysically plausible models of spatiotemporal distribution 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 VS, as well as the VS and dmPFC. This allowed us to test whether the SN/VTA modulated the connections originating from SN/VTA or targeting it. This approach (PEs affecting modulation of intrinsic connectivity) was used in previous studies investigating the effects of PEs on effective connectivity (e.g., 95, 96). We compared six 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 PE) and having one descending modulation, fitter the data better. DCMs were fitted for each subject and run separately, and Bayesian random-effects comparison (94) was used for model comparison.
We found similar effect when using beta weights (which do not take measurement uncertainty into account) instead of t values, and also if we included all voxels irrespective of whether they respond to any of the PE. 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 nonclinical 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 (65) (missing data from one subject). We used the total scores as a dependent variable in a fivefold cross-validated regression (cf. B4, 97). 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 < 0.05 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. 97). 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 with an extended model with VS activations as additional predictors. Permutation tests (by permuting the additional regressors) were again used to assess significance between dmPFC and the full model.

Structural MRI Data Acquisition and Analysis. Structural images were acquired using quantitative multiparameter maps (MPMs) in a 3D multiecho fast low-angle shot (FLASH) sequence with a resolution of 1-mm isotropic voxels (98). Magnetic transfer (MT) images were used for gray matter quantification, as they are particularly sensitive for subcortical regions (99). In total, three different FLASH sequences were acquired with different weightings: predominantly MT (TRi, 23.7 ms6; off-resonance Gaussian MT pulse of 4-ms duration preceded excitation, 2-KHz frequency offset, 220° nominal flip angle), proton density (23.7 ms6), and T1 weighting (18.7 ms20) (100). To increase the signal-to-noise ratio, we averaged signals of six equidistant bipolar gradient echo (TE, 2.2 to 14.7 ms). To calculate the semiquantitative MT maps, we used mean signal amplitudes and additional T1 maps (101), and additionally eliminated influences of B1 inhomogeneity and relaxation effects (102). To normalize functional and structural maps, we segmented the MT maps (using heavy bias regularization to account for the quantitative nature of MPMs), and generated flow fields using DARTEL (79) with the standard settings for SPM12. The flow fields 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 3 mm to correspond to any of the PEs.

To investigate anatomical links between SN/VTA and VS and dmPFC, we performed a (voxel)-based morphometry-based structural covariance analysis (88). The approach assumes that brain regions that are anatomically and functionally related (e.g., form a common network) should covary 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 and normalized gray matter MT maps and applied the Jacobian of the normalization step to preserve total tissue volume (103), 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 PVE < 0.05; 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 density in a linear regression model predicting the gray matter density in every voxel in 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 MIRCron (104). 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 at neurovault.org/collections/IAYMWZjY/. ACKNOWLEDGMENTS. 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 AI Reid for providing the force grippers. A Wellcome Trust Cambridge-UCL Mental Health and Neurosciences Network grant (095844/Z/11/Z) supported all authors. R.J.D. holds a Wellcome Trust Senior Investigator award (098362/Z/12/Z). The Max Planck UCL 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 (091532/Z/10/Z).
Separate mesocortical and mesolimbic pathways encode effort and reward learning signals

Tobias U. Hauser, Eran Eldar & Raymond J. Dolan

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., 1). Each point that the subjects won translated into an approximate 2% increase in the bar. The subjects earned £4.59±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 (2). 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 (3).

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 (4) 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 $\hat{R}_{t+1}$ are learned using a reward PE $\delta_t^R$ and a learning rate $\gamma$. (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 $\delta_t^E$ that adjusts the indifference point of the belief $\omega_t$. The free parameters $\alpha$ and $k$ depicts 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 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; 6). Here, the belief of succeeding is computed using a sigmoidal transformation of the executed effort at trial $t$ $E_t$

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

where $\omega_t$ 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, \( \omega_t \) is updated using an effort prediction error \( \delta^E_t \) and a learning rate \( \alpha \) (free parameter):

\[
\omega_{t+1} = \omega_t - \alpha \delta^E_t
\]  

(1.2)

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

\[
\delta^E_t = O_t - p(O_t \mid E_t; \omega_t, k)
\]  

(1.3)

This learning rule can be seen as a simplified version on a temporal-difference based predictive learner (7), only that we did not incorporate the gradient of our prediction (cf. eq. 2 in (7)) 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 \( \alpha \)). 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 (5) (Fig. S1A), where the expected reward magnitude \( R_t \) is being updated using a reward prediction error \( \delta^R_t \) and a learning rate \( \gamma \) (free parameter):

\[
\hat{R}_{t+1} = \hat{R}_t + \gamma \delta^R_t
\]  

(1.4)

The reward prediction error is the difference between the expected \( \hat{R}_t \) and number of points that was presented \( R_t \):

\[
\delta^R_t = R_t - \hat{R}_t
\]  

(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 (8) and recent studies investigated the effort-discounting function in great detail (9–11). Here, we compared previously suggested functions (hyperbolic, quadratic, sigmoidal) and extended these. For all models, we introduced a utility parameter \( \tau \) that accounts for non-linearities in the subjective representation of reward (9, 12–14) and this improves model fit (Fig. S2).

The quadratic discounting calculated the subjective value of a reward \( v(R \mid e) \) given an effort \( e \), is based on Hartman et al.’s studies (10, 11), and has a free decay kernel parameter \( \kappa \) that depicts the discounting steepness:

\[
v(R \mid e) = R^\tau - \kappa e^2
\]  

(1.6)

The hyperbolic discounting was originally introduced as mirroring hyperbolic temporal discounting and is formalized with a discounting kernel \( \kappa \):

\[
v(R \mid e) = \frac{R^\tau}{1 + \kappa e}
\]  

(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 \( \kappa \). Similar to Klein-Függe et al. (9), 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 \((1 + \frac{1}{e^{xp}})\), it is ensured that the subjective values \(v(R)\) will not become negative for high effort, in keeping with previous modelling of effort discounting (9).

\[
v(R | e) = R^T \left(1 - \frac{1}{1 + e^{-\kappa(e - p)}} - \frac{1}{1 + e^{xp}} \right) \left(1 + \frac{1}{e^{xp}} \right)
\]

(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:

\[
v(R | e) = R^T \left(1 - \frac{1}{1 + e^{-\kappa(e - pR')}} - \frac{1}{1 + e^{xpR'}} \right) \left(1 + \frac{1}{e^{xpR'}} \right)
\]

(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):

\[
B(e) = p(O_t | e; \omega_t, k) v(\hat{R}_t | e)
\]

(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 (9, 11).

Model parameters were estimated by maximizing the probabilistic benefit function, independently for each subject, using a genetic algorithm (15), and model comparison was done using fixed- and random-effects analysis for BIC (16).

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 (logL=-17320, BIC=35523).

Lastly, we found that a model without utility parameter $\tau$ 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 $\tau$ 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; , 17) 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 (18). 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 (19): 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 (18, 20–22), 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 confirm 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 representations 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, (23)). The effort PE activation lies in close proximity with a previous finding (19) of effort outcomes (blue). Interestingly, the effort PEs lie anterior to activations for effort expectation (green). Both, a previous (24) 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

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

* $p=.011$, 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
References


