bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hippocampal-midbrain circuit enhances the pleasure of anticipation in the prefrontal cortex

Kiyohito Iigaya *[1,2,3], Tobias U. Hauser[1,4], Zeb Kurth-Nelson[1,5], John P. O'Doherty[3], Peter Dayan[1,2,6,7], and Raymond J. Dolan[1,4,7]

[1]Max-Planck UCL Centre for Computational Psychiatry and Ageing Research, 10-12 Russell Square, London WC1B 5EH, United Kingdom
[2]Gatsby Computational Neuroscience Unit, University College London, 25 Howland Street, London, W1T 4JG, United Kingdom
[3]Division of Humanities and Social Sciences, California Institute of Technology, 1200 E California Blvd, Pasadena, CA 91125
[4]Wellcome Centre for Human Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom
[5]Deepmind, 6 Pancras Square, London N1C 4AG, United Kingdom
[6]Max Planck Institute for Biological Cybernetics, 72076 Tubingen, Germany
[7]Joint senior authors

March 25, 2019

Having something to look forward to is a keystone of well-being. Anticipation of a future reward, like an upcoming vacation, can be more gratifying than the experience of reward itself. Theories of anticipation have described how it causes behaviors ranging from beneficial information-seeking to harmful addiction. Here, we investigated how the brain generates and enhances anticipatory pleasure, by analyzing brain activity of human participants who received information predictive of future pleasant outcomes in a decision-making task. Using a computational model of anticipation, we show that three regions orchestrate anticipatory pleasure. We show ventromedial prefrontal cortex (vmPFC) tracks the value of anticipation; dopaminergic midbrain responds to information that enhances anticipation, while the sustained activity in hippocampus provides for functional coupling between these regions. This coordinating role for hippocampus is consistent with its known role in the vivid imagination of future outcomes. Our findings throw new light on the neural underpinnings of how anticipation influences decision-making, while also unifying a range of phenomena associated with risk and time-delay preference.

*kiigaya@caltech.edu

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

"Pleasure not known beforehand is half-wasted; to anticipate it is to double it."
– Thomas Hardy, *The Return of the Native*

Standard economic theory suggests a reward is more attractive when it is imminent (e.g. eating now) than when it is delayed (e.g. eating tomorrow), predicting that people always consume a reward immediately. This so-called temporal discounting[43,66,99] has been adapted with great success, such as to design an artificial intelligence that can plan its future efficiently[50,81,108] and to understand aspects of the human mind.[56,80,94,103] However, real life behavior is more complex.[4,33,37,39,75] Humans and other animals sometimes prefer deliberately to postpone pleasant experiences (e.g. saving a piece of cake for tomorrow, or delaying a one-time opportunity to kiss a celebrity[75]), clearly contradicting simple temporal discounting.

An alternative idea in behavioral economics is that we enjoy, or *savor*, the moments leading to reward, a notion known as the utility of anticipation.[4,16,55,67,69,75] This idea accounts for why people occasionally prefer to delay reward (e.g. because we can enjoy the anticipation of eating a cake until tomorrow by saving it now),[75] as well as behavior such as information-seeking and addiction.[16,53] However, with notable exceptions,[55] there have been few studies that have examined the neural basis of the pleasure of anticipation, compared to the pleasure arising from reward itself[5,97,102,114] and temporal discounting.[56,61,65,78,80] For instance, it has been speculated, but never directly tested, that the pleasure of anticipation arises from a vivid imagination of outcomes during waiting.[75]

Here we investigated the neurobiological mechanisms of anticipatory pleasure, by combining computational modeling, a behavioral task, and functional magnetic resonance imaging (fMRI). We fit our computational model[53] of anticipatory utility[75] to task behavior, and made predictions about the timecourse of anticipatory pleasure in the brain for each participant. We then compared this predicted signal with fMRI data, finding that the ventromedial prefrontal cortex (vmPFC) encoded an anticipatory pleasure signal. Additionally, dopaminergic midbrain encoded a signal reporting changes in reward expectation. This so-called reward prediction error (RPE) is widely interpreted as a reinforcement learning signal,[46,82,86,102] but our model predicts it acts to enhance anticipatory pleasure, which in turn drives behavior. We show that in fact hippocampus mediates this enhancement, by enhancing a functional coupling between two regions (the vmPFC and the dopaminergic midbrain). This can link anticipatory pleasure to a vivid imagination of future reward, in light of a strong link of hippocampus to memory and future imagination.[12,15,47,73,90,101]

Results

Anticipation drives a preference for advance information

We previously formalized, and validated behaviorally, a theoretical model of anticipation.[53] Here, we wanted to directly test neurobiological predictions arising out of this model, using the same behavioral task that we previously linked to anticipatory pleasure. In brief, our task examines how participants change their preference for resolving uncertainty about future pleasurable outcomes, according to reward probability and delay duration (please also see the Methods section). Participants made decisions with full knowledge regarding conditions (probability, and delay, of reward outcomes) as these were signaled with simple visual stimuli on each trial.

On each trial, participants chose between an immediate-information target (labeled 'Find out now') and a delayed-information target ('Keep it secret'). If the immediate-information target was

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

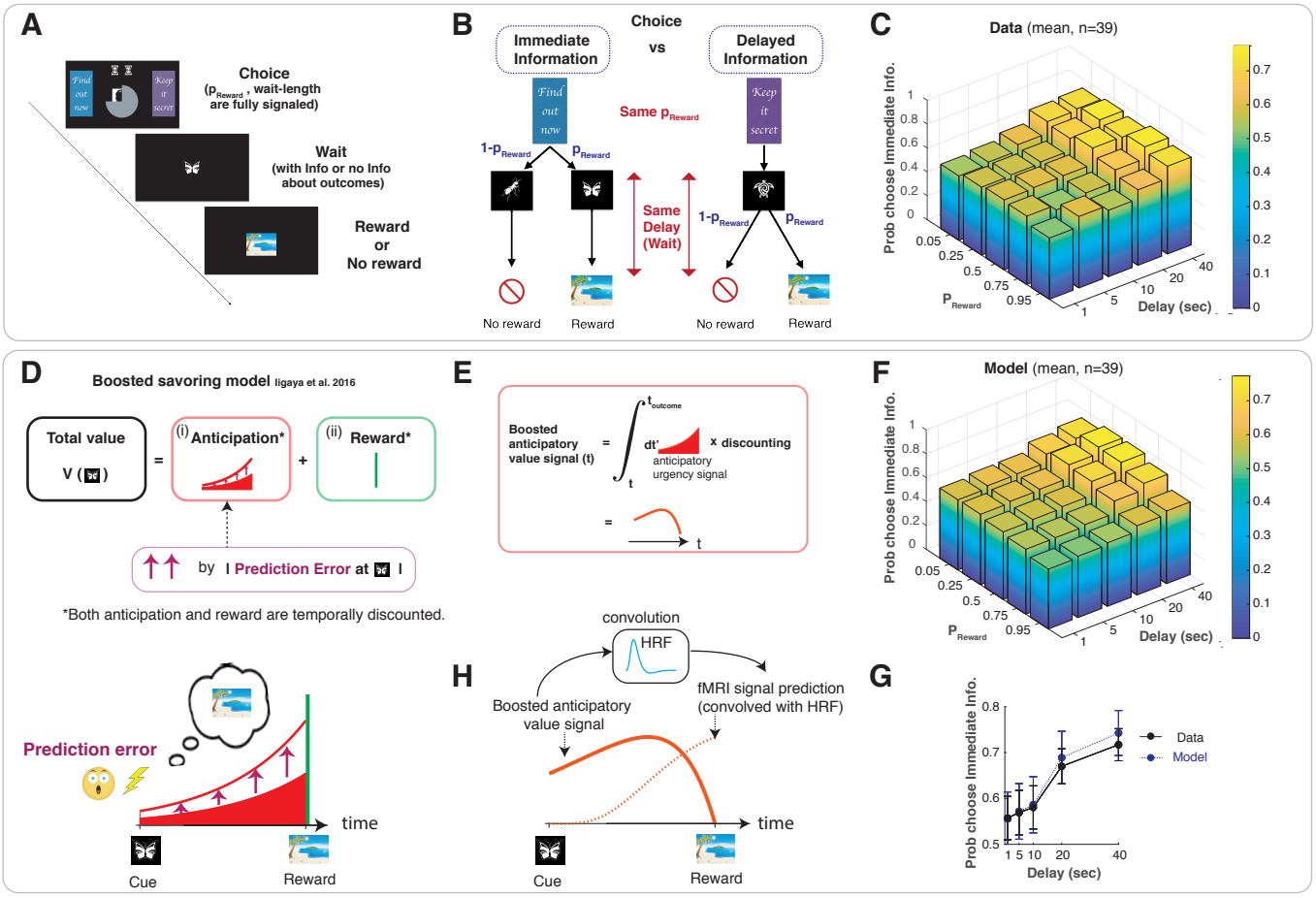

Figure 1: Anticipation drives a preference for advance information. (**A**). Task. On each trial, participants were presented with two lateral choice targets: an Immediate-information target ('Find out now') and a Delayed-information target ('Keep it secret') as well as two central stimuli signaling the probability of reward (sampled uniformly at random from 0.05, 0.25, 0.5, 0.75, 0.95) and the duration of a wait period until reward or no-reward delivery (sampled uniformly at random from 1, 5, 10, 20, 40 s). Once participants chose a target, a symbolic image cue was presented and remained present for the entire wait period. After this period, a rewarding image or an image signaling no reward appeared. (**B**). The Immediate-information-target was followed by symbolic cues that predict upcoming reward or no reward with no remaining ambiguity (reward predictive cue, or no-reward predictive cue). The Delayed-information-target was followed by a cue that implied nothing about the reward outcome (no-information cue), so that the future outcome remained uncertain during this wait period. (**C**). Average behavior. Participants showed a stronger preference for advance information under the longer delay conditions. The effect of reward probability was not significant at a group level (two-way ANOVA, $F_{4,950} = 10.0, p < 0.001$ and $F_{4,950} = 0.35, p > 0.05$, respectively). See **Figure S3** for individual differences. (**D**). Computational model.[53] Following a standard characterization of the utility of anticipation,[75] our model assumes that the value of each predictive cue is determined by the sum of (i) the utility of anticipation that can be consumed while waiting for reward (red), and (ii) the value of reward consumption itself (green). The value of anticipation is boosted by prediction errors associated with the reward predictive cues (purple upward arrows). This boosts the values of predictive cues, capturing the preference of advance information.[53] (**E**). The model's anticipatory value signal at time t is an integral of discounted future anticipation (urgency signal) at $t' > t$. This integrated signal evolves dynamically during the waiting period (red curve). (**F**). The average of modeled preferences across participants. We fit our model to participants' behavior using a hierarchical Bayesian procedure. (**G**). The model (blue) captures the effect of delay conditions in data (black). The error bars indicate the mean and standard errors of participants ($n = 39$). Please see **Figure S1** for how other classical models fail to capture the data. (**H**). The model's prediction for fMRI signals (dotted red) is computed by convolving the value signal (solid red) with the canonical haemodynamic response function (light blue).

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chosen, one of two cues, each of which uniquely signaled if reward would or would not arrive, was shown during the wait period (**Figure 1B**, left). If the delayed-information target was chosen, a separate non-predictive cue that carries no information about an upcoming outcome was shown on the screen during the wait period (**Figure 1B**, right), ultimately followed by reward or non-reward. The reward image was randomly drawn from previously validated rewarding pictures,[22,53] and consequently subject to immediate consumption (by viewing) upon delivery. The no-reward image was a neutral image indicating no-reward.

In this design, participants' choices did not affect final reward outcomes or duration of delays (**Figure 1B**) and both were pre-determined and signaled to participants at the beginning of each trial. Participants' choices influenced when they received information about rewards alone (i.e. immediately, or after delays, i.e. when they actually received outcomes). Conventional temporal-discounting predicts participants should be indifferent between the two choices, regardless of delay (please see **Figure S1**). However, consistent with our previous findings,[53] participants exhibited a preference for advance information in a manner that varied systematically with delay (**Figure 1C**).

We accounted for this information-seeking behavior using a modified version of the standard behavioral economic notion of the pleasure of anticipation, formally known as the utility (or economic value) of anticipation.[4,16,55,67,69,75,106] The standard model quantifies how people feel pleasure from anticipation of a future outcome while waiting. This comes on top of the value that arises from consumption of the outcome itself (**Figure 1D**). In behavioral economics, the anticipatory pleasure arising from desired outcomes (e.g. reward) is known as savoring, while negative feelings arising from anticipation of aversive outcomes (e.g. electric shock) is termed dread.[75]

However, by itself, the utility of anticipation does not capture participants' preference for advanced information (**Figure S1**[53]). We therefore proposed[53] a modification to the original formulation that people feel *enhanced* anticipatory pleasure after they unexpectedly discover a reward is impending (**Figure 1D**). This proposal is based on experimental observations that such unexpected discoveries lead animals to both become and remain more excited, than when animals waited for a certain reward with no such surprising discoveries.[105] Computationally, a surprise associated with finding out about future reward (or no-reward) creates a reward prediction error (RPE), defined by a change in expected value. Such RPEs are represented in the activity of dopamine neurons[46,82,86,102] — a neuromodulator that is also involved in enhanced motivation.[45,85,100,113] Therefore, in our model, a RPE triggers an enhancement (boosting) of anticipatory utility. In our task, the cue predictive of a future outcome that follows the immediate-information target creates a dopaminergic RPE, and it triggers a boosting of the pleasure of anticipation. The non-predictive cue following the delayed-information target does not generate RPE, thus it does not trigger any boosting.

It is important to note that a RPE is a phasic response that lasts only for a short period. However, animals appear to remain excited for whole anticipatory periods,[105] and so in the model, the enhancement of anticipation is sustained throughout a wait period.[53] Further, according to the model, surprise is quantified by the absolute value of the prediction error (and so a boosting is big for large negative RPEs as well as large positive ones). Taking an absolute value is a conventional mapping of prediction error to surprise,[89] and it also avoids unreasonable effects such as turning dread into savoring (by multiplying a negative anticipatory value with a negative RPE-based boost).[53] Therefore we expect the boosting signal to be a prolonged representation of the absolute value of RPE, which will likely be encoded in regions other than the ones encoding phasic RPEs.

We also note that the RPEs in our model are computed on the basis of the value of both anticipation and reward consumption (see the Methods section). With this RPE boosting hypothesis, our model captures a wide range of existing findings about information-seeking behavior, and offers potential links to addiction and gambling[53] (please also see the Discussion).

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The value of anticipation, regardless of boosting, ramps up as the outcome approaches, but the value is also subject to conventional discounting. This implies it carries less weight at the time of the choice as the outcome is further delayed, reflected in a tilted inverted-U shape over time under typical parameter settings (**Figure 1E**).

We fit our model to participants' behavioral data using a hierarchical Bayesian scheme[51,53] (see Methods section). As before,[53] the model captured participants' preferences for advance information (**Figure 1F,G**). We stress that other standard models, such as models with discounted reward but with no anticipatory value, or models with both discounted reward and anticipatory value but no enhancement of anticipation, cannot capture a preference for advance information (**Figure S1**). We formally tested this in a model comparison analysis using integrated Bayes Information Criterion (iBIC; please see the Methods section[53]), which strongly favored our full model over other models (**Figure S2**). However, despite the behavioral explanatory power, the neural roots of anticipation and its boosting are largely unknown.

Here we wanted to elucidate the neurobiological basis of three key components in our boosted anticipation model: namely, dynamic value representations of anticipatory pleasure; anticipation-dependent RPE signal at advance information cues, and a boosting signal of anticipation. We were also interested in how boosting relates to a vivid imagination of outcome, inspired by a long-standing, but never-tested, hypothesis in behavioral economics that a vivid imagination of outcomes determines the impact of anticipatory value.[75]

The vmPFC encodes the value of anticipation

We used each subject's maximum a posteriori (MAP) parameters, based on an hierarchical model fit,[51–53] to estimate subject-specific time courses of several variables, including the two separate value signals in our model: (i) discounted boosted-anticipatory value during wait periods (ii) discounted reward value during the same periods, and (iii) prediction errors at cue presentation. These signals were convolved with SPM's default canonical HRF (**Figure 1H**; see **Figure S4** for an example). As illustrated in the Methods section, we separated predictive signals for savoring (anticipation of reward with a positive value) and dread (anticipation of no reward with a negative value[53]).

We found the model's anticipatory value signal correlated significantly with BOLD in ventro-medial prefrontal cortex (vmPFC) ($p < 0.05$, whole brain FWE peak correction; MNI coordinates $[10, 50, 16]$, $t = 6.02$; **Figure 2A**), as well as dorsal caudate ($p < 0.05$, whole brain peak FWE correction; $[-20, -2, 18]$, $t = 5.81$ **Figure S5**). These results are consistent with value representation of imaginary reward reported in the vmPFC,[8] and of reported anticipatory activity in vmPFC[55,57,60] as well as in caudate.[28,54,63,72,98] Across the brain, we found no significant effect of dread that survived a stringent whole-brain correction (see **Figure S6**). Thus, we focus on savoring (the anticipatory value of future reward) referred to henceforth as anticipatory value.

Given a needed caution against potential false positives from auto-correlations in slowly-changing signals,[32] we conducted non-parametric, phase-randomization, tests wherein we scrambled the phases of signals in a Fourier decomposition[13,71,109] (**Figure S7A**). To do so, we transformed our model's predicted anticipatory value signal for each participant into Fourier space, randomized the phase of each frequency component, and transformed the signal back to the original space. Only the regressor being tested was randomized, while others were kept the same in the full GLM. We then performed a standard analysis on this full GLM for each participant with the scrambled signal, and then conducted a second level analysis. By repeating this procedure many times, we created a null distribution. To protect this test against family-wise error, we constructed the null distribution by taking a maximum value of correlation score across a region of interest, or across the whole brain, from each of our

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

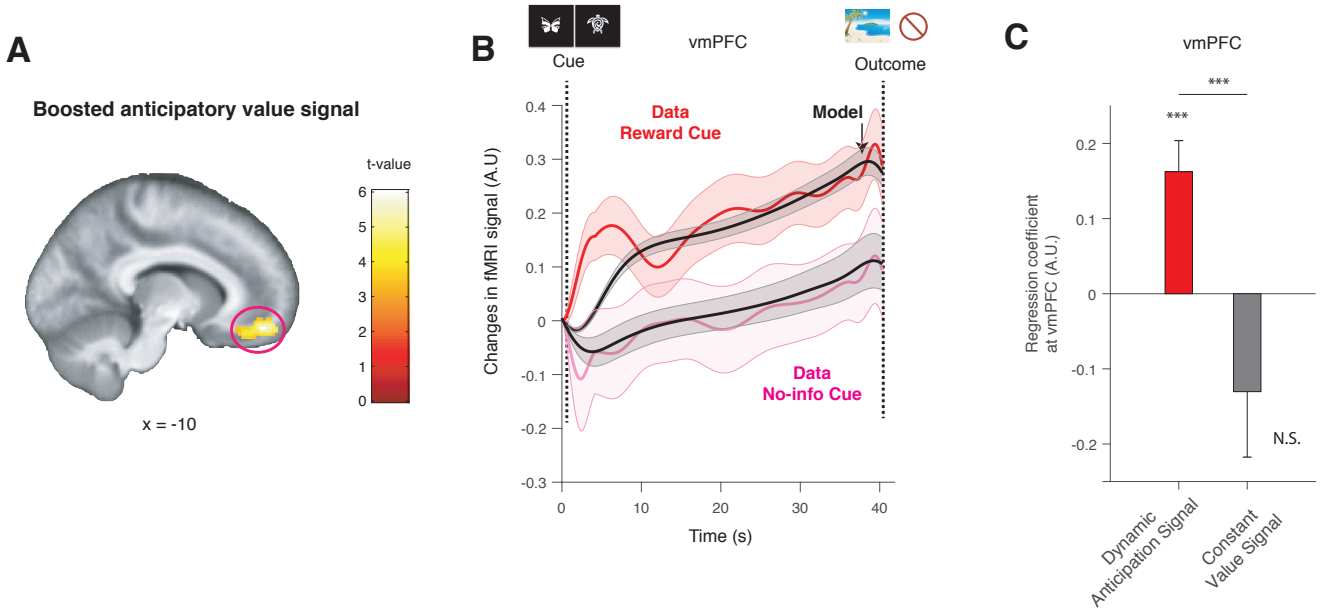

Figure 2: Neural representation of the anticipatory value signal in the vmPFC. (**A**). BOLD in vmPFC positively correlated with an anticipatory value signal ($p < 0.001$ according to a phase-randomization test; see **Figure S7**). (**B**). The temporal dynamics of BOLD signal in vmPFC matched the model's predictions during the anticipation period. Changes in activity averaged over participants following receipt of a reward predictive cue (Red), and after receipt of a no-information cue (magenta), are shown, as well as the model's prediction for each of these conditions (black). The error bar indicates the SEM over participants. The data was upsampled for this plot, following.[2] (**C**). A control, confirmatory, analysis shows that vmPFC did not correlate with a constant expected value signal, defined by a boxcar regressor modulated by the probability of reward, but did so with our model's dynamic anticipatory utility signal. The average regression weights in vmPFC were significantly larger than zero for our model's predicted signal ($p < 0.001$ t-test, $t_{38} = 3.93$), while the average weights were not significantly different from zero for the constant signal. The regression coefficients for the anticipatory value signal was significantly greater than the coefficients to the constant value signal ($p < 0.001$, permutation test). The error bars indicate the mean and SEM. Note that this is a confirmatory analysis, using the average coefficients of cluster in vmPFC, instead of peak voxel. The anticipation signal was generated separately for positive (reward) and negative (no-reward) part; the results for the positive part alone are shown in (**A,B,C**).

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

second level analysis, comparing against the correlation value in the original analysis. We found that the effects in the vmPFC ($p < 0.001$ whole-brain FWE corrected) and the caudate ($p < 0.01$ whole brain FWE corrected) survived this Fourier phase randomization test (**Figure S7B**). We note that this phase-randomization test can be useful in both neuroimaging and electrophysiology studies for avoiding false positive discoveries, particularly when analyzing correlations between slow signals.[32]

A more detailed inspection of these signals during the wait period showed the time course of the vmPFC activity closely resembled our model's predictions. In **Figure 2B**, we plot the time course of fMRI signals in the vmPFC during the wait period separately for two conditions, namely when participants received a reward predictive cue (red), and when participants received a no-information cue (magenta). The time-courses track the model's predictions in each condition (black).

We asked next whether BOLD in the vmPFC during the wait period correlated with a more conventional signal such as expected outcome value. When the immediate-information cue is presented, this is the same as the value of reward or no-reward without discounting or anticipatory modulation; otherwise it is an average of the value of reward and no-reward weighted by their respective probabilities. We examined the singular contribution of this signal by adding it as another parametric boxcar regressor during waiting periods to the original GLM, and then compared the beta values between the anticipatory value, and the expected value, regressor. In this way, we estimated the partial correlation of each regressor. As shown in **Figure 2C**, the vmPFC BOLD still positively correlated with the model's anticipatory value signal, and the effect of an expected value signal was not significant.

Apart from an anticipatory value signal, we found that the model's evolving discounted reward value signal correlated significantly with activity in regions including the superior temporal gyrus ($p < 0.05$, whole brain FWE peak correction; $[-48, -48, 16]$, $t = 5.28$, **Figure S8A**). This also survived a phase-randomization test ($p < 0.001$). For completeness, we report descriptively that an anticipatory urgency signal, which is an anticipation signal before integration (see Methods section) correlated with anterior insular cortex[64] (peak voxel $[34, 30, 2]$, which survived the phase randomization test $p < 0.01$, **Figure S8B**).

The dopaminergic midbrain encodes anticipation-dependent reward prediction errors at the time of predictive cues

The prediction error arising at advance cues is critical in our model, because it triggers the boosting of anticipation. We calculated a full, signed, prediction error signal, occasioned at the onset of advance information cues (reward predictive, no-reward predictive, and no-information cues), calculated by the discounted value of anticipation (of both savoring and dread) and that of outcomes. Because of the boosting, the value of the immediate information target and the amount of prediction error arising from the subsequent predictive cues are recursively computed in a self-consistent manner (see Methods). In this regard it differs from a conventional temporal difference (TD) prediction error,[82,108] which only considers conventionally discounted outcomes and does not involve boosting. We have previously shown computationally that this signal encompasses a previously-described "information prediction error"[9–11,18,79] (see also the Discussion section).

Based on extensive prior studies, we hypothesized that a signed prediction error signal arising at the time of predictive cues would be encoded in the midbrain dopaminergic regions and the ventral striatum (e.g.[24,48,87,102]). Indeed, the model's signal correlated significantly with BOLD in a midbrain dopaminergic region encompassing the ventral tegmental area and substantia nigra (VTA/SN) (**Figure 3A**; $p < 0.05$, small volume FWE correction;[48] $[4, -26, 20]$, $t = 3.78$). We also found that BOLD in the medial posterior parietal cortex (mPPC)[36,110] correlated significantly with the model's predicted signal (**Figure 3A**; $p < 0.05$, cluster-level whole brain FWE correction with the height threshold

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

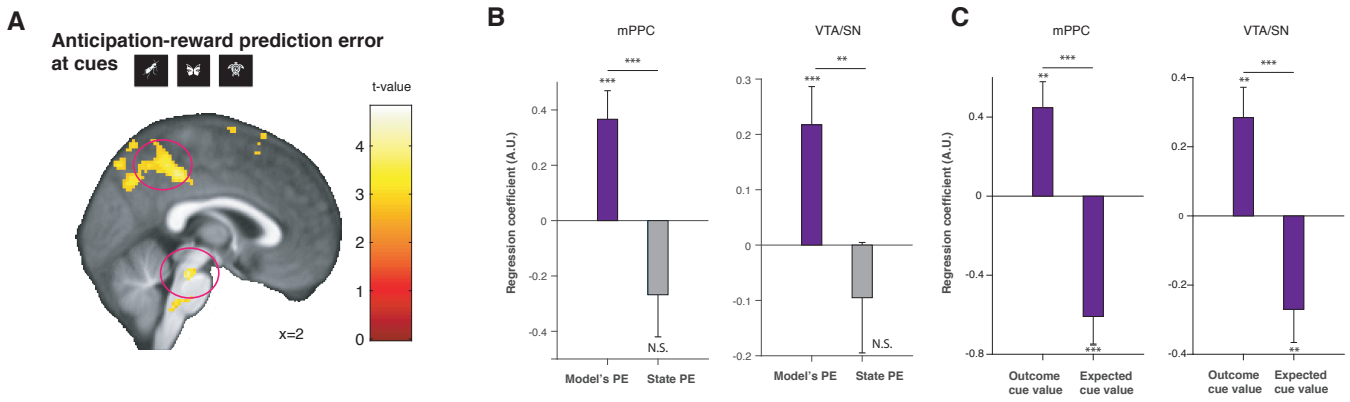

Figure 3: Neural representation of prediction error signals. (**A**) The mPPC and the VTA/SN BOLD positively correlated with the model's anticipatory reward prediction errors at the time of advance information cue presentations (reward predictive, no-reward predictive, and no-information cues) (mPPC: $p < 0.01$, whole-brain FWE. VTA/SN: $p < 0.01$ FWE small volume correction). Voxels at $p < 0.005$ (uncorrected) are highlighted for display purposes. (**B**) Our illustrative, confirmatory, analysis shows that BOLD signals in both the mPPC and the VTA/SN positively correlated with the model's prediction error signal (including anticipation and outcome), but not with a simpler, so-called state prediction error signal ($1 - p_{\text{reward}}$ when reward predictive cue was presented, $|0 - p_{\text{reward}}|$ when no-reward predictive cue was presented). The differences between the average regression coefficients in the mPPC and in the VTA/SN were significant in the mPPC ($p < 0.001$, a standard permutation test in which we permuted the average regression coefficients), and the VTA/SN ($p < 0.001$ permutation test). The average correlation with the model's prediction error signal was significant both in the mPPC and the VTA/SN ($p < 0.001$ for the mPPC and the VTA/SN; t-test $t_{38} = 3.56$ and $t_{38} = 3.15$). (**C**) Another confirmatory analysis shows that at the time of a potentially informative cue, BOLD in the mPPC and the VTA/SN positively correlated with the model's outcome cue value signal, and negatively with the model's expected value signal, indicating both regions express canonical prediction errors. This difference was significant in the mean coefficients in the mPPC ($p < 0.001$ permutation test) and the mean coefficients in the VTA/SN ($p < 0.001$ permutation test). The positive correlation with outcome values and the negative correlation with expected values were all significant in the mean values of the mPPC and the mean values of VTA/SN (outcome: $p < 0.01$ for the mPPC and the VTA/SN by t-test $t_{38} = 3.40$ and $t_{38} = 3.24$, expectation: $p < 0.001$ for the mPPC and $p < 0.01$ for the VTA/SN by t-test $t_{38} = 4.37$ and $t_{38} = 2.82$). Note that (**B**) and (**C**) are confirmatory analysis, using the average coefficients of clusters in the mPPC and the VTA/SN, instead of peak voxels. The three stars indicate $p < 0.001$, and two stars indicate $p < 0.01$.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$p < 0.001$; $k = 166$, peak at $[0, -42, 50]$.). We did not find significant associations in the ventral striatum, perhaps because cue- and reward onsets were unusually temporally distant (up to 40 sec), or it is rather consistent with a previous report that the anticipation of reward triggers a dopamine release in the caudate, but not in the ventral striatum.[98]

BOLD in the mPPC has previously been reported to covary with a simpler prediction error signal, the state prediction error signal.[40] In our experiment, this state-prediction-error signal is the absolute value of the difference between outcome (1 or 0) and expectation (the presented probability of reward). To rule out state prediction error as a driver of our results, we performed a confirmatory analysis, by constructing a GLM that included the model's full prediction error signal and its state prediction error signal, and the compared the beta values of partial correlations associated with these regressors. For both the mPPC (cluster defined by our functional analysis as above, conservatively for this analysis with $p < 0.001$ uncorrelated, which encompasses the peak voxel with FWE $p < 0.05$) and the VTA/SN (cluster defined anatomically[48]), the mean beta values in clusters correlated significantly with the model's prediction error signal, but not with the state prediction error signal (**Figure 3B**).

Further, previous studies suggested that significant correlations reported between fMRI signals and prediction errors might be attributable to a strong correlation with outcomes (reward or no reward) alone, regardless of the presence of negative correlations with expectation at outcome.[2,21,62] To rule out this possibility, we performed another confirmatory analysis by constructing a GLM with separate regressors for the values of outcomes and expectations. BOLD in both the mPPC (cluster defined by our functional analysis as above) and the VTA/SN (anatomically defined cluster[48]) positively correlated with outcome value and negatively correlated with expected value (**Figure 3C**). Thus, responses in these regions had characteristic of canonical prediction error signals.[2,21,62]

The hippocampus correlates with boosting of anticipation

Our computational model predicts people experience enhanced anticipation following a reward prediction error, occasioned by advance information cues in our task. In our model, the magnitude of enhancement (or boosting) is quantified by the amount of surprise participants experienced,[89] defined by the absolute value of reward prediction error. Also, the boosting is expected to be prolonged over the entire duration of a wait period (**Figure 4A**). In original work on the utility of anticipation,[75] the strength of anticipation was speculated to be related to a vivid imagination of desired outcomes. In our computational model, the latter can be re-stated in terms of whether this vivid imagination relates to the boosting of anticipation.

Previous research suggests that the hippocampus is an ideal substrate for this effect. First, in the context of recognition tasks, the hippocampus encodes unsigned error (mismatch, novelty) signal.[19,25,31,70,74,107] Additionally, many studies link hippocampal activity to the imagination of future prospects (e.g.[12,47,73,101]), where prefrontal-medial temporal interactions influence the effects of imagination on valuation,[90] as well as support the mental construction of future events.[15] This suggests that the boosting of anticipatory value is mediated by enhancing hippocampal-dependent imagination of associated outcomes.

To test this we first examined activity in the hippocampus in response to the absolute value of the model's prediction error at the time the advance information cue was presented. As predicted, we found hippocampal activity was significantly correlated with the magnitude of the model's unsigned prediction error ($p < 0.05$ FWE small volume correction by neurosynth; peak voxel at $[32, -24, -12]$, $t = 3.60$, **Figure 4B**).

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

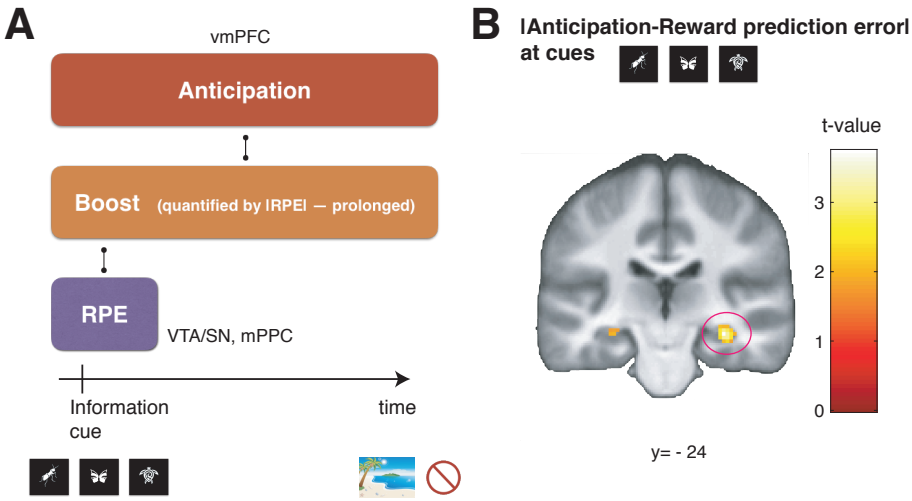

Figure 4: Neural correlates of boosting in the hippocampus. (**A**). Our model predicts that dopaminergic reward prediction error triggers a boosting of anticipation. The boosting is quantified by the absolute value of reward prediction error. In contrast to the phasic, short, reward prediction error signal, the effect of boosting needs to be sustained during the anticipatory period. (**B**). The absolute value of the model's prediction error signal significantly correlated with BOLD in the hippocampus (FWE $p < 0.05$ small volume correction).

Midbrain-hippocampus-vmPFC circuit dynamically computes anticipatory value

We next asked whether boosted hippocampal activity persisted throughout the delay period, as predicted by our computational model. Indeed, we found that the activity in the hippocampus was boosted in rewarded, immediate-information, trials throughout the wait period, compared to uncertain, delayed-information, trials (**Figure 5A**; see also **Figure S9**). Thus, in addition to expressing the magnitude of prediction error at advance information cues, hippocampal BOLD during the wait period expresses a sustained variable related to anticipation. This mixed coding of prediction error and anticipation-like signal suggests that hippocampus connects reciprocally with both to the VTA/SN[74] and to the vmPFC,[15] playing a central role in boosting anticipation. Hippocampal-vmPFC connection has been strongly tied with vivid future imagination[15] and on this basis we hypothesized that hippocampal-dependent prospective process realize boosting, linking computations in the VTA/SN (prediction error) and the vmPFC (boosted-anticipatory value).

To formally test this idea, we analyzed functional connectivity using dual psychophysical interaction (PPI) regressors based on two a priori seed regions: 1) the vmPFC (which encodes anticipatory value) and the model's full, signed, prediction error signal at predictive cues (which is encoded at the VTA/SN) as a psychological variable, and 2) the VTA/SN (which encodes prediction error) as a seed and the model's anticipatory value signal (which is encoded at the vmPFC) as a psychological variable. Because each of these two PPI regressors includes variables relating to both the vmPFC (anticipation) and the VTA/SN (prediction error), and these variables are coupled in our computational model through the notion of boosting, this analysis tests our hypothesis that the hippocampus links the VTA/SN (prediction error) and the vmPFC (anticipation) as a potential medium of boosting. Thus we included these two sets of regressors into the single GLM we used so far used (see Methods section), and tested if hippocampal activity significantly correlated with these PPI regressors.

We found significant correlations in the hippocampus for both PPI regressors. Thus, a functional coupling between the VTA/SN (the area encoding reward prediction errors) and the hippocampus[104]

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

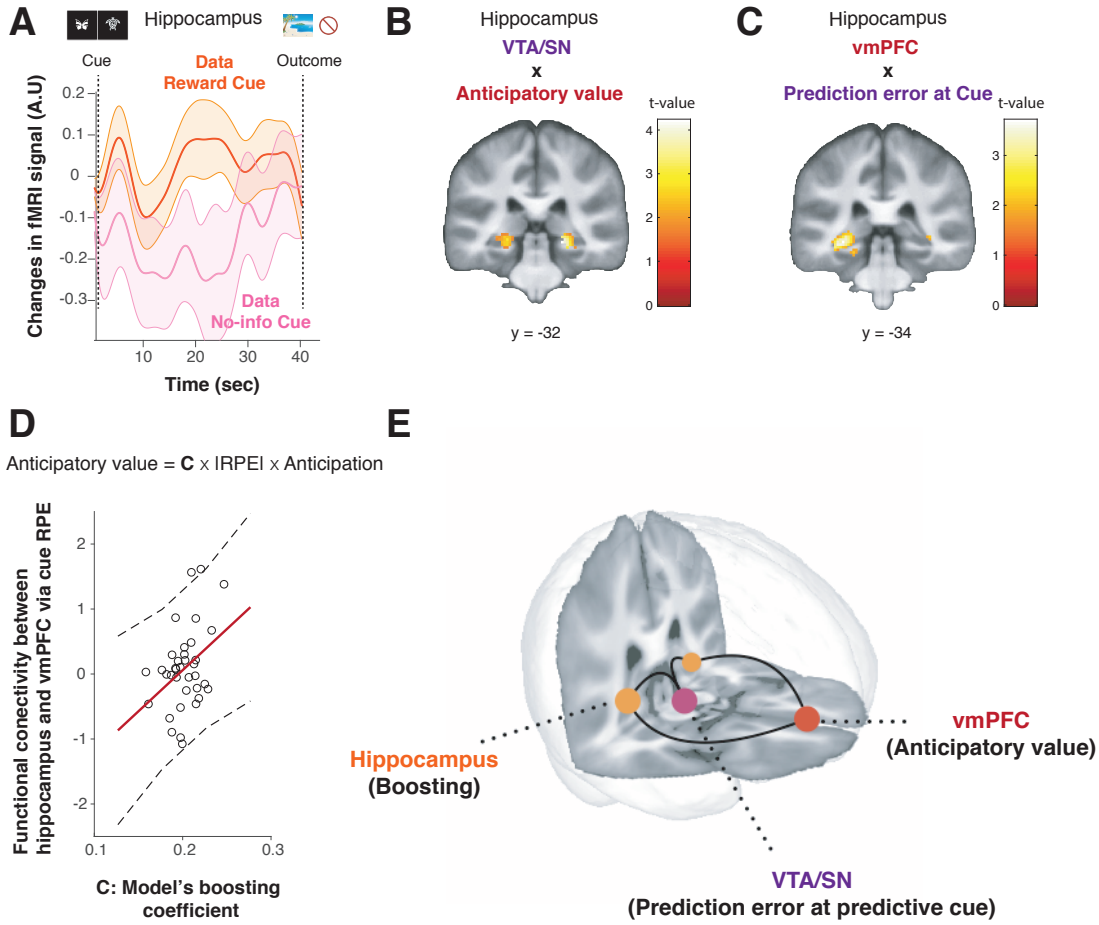

Figure 5: Functional connectivity analysis suggestive of a neural network for anticipatory value computation. (**A**). The temporal dynamics of fMRI signal in the hippocampus. Changes in activity averaged over participants after receiving a reward predictive cue (orange), and after receiving a no-information cue (magenta), are shown. The coding of boosting-related value are prolonged over the course of wait period, which is what our model predicted. The error bar indicates the SEM. Please also see **Figure S9**. (**B**). Functional coupling between the VTA/SN and the hippocampus is positively modulated by the model's anticipatory value signal (PPI regressor: BOLD signal in VTA/SN modulated by model's anticipatory value signal (encoded in vmPFC); $p < 0.05$ FWE small volume correction). (**C**). Functional coupling between the vmPFC and the hippocampus is positively modulated by model's prediction error signal (PPI regressor: BOLD signal in vmPFC modulated by model's prediction error signal. $p < 0.05$ FWE small volume correction). PPI regressors in (**B**) and (**C**) were chosen because each of them includes variables relating to both the VTA/SN (prediction error) and the vmPFC (anticipatory value), allowing us to test if the hippocampus links to both these two regions in a way predicted by our computational model. Note that (**B**) and (**C**) are partial correlation results from the same analysis using a single GLM with the two PPI regressors that are symmetrically orthogonalized. (**D**). The functional coupling strength between the vmPFC and the hippocampus mediated by the model's prediction error signal is positively correlated with the model's boosting coefficient parameter estimated by behavior of participants ($r = 0.37$, $p < 0.05$). (**E**). Functional network for anticipatory value computation. Three distinctive regions play together to construct the anticipatory value, in a manner that is predicted by our computational model. The 3D brain image was constructed by the mean T1 brain images, which were cut at $y = 34$ and $z = 15$ for illustrative purposes.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

was significantly modulated by our model's anticipatory value signal ($p < 0.05$, FWE small volume correction by Neurosynth hippocampus mask; peak voxel at $[22, -32, -6]$, $t = 3.89$, **Figure 5B**). Additionally, a functional coupling between the vmPFC and the hippocampus[111] was significantly modulated by our model's prediction error signal ($p < 0.05$, FWE small volume correction by Neurosynth hippocampus mask; peak voxel at $[-30, -34, -6]$, $t = 3.70$, **Figure 5C**).

If the hippocampal-vmPFC coupling is responsible for the boosting of anticipation, the coupling strength that we estimated in our PPI analysis should relate to the magnitude of boosting. We estimated the magnitude of boosting from participants' behavior by model fitting. The linear boosting coefficient (parameter C) is a constant that is unique to each participant and determines the impact of RPE on boosting. Over participants we found a significant positive correlation between the hippocampal-vmPFC coupling strength, estimated from fMRI data, and the model's linear boosting coefficient (parameter C), estimated from each participant's behavior (**Figure 5D**). Note that the model's boosting coefficient is different from a simple behavioral measure, such as a choice dependence on reward probabilities (**Figure S10**).

These functional connectivity results support our hypothesis that hippocampus plays a key coordinating role in boosting the pleasure of anticipation, linking the vmPFC's encoding of the value of anticipation with the VTA/SN's encoding of prediction errors at advance information. The findings point to these regions functioning as a large-scale neural network for boosting pleasure of anticipation (**Figure 5E**), driving a preference for advance information and curiosity.

Discussion

While the pleasure (utility) of anticipation has long been recognized as a key notion in behavioral economics, its neural mechanism remains surprisingly unexplored. Recent behavioral evidence suggests that anticipatory pleasure is enhanced as the likelihood of a desired experience increases unexpectedly. This psychological process links anticipatory utility theory to a wide range of suboptimal human behavior, such as information-seeking, risk-seeking, and addiction. Here, we took advantage of a new link between computational theory and behavior, and applied it to a laboratory experiment with fMRI, to reveal how anticipatory pleasure arises in the brain (please see **Figure S11** for a visual summary of the Discussion section).

Crucially, we show a network for computing the utility of anticipation, consisting of three distinctive brain regions. First, we show vmPFC represents the time course of an anticipatory utility signal that evolved during a waiting period. Second, dopaminergic midbrain regions, encompassing VTA/SN, encoded the model's prediction error that signals changes in expected value (prediction error). Third, hippocampal activity which indexed an unsigned prediction error was functionally coupled both to vmPFC and to the VTA/SN, in a manner consistent with our model's predicted boosting of anticipation with prediction error. One plausible cognitive account of these findings is that a dopaminergic prediction error signal projected to hippocampus[74] serves to amplify vivid imagination of future outcomes,[12,15,47,73,90,101] and this in turn strengthens the pleasure of anticipation.[75]

Our study provides insights into neural processes underling suboptimal human decision-makings. One example in our current study concerns a preference for early resolution of uncertainty,[16,33,38,68,69] also known as information-seeking,[9,10,18,41,116] or observing.[3,29] Humans and other animals are willing to incur costs to find out their true fate, even if this knowledge does not change actual outcome.[7,14,18,33,59,76] An alternative idea, as oppose to our boosted anticipatory utility, is that people feel pleasure from information itself.[10,42,68] However this so-called intrinsic value of information cannot explain why a preference for advance information is valence-dependent,[18,58] that it depends

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

incorrectly on the reward probability,[23,96] and a sensitivity to delay until reward (as we also demonstrated here).[53,105] All of these findings are natural consequence of the pleasure of anticipation.

Previous neural findings about the intrinsic value of information can be accounted for by our model of anticipation. This so-called information-prediction-error signal is presumed to arise from the value of information,[9–11,18,79] and has been reported in the same midbrain dopaminergic regions as standard reward prediction errors,[9,10] implying that the two signals might be strongly related.[84] Indeed, our model accounts for information prediction errors as a side effect of anticipation-dependent reward predictions, which we show correlates with BOLD signal in dopaminergic midbrain regions[9,10] as well as in the mPPC[36] (see also[110]).

Our results offer alternative accounts for addiction, and the possibility of individually-tailored psychiatric interventions (**Figure S11**). While initial phases of addiction[93] involve excessive dopamine release at the time of drug consumption,[92] later phases involve craving,[91,95] a close relation to savoring of anticipation. Our model implies that people boost craving when the likelihood of drug administration increases (e.g. when purchasing drugs). People may feel greater pleasure from obtaining drugs (which can act as a kind of conditioned stimuli[35]) than from administering them, because the former includes values associated with the anticipation of future administration. Importantly, our model predicts that people with certain parameter values (e.g. large boosting coefficients) could repeatedly over-boost the pleasure of anticipating drugs, resulting in excessive, pathological, drug seeking (see Eq.(12)). By fitting our model to subjects performing the task used here, we can in principle link an individual's tendency toward addiction with unique cause of this disorder (e.g. excessive boosting or unbalance between anticipation and discounting). This can in turn suggest interventions tailored to individual patients, such as cognitive behavioral therapy focusing on controlling anxiety and craving,[34] as well as possible dopaminergic antagonists to control boosting.

Our study also unifies separate notions concerning gambling: preference for risk and time-delay. While these two economic phenomena are often been treated separately there is increasing evidence in favor of an interactive relationship (e.g.[1,17]). Our computational model of anticipation explicitly offers an interaction between risk (prediction error) and delay (anticipation), because the former can enhance the value of latter. This interaction creates well-documented effects, such as nonlinear coding of probabilities of anticipated rewards.[49] Further studies may allow us to design a behavioral task for psychiatric interventions, in which patients can lessen their preference for addicted substances, or even their risk preference in general, because our model can find the optimal task parameters for each individual to achieve this goal.

Our study offers an alternative view to a long-standing problem in neuroscience and machine-learning. The so-called temporal credit assignment problem raises the issue of how neurons operating on a timescale of milliseconds learn relationships on a behaviorally relevant timescale (such as actions and rewards in our task), while designing a machine learning algorithm overcoming this problem also remains to be a challenge. A recent physiological study demonstrated that synaptic plasticity in hippocampal pyramidal neurons can learn associations on a behaviorally relevant timescale, with the aid of ramping-like, slow, external inputs in a realistic setting.[6] Our results suggest that a slow anticipatory value signal in the vmPFC could serve as such an input to neurons in the hippocampus, bridging a gap over behavioral timescales. A dopaminergic input from the VTA/SN to the hippocampus may facilitates this type of learning.[44,74] Therefore our study suggests that anticipation processes may serve to bridge a gap in the temporal credit assignment problem.

Neuroeconomic studies show that people make decisions between goods in different categories, by expressing the value of those goods in a common currency primarily encoded in the vmPFC (e.g.[20]). Here we found that the utility of anticipation is indeed computed in the vmPFC. This invites an alternative interpretation of previously reported ramping activity in the vmPFC while waiting for

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

rewards[30,83,115] in terms of an anticipation-sensitive value signal, which has been instead interpreted as a reward-timing signal.

We also note an influential suggestion[16] that future uncertainty drives other forms of anticipatory utility, such as anxiety. We did not consider this notion directly in our analysis; but in our model an agent can experience a mixture of positive (savoring) and negative (dread) utilities of anticipation according to the probabilities of these outcomes. It would be interesting to study how this mixed anticipatory utility of our model relates to the notion of anxiety,[16] which may help the design of more effective psychiatric interventions for anxiety disorders.

In sum, we identified novel neural substrates for computing the pleasure arising from anticipation, orchestrated by three distinctive brain regions with different associated functions. We suggest this anticipatory pleasure drives a range of behaviors including information-seeking, addiction, and gambling. Our study also provides a seed for individually tailored interventions for psychiatric disorders.

Acknowledgment

We thank Tim Behrens for his insightful suggestion for the Fourier phase randomization test. We thank Colin Camerer, George Loewenstein, Sandro Romani, Ethan Bromberg-Martin, Jackie Gotlieb, Elliot Ludvig, Yunzhe Liu, Bastian Blain, Giles Story, Laurence Hunt, Jeff Cockburn, Vincent Man, Tomas Aquiono, Caroline Charpentier, Bowen Fung, Wolfgang Pauli, Erin Burkett for the most valuable discussions and helpful suggestions for the manuscript. We also thank radiographers at the UCL for their assistance in running fMRI experiments. This work was supported by the Max Planck Society, the Gatsby Foundation, Wellcome Trust Investigator Award, Japan Society for the Promotion of Science, the Swartz Foundation, Wellcome Sir Henry Dale Fellowship (211155/Z/18/Z), the Jacobs Foundation (2017-1261-04), the Medical Research Foundation, and 2018 NARSAD Young Investigator grant (27023) from the Brain & Behavior Research Foundation.

References

[1] V. Anderhub, W. Güth, U. Gneezy, and D. Sonsino. On the interaction of risk and time preferences: An experimental study. *German Economic Review*, 2(3):239–253, 2001.

[2] T. E. Behrens, L. T. Hunt, M. W. Woolrich, and M. F. Rushworth. Associative learning of social value. *Nature*, 456(7219):245, 2008.

[3] U. R. Beierholm and P. Dayan. Pavlovian-instrumental interaction in 'observing behavior'. *PLoS computational biology*, 6(9):e1000903, 2010.

[4] G. S. Berns, J. Chappelow, M. Cekic, C. F. Zink, G. Pagnoni, and M. E. Martin-Skurski. Neurobiological substrates of dread. *Science*, 312(5774):754–8, 2006.

[5] K. C. Berridge. Food reward: brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews*, 20(1):1–25, 1996.

[6] K. C. Bittner, A. D. Milstein, C. Grienberger, S. Romani, and J. C. Magee. Behavioral time scale synaptic plasticity underlies ca1 place fields. *Science*, 357(6355):1033–1036, 2017.

[7] T. C. Blanchard, B. Y. Hayden, and E. S. Bromberg-Martin. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron*, 2015.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[8] S. Bray, S. Shimojo, and J. P. O'Doherty. Human medial orbitofrontal cortex is recruited during experience of imagined as well as real rewards. *Journal of neurophysiology*, 2010.

[9] E. S. Bromberg-Martin and O. Hikosaka. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, 63(1):119–126, 2009.

[10] E. S. Bromberg-Martin and O. Hikosaka. Lateral habenula neurons signal errors in the prediction of reward information. *Nature neuroscience*, 14(9):1209, 2011.

[11] M. Brydevall, D. Bennett, C. Murawski, and S. Bode. The neural encoding of information prediction errors during non-instrumental information seeking. *Scientific reports*, 8(1):6134, 2018.

[12] R. L. Buckner. The role of the hippocampus in prediction and imagination. *Annual review of psychology*, 61:27–48, 2010.

[13] E. Bullmore, C. Long, J. Suckling, J. Fadili, G. Calvert, F. Zelaya, T. A. Carpenter, and M. Brammer. Colored noise and computational inference in neurophysiological (fmri) time series analysis: resampling methods in time and wavelet domains. *Human brain mapping*, 12(2):61–78, 2001.

[14] J. A. M. R. Cabrero, J. Zhu, and E. Ludvig. Costly curiosity: People pay a price to resolve an uncertain gamble early. *PsyArXiv*, 2018.

[15] K. L. Campbell, K. P. Madore, R. G. Benoit, P. P. Thakral, and D. L. Schacter. Increased hippocampus to ventromedial prefrontal connectivity during the construction of episodic future events. *Hippocampus*, 28(2):76–80, 2018.

[16] A. Caplin and J. Leahy. Psychological expected utility theory and anticipatory feelings. *Quarterly Journal of economics*, pages 55–79, 2001.

[17] L. S. Carvalho, S. Prina, and J. Sydnor. The effect of saving on risk attitudes and intertemporal choices. *Journal of Development Economics*, 120:41–52, 2016.

[18] C. J. Charpentier, E. S. Bromberg-Martin, and T. Sharot. Valuation of knowledge and ignorance in mesolimbic reward circuitry. *Proceedings of the National Academy of Sciences*, 115(31):E7255–E7264, 2018.

[19] J. Chen, R. K. Olsen, A. R. Preston, G. H. Glover, and A. D. Wagner. Associative retrieval processes in the human medial temporal lobe: hippocampal retrieval success and ca1 mismatch detection. *Learning & Memory*, 18(8):523–528, 2011.

[20] V. S. Chib, A. Rangel, S. Shimojo, and J. P. O'Doherty. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, 29(39):12315–12320, 2009.

[21] R. Chowdhury, M. Guitart-Masip, C. Lambert, P. Dayan, Q. Huys, E. Düzel, and R. J. Dolan. Dopamine restores reward prediction errors in old age. *Nature neuroscience*, 16(5):648, 2013.

[22] M. J. Crockett, B. R. Braams, L. Clark, P. N. Tobler, T. W. Robbins, and T. Kalenscher. Restricting temptations: Neural mechanisms of precommitment. *Neuron*, 79(2):391–401, 2013.

[23] N. Daddaoua, M. Lopes, and J. Gottlieb. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Scientific reports*, 6:20202, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[24] K. D'ardenne, S. M. McClure, L. E. Nystrom, and J. D. Cohen. Bold responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, 319(5867):1264–1267, 2008.

[25] J. Y. Davidow, K. Foerde, A. Galván, and D. Shohamy. An upside to reward sensitivity: the hippocampus supports enhanced reinforcement learning in adolescence. *Neuron*, 92(1):93–99, 2016.

[26] P. Dayan. How to set the switches on this thing. *Current opinion in neurobiology*, 22(6):1068–1074, 2012.

[27] P. Dayan, Y. Niv, B. Seymour, and N. D. Daw. The misbehavior of value and the discipline of the will. *Neural networks*, 19(8):1153–1160, 2006.

[28] M. Delgado, V. Stenger, and J. Fiez. Motivation-dependent responses in the human caudate nucleus. *Cerebral Cortex*, 14(9):1022–1030, 2004.

[29] J. Dinsmoor. Observing and conditioned reinforcement. *Behav Brain Sc*, 6:693–728, 1983.

[30] N. A. Donnelly, O. Paulsen, T. W. Robbins, and J. W. Dalley. Ramping single unit activity in the medial prefrontal cortex and ventral striatum reflects the onset of waiting but not imminent impulsive actions. *European Journal of Neuroscience*, 41(12):1524–1537, 2015.

[31] K. Duncan, N. Ketz, S. J. Inati, and L. Davachi. Evidence for area ca1 as a match/mismatch detector: A high-resolution fmri study of the human hippocampus. *Hippocampus*, 22(3):389–398, 2012.

[32] L. Elber-Dorozko and Y. Loewenstein. Striatal action-value neurons reconsidered. *eLife*, 7:e34248, 2018.

[33] K. Eliaz and A. Schotter. Paying for confidence: An experimental study of the demand for non-instrumental information. *Games and Economic Behavior*, 70(2):304–324, 2010.

[34] J. B. Engelmann, F. Meyer, E. Fehr, and C. C. Ruff. Anticipatory anxiety disrupts neural valuation during risky choice. *Journal of Neuroscience*, 35(7):3085–3099, 2015.

[35] B. J. Everitt and T. W. Robbins. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature neuroscience*, 8(11):1481, 2005.

[36] N. C. Foley, S. P. Kelly, H. Mhatre, M. Lopes, and J. Gottlieb. Parietal neurons encode expected gains in instrumental information. *Proceedings of the National Academy of Sciences*, 114(16):E3315–E3323, 2017.

[37] S. Frederick, G. Loewenstein, and T. O'donoghue. Time discounting and time preference: A critical review. *Journal of economic literature*, 40(2):351–401, 2002.

[38] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo. Active inference and epistemic value. *Cognitive neuroscience*, 6(4):187–214, 2015.

[39] C. D. Gipson, J. J. Alessandri, H. C. Miller, and T. R. Zentall. Preference for 50% reinforcement over 75% reinforcement by pigeons. *Learning & behavior*, 37(4):289–298, 2009.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[40] J. Gläscher, N. Daw, P. Dayan, and J. P. O'Doherty. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4):585–595, 2010.

[41] J. Gottlieb, P.-Y. Oudeyer, M. Lopes, and A. Baranes. Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in cognitive sciences*, 17(11):585–593, 2013.

[42] S. Grant, A. Kajii, and B. Polak. Intrinsic preference for information. *Journal of Economic Theory*, 83(2):233–259, 1998.

[43] L. Green and J. Myerson. A discounting framework for choice with delayed and probabilistic rewards. *Psychological bulletin*, 130(5):769, 2004.

[44] M. J. Gruber, B. D. Gelman, and C. Ranganath. States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, 84(2):486–496, 2014.

[45] A. A. Hamid, J. R. Pettibone, O. S. Mabrouk, V. L. Hetrick, R. Schmidt, C. M. Vander Weele, R. T. Kennedy, B. J. Aragona, and J. D. Berke. Mesolimbic dopamine signals the value of work. *Nature neuroscience*, 19(1):117–126, 2016.

[46] A. S. Hart, R. B. Rutledge, P. W. Glimcher, and P. E. Phillips. Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *Journal of Neuroscience*, 34(3):698–704, 2014.

[47] D. Hassabis, D. Kumaran, S. D. Vann, and E. A. Maguire. Patients with hippocampal amnesia cannot imagine new experiences. *Proceedings of the National Academy of Sciences*, 104(5):1726–1731, 2007.

[48] T. U. Hauser, E. Eldar, and R. J. Dolan. Separate mesocortical and mesolimbic pathways encode effort and reward learning signals. *Proceedings of the National Academy of Sciences*, 114(35):E7395–E7404, 2017.

[49] M. Hsu, I. Krajbich, C. Zhao, and C. F. Camerer. Neural response to reward anticipation under risk is nonlinear in probabilities. *Journal of Neuroscience*, 29(7):2231–2237, 2009.

[50] M. Hutter. *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. Springer Science & Business Media, 2004.

[51] Q. J. Huys, R. Cools, M. Gölzer, E. Friedel, A. Heinz, R. J. Dolan, and P. Dayan. Disentangling the roles of approach, activation and valence in instrumental and pavlovian responding. *PLoS Comput Biol*, 7(4):e1002028, 2011.

[52] K. Iigaya, M. S. Fonseca, M. Murakami, Z. F. Mainen, and P. Dayan. An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. *Nature communications*, 9(1):2477, 2018.

[53] K. Iigaya, G. W. Story, Z. Kurth-Nelson, R. J. Dolan, and P. Dayan. The modulation of savouring by prediction error and its effects on choice. *Elife*, 5:e13747, 2016.

[54] T. Jia, C. Macare, S. Desrivières, D. A. Gonzalez, C. Tao, X. Ji, B. Ruggeri, F. Nees, T. Banaschewski, G. J. Barker, et al. Neural basis of reward anticipation and its genetic determinants. *Proceedings of the National Academy of Sciences*, 113(14):3879–3884, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[55] K. Jimura, M. S. Chushak, and T. S. Braver. Impulsivity and self-control during intertemporal decision making linked to the neural dynamics of reward value representation. *Journal of Neuroscience*, 33(1):344–357, 2013.

[56] J. W. Kable and P. W. Glimcher. An 'as soon as possible' effect in human intertemporal decision making: behavioral evidence and neural mechanisms. *Journal of Neurophysiology*, 103(5):2513–2531, 2010.

[57] T. Kahnt, J. Heinzle, S. Q. Park, and J.-D. Haynes. The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 107(13):6010–6015, 2010.

[58] N. Karlsson, G. Loewenstein, and D. Seppi. The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty*, 38(2):95–115, 2009.

[59] C. Kidd and B. Y. Hayden. The psychology and neuroscience of curiosity. *Neuron*, 88(3):449–460, 2015.

[60] H. Kim, S. Shimojo, and J. P. O'doherty. Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cerebral cortex*, 21(4):769–776, 2010.

[61] S. Kim, J. Hwang, and D. Lee. Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron*, 59(1):161–172, 2008.

[62] M. C. Klein-Flügge, L. T. Hunt, D. R. Bach, R. J. Dolan, and T. E. Behrens. Dissociable reward and timing signals in human midbrain and ventral striatum. *Neuron*, 72(4):654–664, 2011.

[63] B. Knutson, C. M. Adams, G. W. Fong, and D. Hommer. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*, 21(16):RC159–RC159, 2001.

[64] B. Knutson and S. M. Greer. Anticipatory affect: neural correlates and consequences for choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1511):3771–3786, 2008.

[65] S. Kobayashi and W. Schultz. Influence of reward delays on responses of dopamine neurons. *Journal of neuroscience*, 28(31):7837–7846, 2008.

[66] T. C. Koopmans. Stationary ordinal utility and impatience. *Econometrica: Journal of the Econometric Society*, pages 287–309, 1960.

[67] B. Kőszegi. Utility from anticipation and personal equilibrium. *Economic Theory*, 44(3):415–444, 2010.

[68] D. M. Kreps and E. L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica: journal of the Econometric Society*, pages 185–200, 1978.

[69] A. Kumar, M. A. Killingsworth, and T. Gilovich. Waiting for merlot: Anticipatory consumption of experiential and material purchases. *Psychological science*, 25(10):1924–1931, 2014.

[70] D. Kumaran and E. A. Maguire. Match–mismatch processes underlie human hippocampal responses to associative novelty. *Journal of Neuroscience*, 27(32):8517–8524, 2007.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[71] A. R. Laird, B. P. Rogers, and M. E. Meyerand. Comparison of fourier and wavelet resampling methods. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 51(2):418–422, 2004.

[72] J. Lauwereyns, Y. Takikawa, R. Kawagoe, S. Kobayashi, M. Koizumi, B. Coe, M. Sakagami, and O. Hikosaka. Feature-based anticipation of cues that predict reward in monkey caudate nucleus. *Neuron*, 33(3):463–473, 2002.

[73] H. Lee, J.-W. Ghim, H. Kim, D. Lee, and M. Jung. Hippocampal neural correlates for values of experienced events. *Journal of Neuroscience*, 32(43):15053–15065, 2012.

[74] J. E. Lisman and A. A. Grace. The hippocampal-vta loop: controlling the entry of information into long-term memory. *Neuron*, 46(5):703–713, 2005.

[75] G. Loewenstein. Anticipation and the valuation of delayed consumption. *The Economic Journal*, pages 666–684, 1987.

[76] G. Loewenstein. The psychology of curiosity: A review and reinterpretation. *Psychological bulletin*, 116(1):75, 1994.

[77] G. Loewenstein and R. H. Thaler. Anomalies: intertemporal choice. *Journal of Economic perspectives*, 3(4):181–193, 1989.

[78] K. Louie and P. W. Glimcher. Separating value from choice: delay discounting activity in the lateral intraparietal area. *Journal of Neuroscience*, 30(16):5498–5507, 2010.

[79] C. B. Marvin and D. Shohamy. Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, 145(3):266, 2016.

[80] S. M. McClure, K. M. Ericson, D. I. Laibson, G. Loewenstein, and J. D. Cohen. Time discounting for primary rewards. *The Journal of Neuroscience*, 27(21):5796–5804, 2007.

[81] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[82] P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of neuroscience*, 16(5):1936–1947, 1996.

[83] H. Niki and M. Watanabe. Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain research*, 171(2):213–224, 1979.

[84] Y. Niv and S. Chan. On the value of information and other rewards. *Nature neuroscience*, 14(9):1095, 2011.

[85] Y. Niv, N. D. Daw, D. Joel, and P. Dayan. Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3):507–520, 2007.

[86] J. P. O'Doherty, P. Dayan, K. Friston, H. Critchley, and R. J. Dolan. Temporal difference models and reward-related learning in the human brain. *Neuron*, 38(2):329–337, 2003.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[87] J. P. O'Doherty, R. Deichmann, H. D. Critchley, and R. J. Dolan. Neural responses during anticipation of a primary taste reward. *Neuron*, 33(5):815–826, 2002.

[88] S. Palminteri, V. Wyart, and E. Koechlin. The importance of falsification in computational cognitive modeling. *Trends in cognitive sciences*, 21(6):425–433, 2017.

[89] J. M. Pearce and G. Hall. A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological review*, 87(6):532–552, 1980.

[90] J. Peters and C. Büchel. Episodic future thinking reduces reward delay discounting through an enhancement of prefrontal-mediotemporal interactions. *Neuron*, 66(1):138–148, 2010.

[91] C. L. Pickens, M. Airavaara, F. Theberge, S. Fanous, B. T. Hope, and Y. Shaham. Neurobiology of the incubation of drug craving. *Trends in neurosciences*, 34(8):411–420, 2011.

[92] A. D. Redish. Addiction as a computational process gone awry. *Science*, 306(5703):1944–1947, 2004.

[93] A. D. Redish, S. Jensen, A. Johnson, and Z. Kurth-Nelson. Reconciling reinforcement learning models with behavioral extinction and renewal: implications for addiction, relapse, and problem gambling. *Psychological review*, 114(3):784, 2007.

[94] S. Reimers, E. A. Maylor, N. Stewart, and N. Chater. Associations between a one-shot delay discounting measure and age, income, education and real-world impulsive behavior. *Personality and Individual Differences*, 47(8):973–978, 2009.

[95] T. E. Robinson and K. C. Berridge. The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain research reviews*, 18(3):247–291, 1993.

[96] K. L. Roper and T. R. Zentall. Observing behavior in pigeons: The effect of reinforcement probability and response cost using a symmetrical choice procedure. *Learning and Motivation*, 30(3):201–220, 1999.

[97] R. B. Rutledge, N. Skandali, P. Dayan, and R. J. Dolan. A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, 111(33):12252–12257, 2014.

[98] V. N. Salimpoor, M. Benovoy, K. Larcher, A. Dagher, and R. J. Zatorre. Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nature neuroscience*, 14(2):257–262, 2011.

[99] P. A. Samuelson. A note on measurement of utility. *The review of economic studies*, 4(2):155–161, 1937.

[100] T. Satoh, S. Nakai, T. Sato, and M. Kimura. Correlated coding of motivation and outcome of decision by dopamine neurons. *Journal of neuroscience*, 23(30):9913–9923, 2003.

[101] D. L. Schacter, D. R. Addis, and R. L. Buckner. Remembering the past to imagine the future: the prospective brain. *Nature reviews neuroscience*, 8(9):657, 2007.

[102] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, Mar 1997.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[103] R. Shadmehr, J. J. O. de Xivry, M. Xu-Wilson, and T.-Y. Shih. Temporal discounting of reward and the cost of time in motor control. *Journal of Neuroscience*, 30(31):10507–10516, 2010.

[104] D. Shohamy and A. D. Wagner. Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, 60(2):378–389, 2008.

[105] M. L. Spetch, T. W. Belke, R. C. Barnet, R. Dunn, and W. D. Pierce. Suboptimal choice in a percentage-reinforcement procedure: Effects of signal condition and terminal-link length. *Journal of the experimental analysis of behavior*, 53(2):219–234, 1990.

[106] G. W. Story, I. Vlaev, B. Seymour, J. S. Winston, A. Darzi, and R. J. Dolan. Dread and the disvalue of future pain. *PLoS computational biology*, 9(11):e1003335, 2013.

[107] B. A. Strange, A. Duggins, W. Penny, R. J. Dolan, and K. J. Friston. Information theory, novelty and hippocampal responses: unpredicted or unpredictable? *Neural Networks*, 18(3):225–230, 2005.

[108] R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine learning*, 3(1):9–44, 1988.

[109] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Physica D: Nonlinear Phenomena*, 58(1-4):77–94, 1992.

[110] L. L. van Lieshout, A. R. Vandenbroucke, N. C. Müller, R. Cools, and F. P. de Lange. Induction and relief of curiosity elicit parietal and frontal activity. *Journal of Neuroscience*, pages 2816–17, 2018.

[111] R. A. Weilbächer and S. Gluth. The interplay of hippocampus and ventromedial prefrontal cortex in memory-based decision making. *Brain sciences*, 7(1):4, 2016.

[112] N. Weiskopf, C. Hutton, O. Josephs, and R. Deichmann. 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, 2006.

[113] R. A. Wise. Dopamine, learning and motivation. *Nature reviews neuroscience*, 5(6):483, 2004.

[114] R. A. Wise and P.-P. Rompre. Brain dopamine and reward. *Annual review of psychology*, 40(1):191–225, 1989.

[115] M. Xu, S.-y. Zhang, Y. Dan, and M.-m. Poo. Representation of interval timing by temporally scalable firing patterns in rat prefrontal cortex. *Proceedings of the National Academy of Sciences*, 111(1):480–485, 2014.

[116] J.-Q. Zhu, W. Xiang, and E. A. Ludvig. Information seeking as chasing anticipated prediction errors. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Participants

Thirty-nine self-declared heterosexual male participants were recruited from the University College London (UCL) community. Participants provided informed consent for their participation in the study, which was approved by the UCL ethics committee.

Experimental task

The task was a variant of that in,[53] which itself was inspired by a series of animal experiments into information-seeking or observing behavior (e.g.[9,105]). At the beginning of each trial, a pair of task-information stimuli (hourglass and partially-covered human silhouette) were shown along with two choice targets. The number on the hour-glass indicated how long the participants had to wait until to see a reward or no-reward, where $1/2$, 1, 2 4, 8 hour-glass meant 1, 5, 10, 20, 40 sec of waiting time, respectively. The other stimulus, a partially-covered human silhouette, indicated the probability of seeing a reward, specified by the area of uncovered semi-circle (5, 25, 50, 75, 95 % chance of rewards). Two lateral rectangular targets were presented as choices: the Immediate-Information target marked as 'find out now', and the Delayed-Information target marked as 'keep it secret'. The positions of the hourglass, and the covered silhouette, were kept the same every trial, but the locations of choice targets were randomly alternated between left and right on each trial.

The participants were required to choose between left and right targets by pressing a button within three seconds. Once the participants chose a target, one of the three cues appeared in the center of the screen. If the participants chose the Immediate-information target, then a cue that signalled upcoming reward or no-reward appeared on the screen until the onset of reward or no-reward. If the participant chose the Delayed-information target, a cue that signalled no-information about reward appeared on the screen. The meaning of the cues were fully instructed to participants beforehand. The meanings of the cues were counter-balanced across subjects. In order to ensure immediate consumption, rewards were images of attractive female models from a set that had previously been validated as being suitably appetitive to heterosexual male subjects;[22,53] reward images were presented for 1s. Images were chosen randomly from the top 100 highest rated pictures that were introduced in.[22] No image was presented more than twice to the same participants. In case of no-reward, an image signaling absence of a reward was presented for 1s. In either case, a blank screen was presented for 1s before starting a new trial. These timings were set to reduce the timing uncertainty which may cause prediction error that can interfere with our model's value computation.[62]

Participants were fully instructed about the task structure including the meaning of stimuli about the probability and delay conditions, as well as the advance information cues. Then participants underwent extensive training that consisted of three tasks: a variable-delay but fixed probability task, a fixed-delay but variable probability task, a variable-delay and variable-probability task. This ensured that participants had fully learned the task and had adequately developed preferences before being scanned. Scanning was split into three separate runs, each of which consisted of 25 trials that covered all conditions once. Trial orders were randomized across participants. Subjects had a break of approximately 30 sec between runs.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Computational model

The model for the task was fully described in.[53] Briefly, following Loewenstein's suggestion that the anticipation of rewards itself has hedonic value[4,75] (e.g. subjects enjoy thinking about rewards while waiting for them), we extended a standard reinforcement learning framework to include explicit reward anticipation, i.e., savoring. The model's innovation is to suggest that the value of anticipation can be boosted by reward prediction errors associated with advance information about upcoming rewards.[53]

To describe the model formally, consider a task in which if a subject chooses the Immediate-Information target, then they receive at $t = 0$ a reward predictive cue S^+ with a probability of q, or a negative-reward predictive cue S^- with a probability of $1 - q$. Subsequently, the subject receives a reward or no-reward at $t = T(= T_{\text{Delay}})$, with a value of R^+ or R^-, respectively. In our recent experiment we have found that subjects assigned a negative value to an absence of reward;[53] but this is not necessary to account for preference for advance information that has been observed in animals.[39,105]

Based on the observation that subjects prefer to delay consumption of certain types of rewards, Loewenstein proposed that subjects enjoy anticipation while waiting to enjoy the actual reward.[4,75,106] Formally, the anticipation of the reward at time t is worth $a(t) = Re^{-\nu(T-t)}$, where ν governs its rate. Including R itself, and taking temporal discounting into account, the total value of the reward predictive cue, Q_{S^+}, is

$$\begin{aligned} Q_{S^+} &= \eta \; V^{[\text{Anticipation}]} + V^{[\text{Reward}]} \\ &= \eta \int_0^T e^{-\gamma t'} a(t') dt' + R^+ e^{-\gamma T} \\ &= \eta \frac{R^+}{\nu^+ - \gamma} \left(e^{-\gamma T} - e^{-\nu^+ T} \right) + R^+ e^{-\gamma T}, \end{aligned} \tag{1}$$

where η is the relative weight of anticipation, γ^+ is the discounting rate, and T is the duration of delay until the reward is delivered. In prior work, η had been treated as a constant that relates to subject's ability of imagine future outcomes[77] ; however, we proposed that it can vary with the prediction error δ_{pe} at the time of the predicting cue.[53] Our proposal was inspired by findings of a dramatically enhanced excitement that follows such cues.[105] A simple form of boosting arises from the relationship

$$\eta = \eta_0 + C|\delta_{pe}| \tag{2}$$

where η_0 specifies the base anticipation, and C determines the gain. That anticipation is boosted by the *absolute value* of RPE is important in applying our model to comparatively unpleasant outcomes.[53] We also note that the boosting is sustained throughout a wait period, unlike phasic RPE signals.[102]

The total value of the no-reward predictive cue, Q_{S^-}, is then

$$\begin{aligned} Q_{S^-} &= \eta \int_0^T e^{-\gamma^- t'} a(t') dt' + R^- e^{-\gamma^- T} \\ &= \eta \frac{R^-}{\nu^- - \gamma^-} \left(e^{-\gamma^- T} - e^{-\nu^- T} \right) + R^- e^{-\gamma^- T}. \end{aligned} \tag{3}$$

Following our previous work, we assumed that $\gamma = \gamma^+ = \gamma^-$.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Note that in our model, RPE affects the total cue values Q_{S^+} and Q_{S^-}, which in turn also affect subsequent RPEs. Therefore the linear ansatz for the boosting of anticipation on RPE (Eq.(2) could lead to instability due to unbounded boosting. This instability could account for maladaptive behavior such as addiction and gambling. However, in a wide range of parameters, this ansatz has a stable, self-consistent, solution. In our experiment, the prediction error for the reward and no-reward predictive cues can be expressed as

$$\delta_{pe}^{S^+} = Q_{S^+} - \left(qQ_{S^+} + (1-q)\, Q_{S^-}\right) \tag{4}$$

$$\delta_{pe}^{S^-} = Q_{S^-} - \left(qQ_{S^+} + (1-q)\, Q_{S^-}\right), \tag{5}$$

which are, assuming the linear ansatz,

$$\begin{cases} \delta_{pe}^{S^+} &= (1-q)\left(\left(\eta_0 + C|\delta_{pe}^{S^+}|\right) A^+ + B^+ - \left(\left(\eta_0 + C|\delta_{pe}^{S^-}|\right) A^- + B^-\right)\right) \\ \delta_{pe}^{S^-} &= -q\left(\left(\eta_0 + C|\delta_{pe}^{S^+}|\right) A^+ + B^+ - \left(\left(\eta_0 + C|\delta_{pe}^{S^-}|\right) A^- + B^-\right)\right) \end{cases} \tag{6}$$

where

$$\begin{cases} A^+ &= \frac{R^+}{\nu^+ - \gamma}\left(e^{-\gamma T} - e^{-\nu^+ T}\right) \\ A^- &= \frac{R^-}{\nu^- - \gamma}\left(e^{-\gamma T} - e^{-\nu^- T}\right) \\ B^+ &= R^+ e^{-\gamma T} \\ B^- &= R^- e^{-\gamma T} \end{cases} \tag{7}$$

Assuming that $R^- \leq 0$ and $0 \leq R^+$, Equations (6) impose that $\delta_{pe}^{S^+} > 0$ and $\delta_{pe}^{S^-} < 0$. With this, Equations (6) can be reduced to

$$\begin{cases} \delta_{pe}^{S^+} &= \frac{(1-q)\left(\eta_0\left(A^+ - A^-\right) + B^+ - B^-\right)}{1 - C((1-q)A^+ - qA^-)} \\ \delta_{pe}^{S^-} &= \frac{-q\left(\eta_0\left(A^+ - A^-\right) + B^+ - B^-\right)}{1 - C((1-q)A^+ - qA^-)} \end{cases} \tag{8}$$

Because $\left(\eta_0\left(A^+ - A^-\right) + B^+ - B^-\right) > 0$, in order that $\delta_{pe}^{S^+} > 0$ and $\delta_{pe}^{S^-} < 0$ hold for all q and T, the denominators must be positive for all $0 \leq q \leq 1$ and $0 \leq T$. In other words,

$$1 - C\left((1-q)\, A^+ - qA^-\right) > 0. \tag{9}$$

for $0 \leq q \leq 1$ and $0 \leq T$, meaning that

$$C < \frac{1}{\left((1-q)\, A^+ - qA^-\right)}, \tag{10}$$

for $0 \leq q \leq 1$ and $0 \leq T$. This means that

$$C < \frac{1}{\max\left(A^+, |A^-|\right)} \tag{11}$$

for $0 \leq T$. It is straightforward to show that A^+ takes its maximum at $T = \frac{ln\left(\frac{\gamma}{\nu^+}\right)}{\gamma - \nu^+}$, and $|A^-|$ at $T = \frac{ln\left(\frac{\gamma}{\nu^-}\right)}{\gamma - \nu^-}$. Thus the condition that the linear ansatz gives a stable self-consistent solution is

$$C < \min\left(\frac{\gamma}{R^+}\left(\frac{\gamma}{\nu^+}\right)^{\frac{\nu^+}{\gamma - \nu^+}}, \frac{-\gamma}{R^-}\left(\frac{\gamma}{\nu^-}\right)^{\frac{\nu^-}{\gamma - \nu^-}}\right) \tag{12}$$

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In our model-fitting, we imposed this stability condition. However, we note that it is also possible that some people violate this condition. This could account for maladaptive behavior such as addiction and pathological risk seeking.

An alternative to imposing such a stability condition would be to assume that boosting saturates in a non-linear manner:[53]

$$\eta = \eta_0 + c_1 \tanh\left(c_2|\delta_{pe}|\right).$$

However, the model's qualitative behavior does not depend strongly on the details of the RPE dependence of anticipation.[53] Hence we only used the linear ansatz in our analysis in the current study.

We note that our current computational model[53] was originally introduced in light of an experimental refutation[10] of an earlier reinforcement-learning model for information-seeking (observing) behavior.[3] This early computational model[3] appealed to the same underlying idea discussed here, namely that of a Pavlovian influence[27] of a prediction error signal over actions.[26] The present model considers an anticipatory process (a type of an internal action[26]), in contrast to this earlier model that characterized this effect in terms of suppressed forgetting.

We generated choice probability[88] from our model by taking a difference between the expected value of immediate information target and that of delayed information target and taking it through sigmoid with a noise parameter σ.[53]

For our model comparison, we also fit a model with no anticipation $\eta = 0$, and a model with an anticipation but that is not boosted by RPE $C = 0$.

Behavioral Model fitting

We used a hierarchical Bayesian, random effects analysis.[51–53] In this, the (suitably transformed) parameters $\mathbf{h}_i$ of participant i are treated as a random sample from a Gaussian distribution with means and variance $\boldsymbol{\theta} = \{\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta\}$ characterising the whole population of subjects; and we find the maximum likelihood values of $\boldsymbol{\theta}$.

The prior distribution $\boldsymbol{\theta}$ can be set as the maximum likelihood estimate:

$$\begin{aligned}\boldsymbol{\theta}^{ML} &\approx \text{argmax}_{\boldsymbol{\theta}} \left\{p\left(D|\boldsymbol{\theta}\right)\right\} \\ &= \text{argmax}_{\boldsymbol{\theta}} \left\{\prod_{i=1}^{N} \int d\mathbf{h_i}\, p\left(D_i|\mathbf{h}_i\right) p\left(\mathbf{h}_i|\boldsymbol{\theta}\right)\right\}\end{aligned} \tag{13}$$

We optimized $\boldsymbol{\theta}$ using an approximate Expectation-Maximization procedure. For the E-step of the k-th iteration, a Laplace approximation gives us

$$\mathbf{m}_i^k \approx \text{argmax}_{\mathbf{h}} \left\{p\left(D_i|\mathbf{h}\right) p\left(\mathbf{h}|\boldsymbol{\theta}^{k-1}\right)\right\} \tag{14}$$

$$p\left(\mathbf{h}_i^k|D_i\right) \approx \mathcal{N}\left(\mathbf{m}_i^k, \boldsymbol{\Sigma}_i^k\right), \tag{15}$$

where $\mathcal{N}\left(\mathbf{m}_i^k, \boldsymbol{\Sigma}_i^k\right)$ is a Normal distribution with mean $\mathbf{m}_i^k$ and covariance $\boldsymbol{\Sigma}_i^k$ that is obtained from the inverse Hessian around $\mathbf{m}_i^k$. For the M step:

$$\boldsymbol{\mu}_\theta^{k+1} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{m}_i^k \tag{16}$$

$$\boldsymbol{\Sigma}_\theta^{k+1} = \frac{1}{N}\sum_{i=1}^{N} \left(\mathbf{m}_i^k \mathbf{m}_i^{k\mathbf{T}} + \boldsymbol{\Sigma}_i^k\right) - \boldsymbol{\mu}_\theta^{k+1} \boldsymbol{\mu}_\theta^{k+1\mathbf{T}}. \tag{17}$$

For simplicity, we assumed that the covariance Σ_θ^k had zero off-diagonal terms, assuming that the effects were independent.[51]

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Model comparison

We compared the goodness of fit for different computational models according to their integrated Bayes Information Criterion (iBIC) scores.[51,53] We analyzed log-likelihood of data D given a model M, $\log\ p(D|M)$:

$$\log p\left(D|M\right) = \int d\theta p\left(D|\theta\right) p\left(\theta|M\right) \tag{18}$$

$$\approx -\frac{1}{2}\text{iBIC} = \log p\left(D|\theta^{ML}\right) - \frac{1}{2}|M|\log|D|, \tag{19}$$

where iBIC is the *integrated* Bayesian Information Criterion, $|M|$ is the number of fitted prior parameters and $|D|$ is the number of data points (total number of choice made by all subjects). Here, $\log p\left(D|\theta^{ML}\right)$ can be computed by integrating out individual parameters:

$$\log p\left(D|\theta^{ML}\right) = \sum_i \log \int d\mathbf{h} p\left(D_i|\mathbf{h}\right) p\left(\mathbf{h}|\theta^{ML}\right) \tag{20}$$

$$\approx \sum_i \log \frac{1}{K}\sum_{j=1}^{K} p\left(D_i|\mathbf{h}^j\right), \tag{21}$$

where we approximated the integral as the average over K samples $\mathbf{h}^j$'s generated from the prior $p\left(\mathbf{h}|\theta^{ML}\right)$.

Model's fMRI predictions

Our computational model makes specific predictions about temporal dynamics of anticipatory, reward, value signals during wait periods, as well as unique anticipatory reward prediction error signals at predictive cue onsets. Using the parameters (MAP estimates) for each participant, we generated the following variables for each participant as parametric regressors for the fMRI analysis.

The temporal dynamics of anticipatory value signal for positive domain at time t during wait period, until reward onsets $t = T$ are:

$$V_{\text{Anticipation},+}\left(t\right) = \frac{R^+\left(\eta_0 + C|\delta_{pe}^{\left[S^+,q,T\right]}|\right)}{\nu^+ - \gamma}\left(e^{-\gamma(T-t)} - e^{-\nu^+(T-t)}\right) \tag{22}$$

for the negative domain:

$$V_{\text{Anticipation},-}\left(t\right) = \frac{R^-\left(\eta_0 + C|\delta_{pe}^{\left[S^-,q,T\right]}|\right)}{\nu^- - \gamma}\left(e^{-\gamma(T-t)} - e^{-\nu^-(T-t)}\right) \tag{23}$$

We expressed these as two separate regressors.

The prediction errors $\delta_{pe}^{[+/-,q,T]}$ are determined for each delay T and reward probability q condition self-consistently (see below). After a Delayed-Information choice, these signals are scaled by the probability of reward q or no-reward $1-q$ (and no prediction errors). Note that we set $R^+ = 1$ without loss of generality.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The discounted reward signal at t during the wait period is expressed as

$$V_{\text{Reward},+}(t) = R^{+}e^{-\gamma(T-t)}, \tag{24}$$

while the discounted no-reward signal at t is

$$V_{\text{Reward},-}(t) = R^{-}e^{-\gamma(T-t)}. \tag{25}$$

Note that the anticipatory value signal is an integral of (discounted) anticipation urgency signal:

$$V_{\text{Ant. Urgency},+}(t) = R^{+}\left(\eta_0 + C|\delta_{pe}^{\left[S^{+},q,T\right]}|\right)e^{-\nu^{+}(T-t)}, \tag{26}$$

and

$$V_{\text{Ant. Urgency},-}(t) = R^{-}\left(\eta_0 + C|\delta_{pe}^{\left[S^{-},q,T\right]}|\right)e^{-\nu^{-}(T-t)}, \tag{27}$$

which we also included to the GLM.

The reward prediction errors at information cue onsets are computed for each condition (q,T) self-consistently according to Eqs.(8). That is

$$\begin{cases} \delta_{pe}^{\left[S^{+},q,T\right]} &= \frac{(1-q)\left(\eta_0\left(A^{+}-A^{-}\right)+B^{+}-B^{-}\right)}{1-C((1-q)A^{+}-qA^{-})} \\ \delta_{pe}^{\left[S^{-},q,T\right]} &= \frac{-q\left(\eta_0\left(A^{+}-A^{-}\right)+B^{+}-B^{-}\right)}{1-C((1-q)A^{+}-qA^{-})} \end{cases} \tag{28}$$

where $A^{+/-}$, $B^{+/-}$, are given by Eqs.(7). In our analysis we put positive and negative prediction errors as a single parametric regressor at information cue onsets. Note that the anticipatory/reward prediction error signal is different from other conventional prediction error signals, including the so-called state prediction errors:[40]

$$\begin{cases} \delta_{spe}^{S^{+}} &= |1-q| \\ \delta_{spe}^{S^{-}} &= |0-q|, \end{cases} \tag{29}$$

which we used for a control analysis.

fMRI data acquisition

We acquired MRI data using a Siemens Trio 3Tesla scanner with a 32-channel head coil. The EPI sequence was optimized for minimal signal dropout in striatal, medial prefrontal, and brainstem regions:[112] 40 slices with 3-mm isotropic voxels with repetition time (TR) 2.8 s and echo time (TE) 30 ms, and slice tilt 30 degrees. Additionally, field maps (3-mm isotropic, whole-brain) were acquired to correct the EPIs for field-strength inhomogeneity.

fMRI analysis

We used SPM12 (Wellcome Trust Centre for Neuroimaging, UCL, London) for standard preprocessing and image analysis. The first 5 scans of each session were discarded to account for T1-saturation effects. The standard preprocessing includes: slice-timing correction; realigned and unwarped with the field maps that were obtained before the task; co-registration of structural T1 weighted images

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to the sixth functional image of each subject; segmenting structural images into grey matter, white matter and cerebral spinal fluid; normalising structural and functional images spatially to the Montreal Neurological Institute (MNI) space; spatially smoothing with a Gaussian kernel with full-width half-maximum of 8mm. The motion correction parameters were estimated from the realignment procedure, and were included to the first level GLM analysis.

We regressed fMRI time series with GLMs that consist of onset regressors and our model's signals, as well as nuisance regressors. Onset regressors included the presentations of the initial screen, the presentations of cues, the presentation of outcomes. The parametric modulated regressors were added to the presentation of cues with model's prediction error signal, and outcome onset with the value of reward or no reward. The onsets of cues preceding the shortest delay ($1s$) was separately modeled so that the prediction errors at the cues were not affected by reward. The time course of our model's signals included the anticipatory value signals for positive and negative domains, the discounted value signals for positive and negative domains, anticipatory urgency signal for positive and negative domains. The model's predictive signals were generated for each of the anticipatory periods, using the model that was fit to each participant, which were then convolved with the canonical HRF function. We added nuisance parameters that consist of movement estimated from preprocessing, large derivatives of movement between volumes that were larger than 1 mm, box function during the anticipatory periods, box function for each experimental run. In our control analysis, we also added box function during the anticipatory periods that was parametrically modulated by constant expectation of reward, parametrically modulated cue presentation with state prediction errors.

Regions of interests

The region of interests for VTA/SN was taken from.[48] The region of interests for hippocampus was taken from Neurosynth. Other cluster were taken from our functional results, as described in the Results sections.

PPI analysis

We performed PPI analysis with a single GLM, which contained 1) BOLD signal of VTA/SN 2) a PPI regressor that is an interaction between the BOLD signal of VTA/SN and model's anticipatory value signal 3) the BOLD signal of vmPFC 4) a PPI regressor that is an interaction between the BOLD signal of vmPFC and model's prediction error signal, as well as other onset/movement regressors that we included in our original analysis.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemantary figures

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

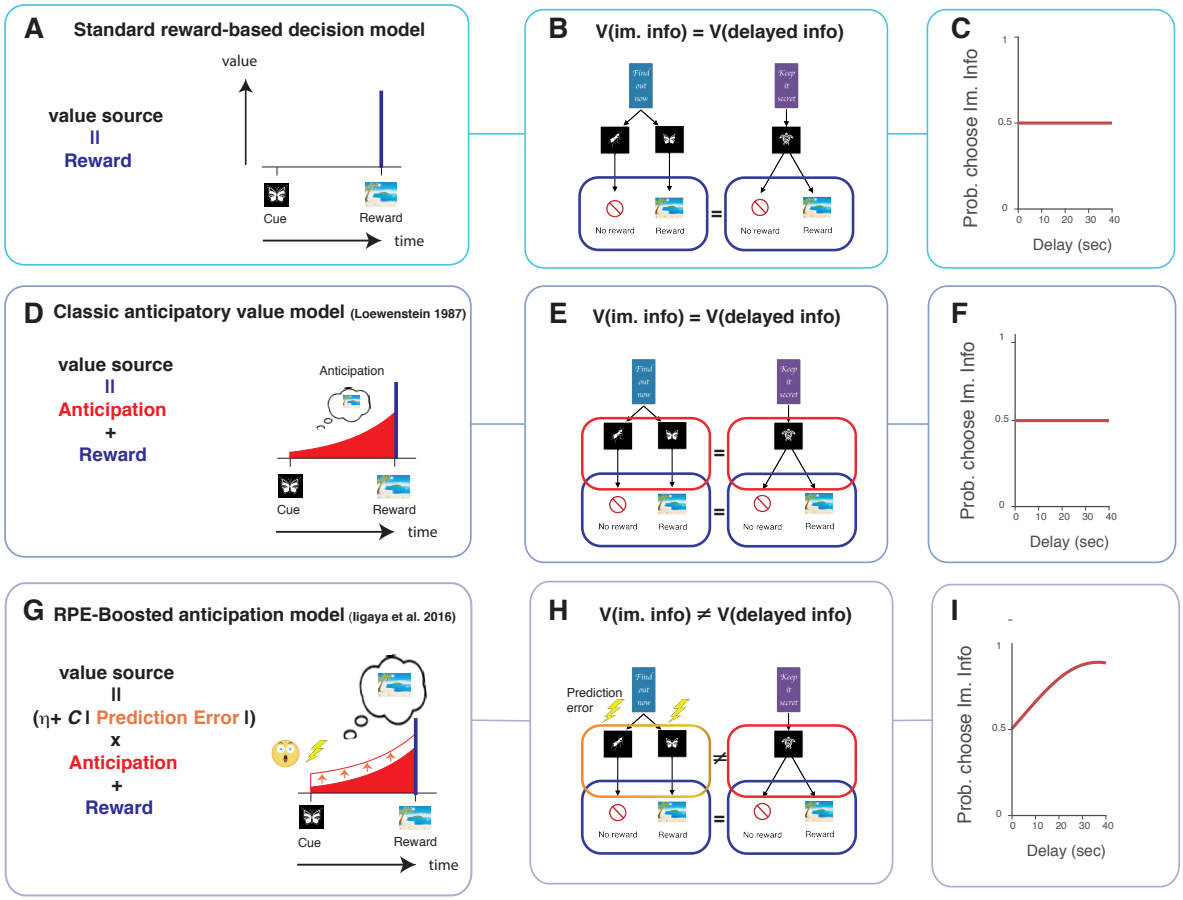

Figure S1: The mechanism by which the model predicts increasing preference for advance information, as a function of the duration of the wait period. (**A,B,C**). A classical reward-based decision-making model. This class of models assumes that a choice is made according to the attractive value of discounted future reward (A). Because there is no difference in the probability of obtaining reward between the Immediate-info-target and the Delayed-Info-target, the model assigns the same value to the two choice targets (B). As a result, the model predicts no preference between the two targets across different delay conditions (C). (**D,E,F**). The classical behavioral economic model of the utility of anticipation.[75] The model assumes that people experience pleasure from the anticipation of future reward, in addition to from consumption of the reward itself. Although this model can capture the well-documented behavior that subjects delay reward consumption, it assigns the same values to the Immediate-info-target and Delayed-info-target in our task (E). As a result, this model also predicts indifferent choice between the two targets (F). (**G,H,I,J**). Our recently proposed model.[53] Inspired by an observation of a dramatic increase in excitement after receiving the information that resolves uncertainty about upcoming reward,[105] this model hypothesizes that the value of anticipation can be boosted by prediction errors associated with the reward predictive cues (G). Consequently, the value of the Immediate-info target can become greater than the value of the Delayed-info target, as the duration of the wait becomes longer (H). As a result, the model predicts that subjects show stronger preference of the Immediate-info-target in longer delay conditions. This model has been previously validated in a series of behavioral experiments[53] (I).

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

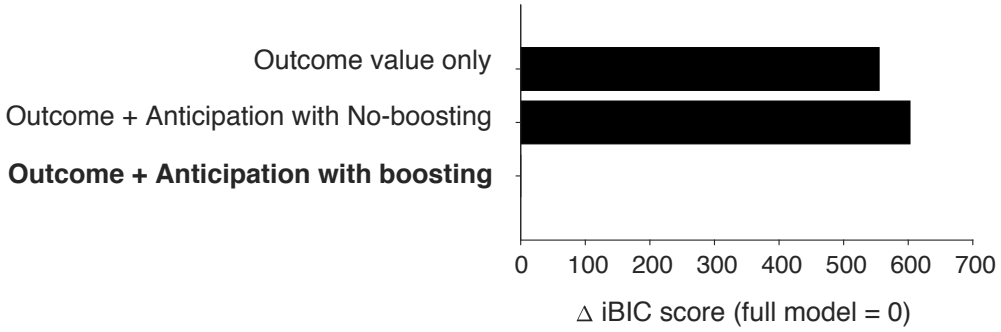

Figure S2: Our model comparison using integrated Bayesian Information Criterion (iBIC) strongly favors our full model with values of outcomes and RPE-boosted anticipation, over a model with outcome values but no anticipatory utility, as well as a model with values of outcomes and anticipation that is not boosted by RPE. All models include temporal discounting. A smaller score indicates a better model. The score is shown in log-scale.

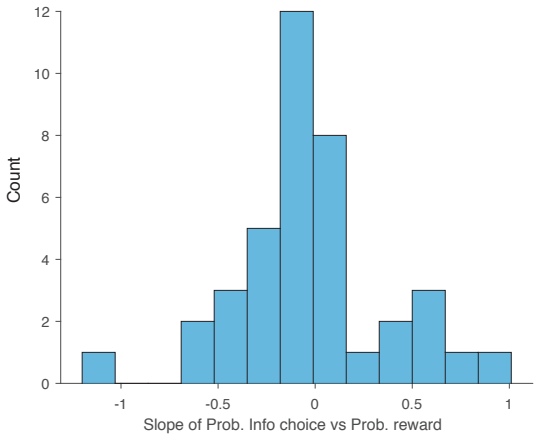

Figure S3: Heterogeneity amongst the participants in how their preferences for Info-target depended on the reward probability. The plot shows a histogram across the subjects of the slope of a linear fit to the probability of choosing the Info-target versus the probability of reward.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

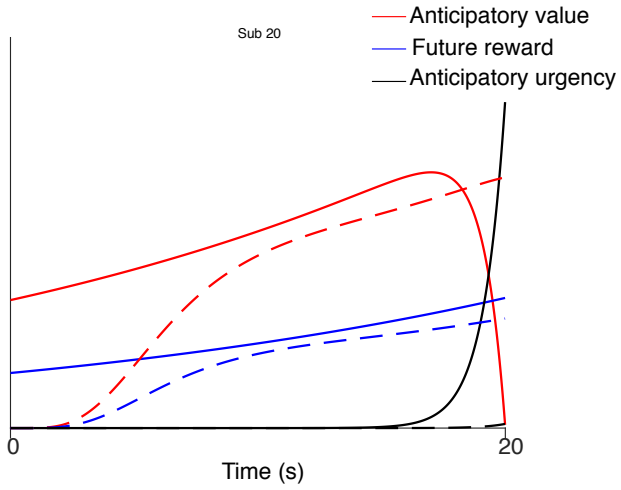

Figure S4: Prediction of fMRI signal in one subject (subjust 20). The anticipatory value signal (red). The discounted reward signal (blue). The anticipation stream signal (black). The dashed curves indicate HRF convoluted predictions for fMRI.

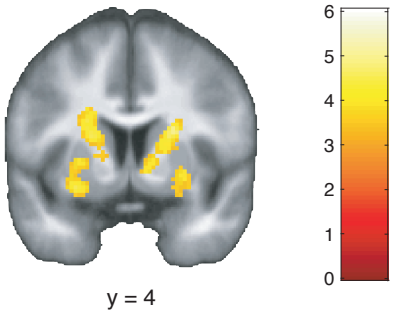

Figure S5: Voxels in dosal caudate correlated with anticipatory value signal. The effects in caudate survived our phase-randomization test ($p < 0.001$ whole-bran FWE correction). The effects in posterior putamen did not survived the whole-brain correction. Voxels at $p < 0.001$ (uncorrelated) are shown for display purposes.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

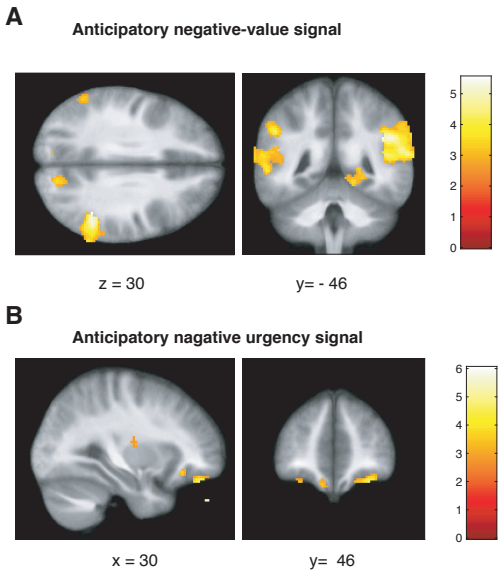

Figure S6: fMRI correlates in negative (no-reward) domain. (**A**) Correlation with anticipatory value signal for no-reward, which is in the negative domain. (**B**) Correlation with anticipatory urgency signal in the negative domain. In all panels, voxels at $p < 0.005$ uncorrected are shown for display purposes.

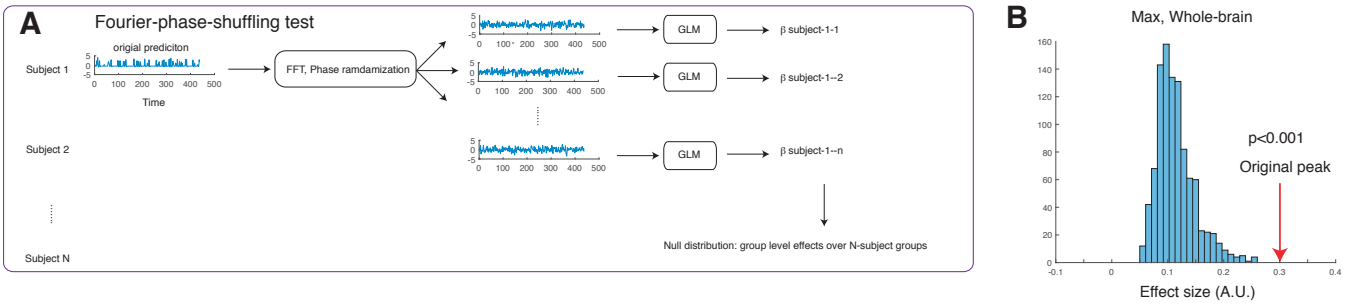

Figure S7: Our control analysis using phase-randomization validates the correlation between the model's anticipatory value signal and the fMRI signal. (**A**) Schematics of the analysis. In response to recent article about false-positive correlations between slow signals in neuroscience,[32] we performed a new analysis using phase-randomization of signals. For this, we first transformed our model's predicted anticipatory signal into the Fourier space. Then we randomized the phase of each frequency without disturbing the power, before transforming back to the original space. We then ran the standard GLM analysis using this regressor as a model's prediction to estimate the regression coefficient. We repeated this for each participant over 100 times (3,900 GLMs in total). We then randomly selected GLM results over participants (one from each participant) to perform a standard second level analysis. We repeated this second level analysis for 1,000 times to create a null distribution of the effect. The null distribution was constructed by taking the maximum correlation over each GLM result, and this was compared against the original. (**B**) Our test shows that our original correlation is significantly greater than by chance, compared to the null distribution constructed by the phase-randomization method ($p < 0.001$)

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

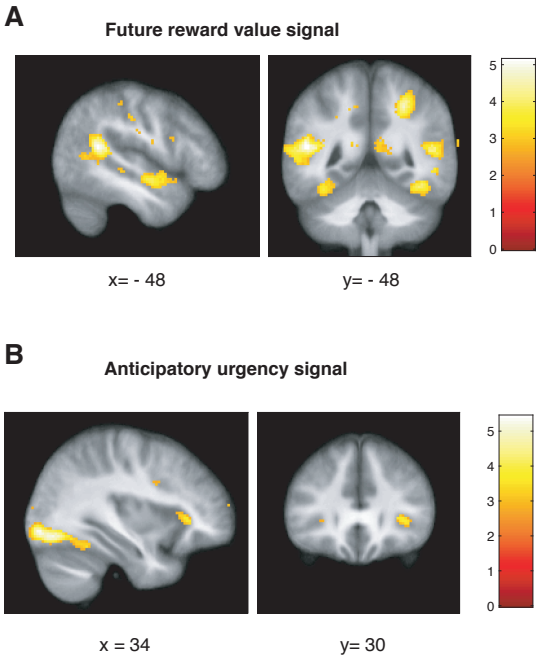

Figure S8: (**A**) Correlations with discounted future reward signal. Regions in superior temporal gyrus survives the whole brain FWE correction $p < 0.05$. (**B**) Correlations with anticipatory urgency signal. In all panels, voxels at $p > 0.005$ are shown for display purposes.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

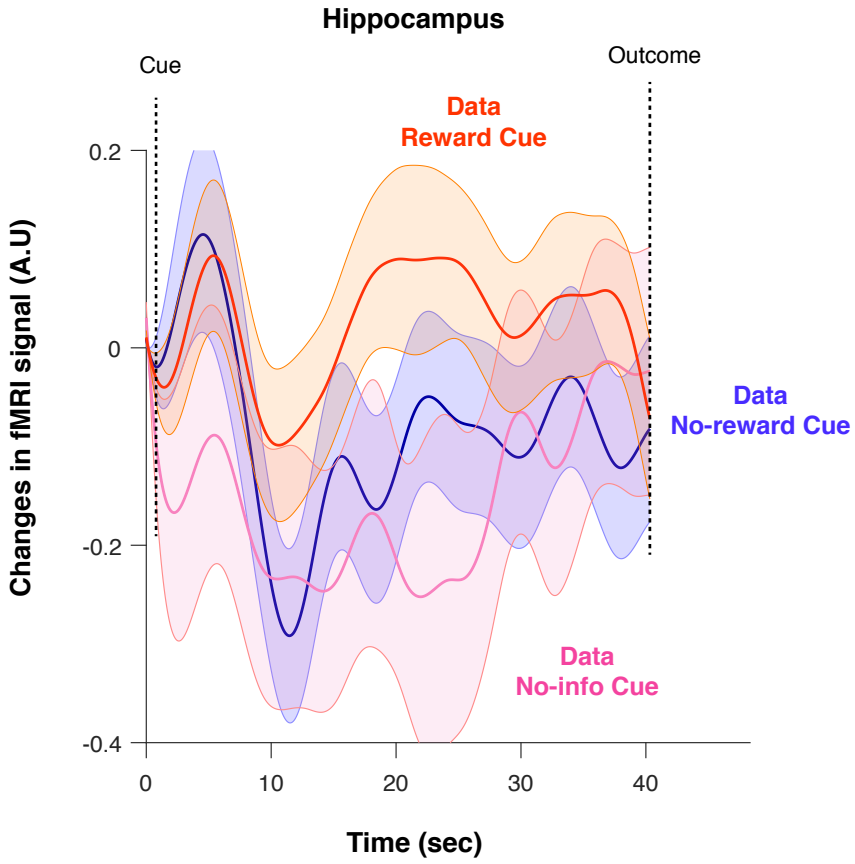

Figure S9: The temporal dynamics of the fMRI signal in hippocampus during anticipatory periods. Changes in activity averaged over participants after receiving a reward predictive cue (orange), after receiving a no-information cue (magenta), and after receiving a no-reward predictive cue (blue) are shown. The error bar indicates the SEM.

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

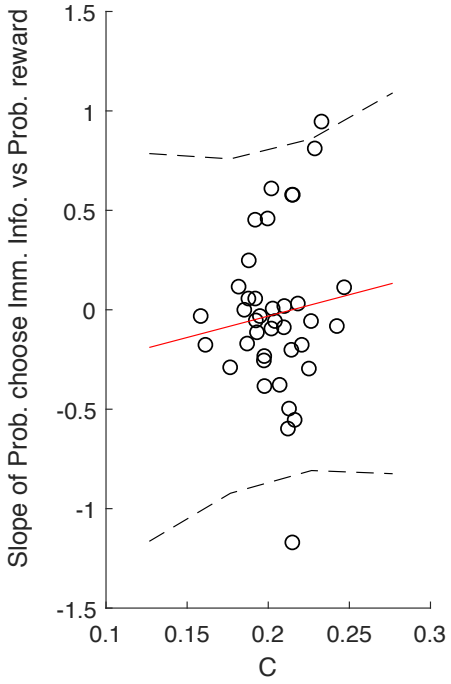

Figure S10: There is no correlation between model's parameter C and the slope of a linear fit to the preference for Immediate Info target vs reward probability ($r = 0.10$, $p > 0.5$).

bioRxiv preprint doi: https://doi.org/10.1101/588699; this version posted March 26, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

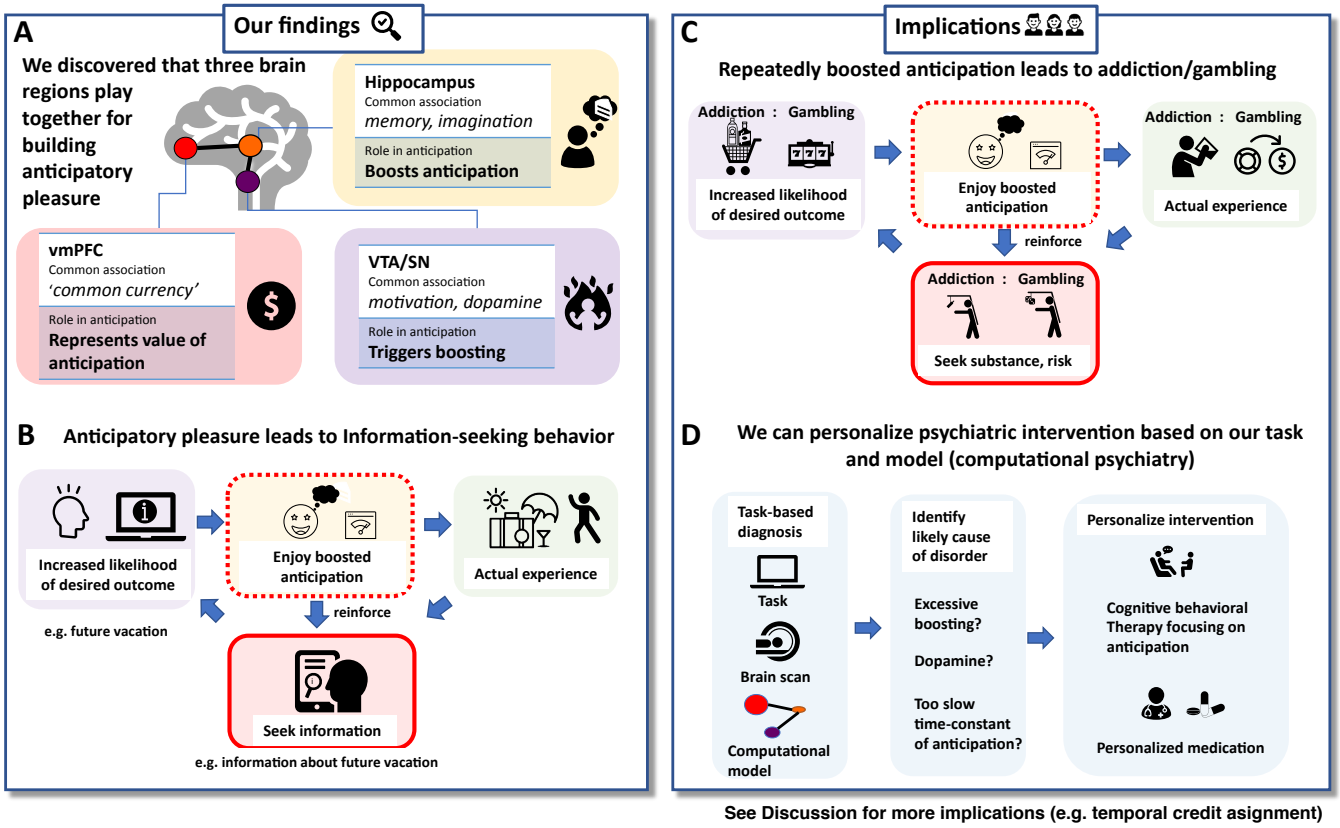

Figure S11: Summary of our findings and implications. (**A,B**). Our primary findings. (**A**). We uncovered three distinctive regions that play together to compute anticipatory pleasure in the brain. The vmPFC, often described as brain's common currency region, represented the value of anticipation. The VTA/SN, regions associated with dopamine and motivation, triggers the boosting of anticipation when the likelihood of desired outcome increases. The hippocampus, a strong associate of memory and imagination, realized the boosting of anticipation. (**B**). We showed that anticipatory pleasure can drive information-seeking behavior. Advance information can increase the likelihood of a desired outcome, which in turn boosts the pleasure of anticipation. As a result, people can feel enhanced pleasure from anticipation after receiving advance information. Therefore people seek advance information of their desired outcomes (information-seeking, or observing), as we confirmed in our current and past experiments.[53] (**C,D**). Implication of our study. (**C**). Over-boosted anticipation could lead to addiction and gambling. Purchasing alcohol or seeing '7-7-7' in a slot machine increases the likelihood of receiving desired outcome (e.g. drinking alcohol, receiving money from gambling). This can boost the anticipatory pleasure. By repeating this many times, the subjective value of alcohol or gambling can also be boosted over and over, leading to pathological seeking for substance (addiction) and risk (gambling). Note that our computational model predicts that this over-boosting can happen only to individuals with particular set of parameter values (e.g. strong boosting and weak discounting). (**D**). Our study can help to design personalized psychiatric interventions (computational psychiatry). Subjects perform a behavioral task in a MRI scanner and we fit our computational model to the behavior. We can identify likely causes of psychiatric disorders (e.g. addiction), by the subject's parameters estimated by our computational model and brain data. This will help design personalized psychiatric intervention, for example cognitive behavioral therapy focusing on aspects of anticipation, as well as medication targeted to specific neuromodulators (e.g. dopamine). Please see the Discussion section for further details and other implications of our study.