Report

Current Biology

Decodability of Reward Learning Signals Predicts Mood Fluctuations

Highlights

- Choices in a week-long reward learning task reveal slow- and fast-learning processes
- Both processes' prediction errors can be decoded from EEG and heart-rate responses
- Greater fast-process decodability predicts positive mood change a few hours later
- Greater slow-process decodability predicts positive mood change one day later

Authors

Eran Eldar, Charlotte Roth,
Peter Dayan, Raymond J. Dolan

Correspondence

e.eldar@ucl.ac.uk

In Brief

In a week-long smartphone experiment, Eldar et al. show that reward-prediction errors indicative of fast and slow reward-learning processes can be decoded from EEG and heart-rate signals. Moreover, fast and slow mood fluctuations are predicted by how well fast and slow learning can be decoded—positive mood changes follow greater decodabilities.

Eldar et al., 2018, Current Biology *28*, 1433–1439
May 7, 2018 © 2018 The Author(s). Published by Elsevier Ltd.
https://doi.org/10.1016/j.cub.2018.03.038

Current Biology
Report

Decodability of Reward Learning Signals Predicts Mood Fluctuations

Eran Eldar,[1,2,5,*] Charlotte Roth,[1,2] Peter Dayan,[3,4] and Raymond J. Dolan[1,2,4]
[1]Max Planck University College London Centre for Computational Psychiatry and Ageing Research, London WC1B 5EH, UK
[2]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, UK
[3]Gatsby Computational Neuroscience Unit, University College London, London W1T 4JG, UK
[4]These authors contributed equally
[5]Lead Contact
*Correspondence: e.eldar@ucl.ac.uk
https://doi.org/10.1016/j.cub.2018.03.038

SUMMARY

Our mood often fluctuates without warning. Recent accounts propose that these fluctuations might be preceded by changes in how we process reward. According to this view, the degree to which reward improves our mood reflects not only characteristics of the reward itself (e.g., its magnitude) but also how receptive to reward we happen to be. Differences in receptivity to reward have been suggested to play an important role in the emergence of mood episodes in psychiatric disorders [1–16]. However, despite substantial theory, the relationship between reward processing and daily fluctuations of mood has yet to be tested directly. In particular, it is unclear whether the extent to which people respond to reward changes from day to day and whether such changes are followed by corresponding shifts in mood. Here, we use a novel mobile-phone platform with dense data sampling and wearable heart-rate and electroencephalographic sensors to examine mood and reward processing over an extended period of one week. Subjects regularly performed a trial-and-error choice task in which different choices were probabilistically rewarded. Subjects' choices revealed two complementary learning processes, one fast and one slow. Reward prediction errors [17, 18] indicative of these two processes were decodable from subjects' physiological responses. Strikingly, more accurate decodability of prediction-error signals reflective of the fast process predicted improvement in subjects' mood several hours later, whereas more accurate decodability of the slow process' signals predicted better mood a whole day later. We conclude that real-life mood fluctuations follow changes in responsivity to reward at multiple timescales.

RESULTS AND DISCUSSION

10 human volunteers reported their mood four times a day, and performed a reward learning task twice a day, for a period of one week (Figures 1A–1C). Overall, each subject completed a total of 2,316 task trials. On each trial, subjects chose between two available images and were rewarded with a coin depending on a reward probability associated with the chosen image. Each of the two daily sessions included two "games" in which trials involving choices between new images and explicit reward feedback ("feedback" trials) were interleaved with trials involving choices between familiar images taken from previous sessions ("no feedback" trials). In each game, the feedback trials involved a set of three images associated with reward with fixed probabilities of 0.25, 0.50, and 0.75. These probabilities were unknown to the subjects and thus could only be learned by trial and error based on obtained rewards. Thus, subjects' performance improved over the course of each game such that by the end of the game, they were choosing the image associated with a higher reward probability 78% of the time (±2% SEM). We tested how well subjects maintained the information they had learned in previous sessions by means of no-feedback trials. In these trials, rewards were administered as before but were not shown to the subject so as to avoid further reward-based learning. Subjects maintained comparable levels of performance on these no-feedback trials even when outcomes associated with the images had not been observed for a period of 3 days (Figure 1D).

Relatively little is known about how learning over a timescale of minutes translates to a timescale of days [19], but previous work suggests that humans might learn separately about short and long timescales [20, 21]. Therefore, we first asked whether the choices a subject made over the course of the experiment reflected a single learning process or, alternatively, an additive combination of multiple learning processes that operate over different timescales. In particular, we compared several learning models in terms of how well each model fitted subjects' choices (see STAR Methods; Figure S1). We found that subjects' choices were best explained by a combination of two learning processes: one that learns quickly but forgets what it has learned by the end of the day and another that learns slowly and does not forget (Figures 2A–2C). In fact, the latter, slower process was best fitted with a negative expectation decay parameter, which entails that what is learned is not only maintained but is actually consolidated or amplified [22] at an average rate of 5.4% per day (Figure 2B). Indeed, the impact of the rewards associated with an image on subjects' choices increased with time even though during this time the image was not associated with additional rewards ($\bar{\beta}_{\text{rwd} \times \text{time}} = 0.12 \pm 0.05$, $p_{\text{boostrap}} = 0.04$, logistic

© 2018 The Author(s). Published by Elsevier Ltd.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

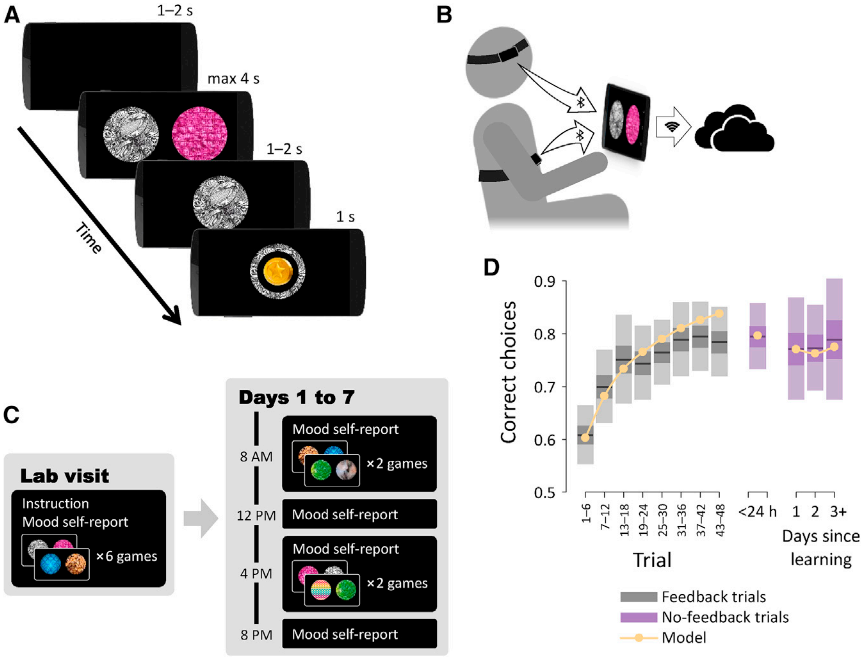

Figure 1. Experimental Task

(A) Subjects chose between two images and either received or did not receive a coin reward depending on the probability associated with the chosen image. Each game included 48 such feedback trials, as well as 24 trials in which outcomes were not revealed (no feedback trials). Every image first appeared on two consecutive sessions with feedback and thereafter only appeared again on no feedback trials.

(B) Subjects performed the experimental task on a smartphone while a chest strap and a headband transmitted heart rate and EEG signals to the phone. Data were then uploaded to a dedicated online server.

(C) Following an initial lab visit, subjects performed two experimental task sessions every day and reported their mood four times a day.

(D) Task performance computed as the proportion of choices of the image associated with a higher reward probability. Also shown is simulated performance of the computational model (see Figure 2). Performance on 'no feedback' trials is shown as function of the time that passed since images appeared with feedback. Shaded areas: SEM (dark) and SD (light).

$n = 10$ subjects.

regression of choices between images about which subjects learned at least one day ago as a function of sum of observed rewards, the average time since these rewards were observed, and their interaction). Importantly, the multi-timescale dynamics captured by the two-process model could not be captured using more complex models that allow for multiple timescales but learn only a single set of expectations (BIC difference of 1377; Figure S1B). Thus, the modeling results revealed fast- and slow-learning processes, each with its own set of learned expectations.

Building on the insight afforded by our learning model about how subjects solved the task, we next asked how subjects' processing of rewards changed from session to session. We first tested variants of the model in which different aspects of the fast- or slow-learning processes were allowed to vary from session to session. These aspects included the learning rates, the decision temperatures, and the subjective value of reward outcomes. We found that variability in subjects' choices across the experiment was best explained by assuming that, for the slow (but not the fast) process, the subjective value of a coin obtained in one session could differ from that of an identical coin obtained in a different session (Figure S1C; Figure 2D). These fluctuations in subjective value during learning explained subjects' later preferences when they were asked to choose between images from different sessions (Figure 2E).

This session-by-session behavioral measure of reward sensitivity, which is based on subjects' choices, did not significantly correlate with subsequent mood changes or with current mood ($p_{\text{bootstrap}} > 0.1$; see STAR Methods). This is despite the fact that subjects' reported mood did vary considerably over the course of the week (mean range 61%; Figure S2B). However, receptivity to reward has at least two aspects. First, there is *sensitivity* [1], which is reported above and which maps objective reward values into subjective utilities. Second, there is *responsivity*, which reflects the attention paid to the dimension of reward and which we operationalized as the degree to which physiological responses (e.g., Figure S3) reflect signals indicative of reward processing. Reward prediction errors, in particular, have been suggested to mediate the emotional impact of reward [3, 23–25], and thus, we next examined whether the reward prediction errors that drove learning according to the model were manifested in subjects' physiological responses and whether this physiological responsivity provided a measure more closely reflective of the dynamics underlying mood changes.

To examine this possibility, we first tested whether physiological responses were consistently modulated by the two elements that compose a prediction error—namely, actual and expected outcome [17, 18]. For this purpose, we computed the average time series of the heart rate (from 1 s before to 10 s after each outcome) and of the EEG signal (from 0.5 s before to 1.5 s after each outcome) during each session for each of six types of outcomes: reward and no-reward outcomes where reward probability was 0.25, 0.50, or 0.75. We then measured the similarity between responses from different sessions (see STAR Methods), and, indeed, we found that physiological responses to the same type of outcome were more similar than responses to different types of outcome (Figures 3A and 3B; Heart rate : $\bar{r}_{\text{same}} = 0.038$, $\bar{r}_{\text{different}} = -0.001$, $p_{\text{bootstrap}} = 10^{-6}$; EEG : $\bar{r}_{\text{same}} = 0.005$, $\bar{r}_{\text{different}} = -0.001$, $p_{\text{bootstrap}} = 10^{-7}$).

This result indicates that the components of the reward prediction error were consistently reflected in subjects' physiological responses. However, this analysis can be improved on in several ways through the medium of the model. First, subjects' expectations of each image were not fixed. Instead, they were dynamically updated as a function of observed outcomes, and the model provides trial-by-trial estimates of these changing expectations. Second, the model indicated that subjects

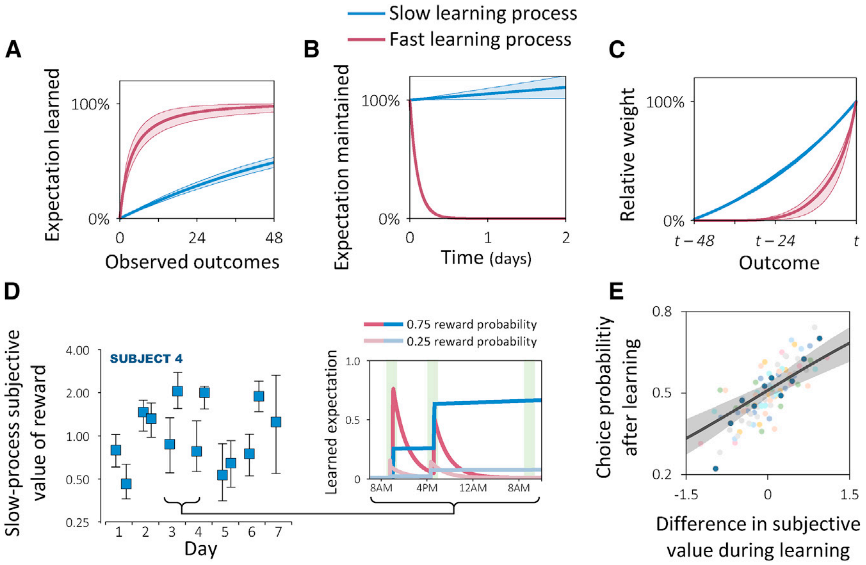

Figure 2. Fast and Slow Learning

Subjects' choices were best explained by a model that involves two simultaneous learning processes. $n = 10$ subjects. see STAR Methods and Figure S1 for details of modeling procedures.

(A) Expectations learned by the two processes given a fixed repeating outcome. Shaded areas indicate spread across subjects (95% interval of fitted group-level distribution).

(B) Decay of expectations as a function of time.

(C) Weights assigned by the two processes to different outcomes as a percentage of the most recent outcome's weight.

(D) Subjective value of reward for the slow process in an exemplar subject. Inset shows the subject's expectations for two images for the slow and fast processes over three sessions (green shading). The images appeared with feedback only in the first two sessions. Error bars: 95% credible interval.

(E) Image choice probability in 'no feedback' trials as a function of the subjective value of reward during learning about the image, minus the subjective value during learning about the alternative image. For visualization, trials were divided into ten quantiles of subjective value differences (each circle represents 10% of a subject's trials). Subjects are color-coded (dark blue: subject from [D]). Choice probabilities are corrected for the number of reward outcomes observed for each image. Shaded areas: 95% bootstrap CI.

maintained two sets of expectations, and therefore, two sets of prediction errors should be reflected in physiological responses. Third, the model indicated that for the slow-learning process, the subjective value of reward also varied, implying that prediction errors should be computed with respect to this subjective value.

To account for these nuances in subjects' learning, we derived trial-by-trial prediction errors for each subject from the fast- and slow-learning processes of the model, with parameters fitted to that subject's choices. We then measured the degree to which each series of prediction errors was reflected in subjects' physiological responses by attempting to decode them from the physiological data using support vector regression with radial basis functions. The degree of success in decoding using this nonlinear method provided us with a single measure of physiological reward prediction error signaling that accounts not only for simple effects of intensity, but also for individualized changes in the shape, timing, and sign of the physiological response. To prevent overfitting in this procedure, we decoded prediction errors for each trial using a decoder trained on a separate set of trials (i.e., using nested cross-validation), and we compared the resultant decoding accuracy to that obtained by applying the same procedure to randomly permuted data (see STAR Methods).

We found that both heart-rate and EEG responses to outcomes reflected the predictions errors generated by the slow- and fast-learning processes of the model (Figure 3C). Moreover, we found that the two components that compose prediction errors, namely actual and expected outcomes, were each separately decodable from subjects' physiological responses (Figure 3D). In addition, combining the decoding from the heart-rate and the EEG responses yielded statistically significant decoding accuracy ($p_{\text{premutation}} < 0.05$) for each individual subject for the slow process, and for 8 out 10 subjects for the fast process. Since both processes learned from the same series of choices and outcomes, and thus their prediction errors were correlated ($\bar{r} = 0.75$, $p_{\text{bootstrap}} < 10^{-5}$), we tested whether decoded prediction errors specifically reflected the learning process from which they were derived. For this purpose, we examined the correlation between the decoded prediction errors of one process and the prediction errors of the other process. We found no such correlations for either the heart-rate or EEG responses (all $\bar{r} < 0.006$, $p_{\text{bootstrap}} > 0.6$). Interestingly, by computing decoding accuracy separately for each experimental session, we found that decoding from heart rate was not significantly correlated across sessions with decoding from EEG, for either the fast ($\bar{r} = 0.09$, $p_{\text{bootstrap}} = 0.34$) or slow ($\bar{r} = 0.03$, $p_{\text{bootstrap}} = 0.71$) processes. However, for each of the two physiological sources, decoding accuracies for the fast and slow processes were correlated with one another (Heart rate : $\bar{r} = 0.28$, $p_{\text{bootstrap}} < 10^{-6}$; EEG : $\bar{r} = 0.24$, $p_{\text{bootstrap}} = 10^{-5}$).

We next tested whether more robust physiological reward-prediction-error signaling (i.e., high responsivity to reward) was followed by improvement in subjects' mood. For this purpose, we tested the relationship between the decodability of prediction errors in a given experimental session and how subjects' self-reported mood changed following the session. Thus, we examined changes in self-reported mood 4 hr following each session, when subjects were next asked to report their mood. In addition, to control for possible diurnal variations in mood [23], we also examined mood 24 hr following each session. Since we were agnostic as to which physiological source (heart rate or EEG) would best reflect future mood change and what timescale of mood change would be reflected (4 or 24 hr), we corrected for the four possible combinations using Bonferroni correction for multiple comparisons. We found that EEG signals reflecting the reward prediction errors derived from the fast process predicted 4-hr mood changes ($\bar{\beta} = 0.025 \pm 0.009$, $p_{\text{bootstrap}} = 0.003$, linear regression controlling for current mood; Figures 4A and 4B),

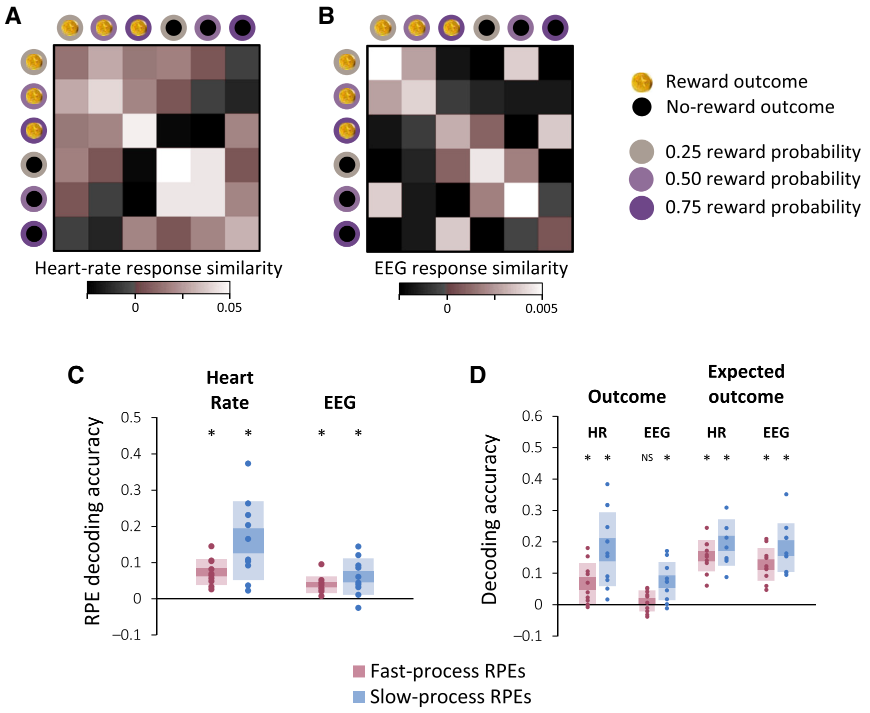

Figure 3. Heart-Rate and EEG Responses to Outcomes

$n = 10$ subjects.

(A) Similarity between heart-rate responses recorded in different sessions following different types of outcomes. Similarity was computed as the average temporal (Pearson) correlation between heart-rate responses for six types of outcomes: reward and no-reward outcomes following choices of images associated with a 0.25, 0.50, or 0.75 reward probability. Similarity was computed separately for each subject and then averaged across subjects.

(B) Similarity between EEG responses recorded in different sessions following different types of outcomes. See Figure S3 for time courses of heart-rate and EEG responses for exemplar subjects.

(C) Reward prediction errors (RPEs) of the fast- and slow-learning processes were decoded from the physiological response to outcomes. The y axis denotes decoding accuracy, computed as the correlation between decoded and actual values. RPEs were derived using the model (see Figure 2) and decoded with cross-validated support vector regression (see STAR Methods).

(D) Correlation between actual and decoded outcomes and between actual and decoded expectations for the slow and fast processes. In (C) and (D), circles correspond to individual subjects. Shaded areas: SEM (dark) and SD (light).

*: $p_{\text{permutation}} < 0.01$, NS: $p_{\text{permutation}} = 0.2$.

whereas EEG signals derived from the slow process predicted 24-hr mood changes ($\overline{\beta} = 0.053 \pm 0.021$, $p_{\text{bootstrap}} = 10^{-4}$; Figures 4C and 4D). In both cases, higher prediction-error decodability predicted more positive mood, and lower decodability predicted worse mood. Neither of these predictive relationships reflected fluctuations in task performance ($p_{\text{bootstrap}} < 0.009$ when including task performance as a control regressor). In contrast, the fast-process signals did not predict 24-hr mood changes ($\overline{\beta} = 0.018 \pm 0.022$, $p_{\text{bootstrap}} = 0.4$; difference from slow process: $p_{\text{bootstrap}} = 0.001$), nor did the slow-process signals predict 4-hr mood changes ($\overline{\beta} = -0.004 \pm 0.009$, $p_{\text{bootstrap}} = 0.7$; difference from fast process: $p_{\text{bootstrap}} = 0.18$). Thus, we found a significant interaction between the timescale of the learning process and the timescale of subsequent mood changes ($p_{\text{bootstrap}} = 0.001$). A complementary analysis involving all time lags up to 24 hr showed similar results (Figure 4E). No such relationship was found between mood changes and the heart-rate signals ($p_{\text{bootstrap}} > 0.1$), which were also not correlated with the EEG signals ($p_{\text{bootstrap}} > 0.1$). These findings establish a striking double dissociation between fast- and slow-learning EEG signals in predicting fast and slow mood fluctuations.

We have shown that responsivity to reward, manifesting as reward prediction error signals in EEG, is predictive of subsequent mood changes. Moreover, this predictive relationship reflects multiple timescales in both reward learning and the dynamics of mood. The finding of multiple timescales adds to previous theoretical accounts of mood as reflecting changes in the availability of reward [3, 26], suggesting that fast and slow changes in mood track short-term and long-term changes in this availability. Future research could investigate whether the distinction between fast- and slow-learning processes evident here reflects the operation of separate brain systems [27, 28] or complex multi-timescale dynamics arising within the same neural population [21, 29]. More importantly, our results show that people's responsivity to reward prediction errors changes from day to day and that greater responsivity is followed by elevated mood, whereas lower responsivity is followed by depressed mood. These findings suggest that day-to-day changes in reward responsivity may play an important role in the generation of natural daily mood fluctuations.

We found leading indicators of changes in mood over two timescales. The precise psychological nature of these indicators, which are based on EEG decoding accuracy, remains to be determined. In the present experiment, these indicators did not consistently reflect task performance (correlation with accuracy: $p_{\text{bootstrap}} \geq 0.27$) nor long-term learning (correlation with model parameter ψ: $p_{\text{bootstrap}} \geq 0.34$). Thus, the processes that impair the accuracy of decoding, the influence of those processes on momentary computations involving reward, and the interaction between these processes and the internal thoughts and external events that can influence subsequent mood become tempting targets for future investigation. Importantly, we note that our results do not rule out the possibility that similar mood-predicting signals also manifest in heart-rate responses or choice behavior (Figures S2C and S2D).

Our EEG measures provide an ecological and scalable means to assess fluctuations in reward responsivity that might prove useful for investigating and predicting how pathological mood episodes evolve—for instance, in major depression and bipolar disorder. The therapeutic effect of existing drug and talk therapies is suggested to reflect their impact on patients' processing

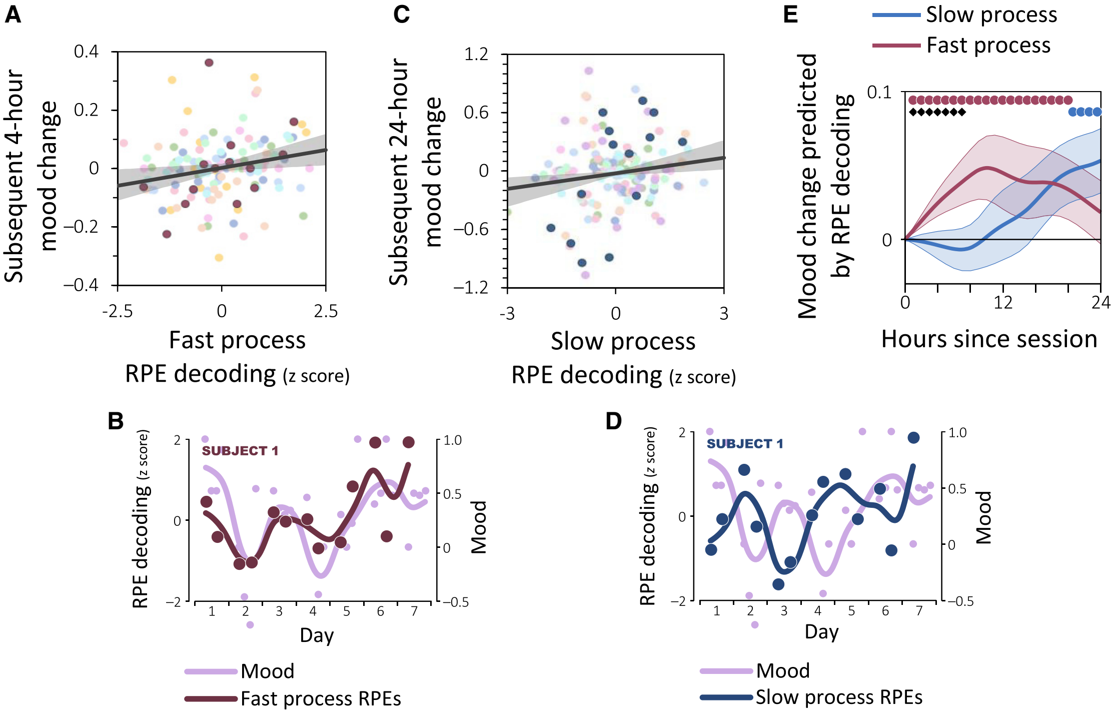

Figure 4. RPEs Evident in EEG and Subsequent Mood Changes

$n = 10$ subjects.

(A and C) Change in self-reported mood as a function of RPE decoding accuracy for the fast- (A) and slow- (C) learning processes. Decoding accuracy was computed separately for each session (denoted by circles). Subjects are color coded, with the subject from (B) and (D) highlighted in dark red (A) and dark blue (C).

(B and D) Relationship between RPE decoding and mood in an exemplar subject for the fast- (B) and slow- (D) learning processes. Shifts in mood follow the fast process's PE signaling almost immediately but substantially lag the slow process's signals.

(E) Average change in mood following each experimental session as a function of reward PE decoding. Magnitude of change is shown per one standard deviation of decoding accuracy. ●: difference from zero, ◆: difference between processes ($p_{\text{corrected}} < 0.05$). Shaded areas: SEM.

See Figures S2C and S2D for a similar analysis with respect to heart-rate responses and reward sensitivity.

of reward [30, 31], and this may serve as a target for the development of new therapeutic approaches.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - ○ Subjects
- METHOD DETAILS
  - ○ Experimental design
  - ○ Mobile platform
  - ○ Daily schedule
  - ○ Experimental task
  - ○ Modeling: learning and forgetting
  - ○ Modeling: multiple timescales
  - ○ Modeling: session-to-session variability
  - ○ Additional alternative models
  - ○ Heart rate data collection
  - ○ Heart rate preprocessing
  - ○ EEG data collection
  - ○ EEG preprocessing
  - ○ Physiological responses similarity
  - ○ Physiological responses decoding
  - ○ Mood self-reports
  - ○ Movement tracking
  - ○ Circle drawing
  - ○ Initial lab visit
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ○ Model fitting
  - ○ Session-by-session parameter fits
  - ○ Trial-by-trial prediction errors
  - ○ Model comparison
  - ○ Physiological responses decoding
  - ○ Regression Analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and two tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.03.038.

ACKNOWLEDGMENTS

We thank Robb B. Rutledge and Jochen Michely for helpful comments on previous versions of this manuscript. This work was funded by a Wellcome

Investigator Award 098362/Z/12/Z (R.J.D.) and the Gatsby Charitable Foundation (P.D.). The Max Planck UCL Centre is a joint initiative supported by UCL and the Max Planck Society. C.R. is currently at the ICM, Hôpital de la Pitié-Salpêtrière. P.D. is on a leave of absence from UCL at Uber Technologies. Neither body was involved in this study.

AUTHOR CONTRIBUTIONS

Conceptualization, E.E.; Methodology, E.E., C.R., and P.D.; Investigation, E.E., and C.R.; Writing – Original Draft, E.E.; Writing – Review & Editing, E.E., P.D. and R.J.D.; Funding Acquisition, R.J.D.; Supervision, P.D., and R.J.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 14, 2017
Revised: March 5, 2018
Accepted: March 16, 2018
Published: April 26, 2018

REFERENCES

1. Huys, Q.J., Pizzagalli, D.A., Bogdan, R., and Dayan, P. (2013). Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biol. Mood Anxiety Disord. *3*, 12.
2. Eldar, E., and Niv, Y. (2015). Interaction between emotional state and learning underlies mood instability. Nat. Commun. *6*, 6149.
3. Eldar, E., Rutledge, R.B., Dolan, R.J., and Niv, Y. (2016). Mood as representation of momentum. Trends Cogn. Sci. *20*, 15–24.
4. Costello, C.G. (2016). Depression: Loss of Reinforcers or Loss of Reinforcer Effectiveness? - Republished Article. Behav. Ther. *47*, 595–599.
5. Gray, J.A. (1994). Framework for a taxonomy of psychiatric disorder. In Emotions: Essays on emotion theory, S.H.M. van Goozen, N.E. Van de Poll, and J.A. Sergeant, eds. (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 29–59.
6. Pizzagalli, D.A. (2014). Depression, stress, and anhedonia: toward a synthesis and integrated model. Annu. Rev. Clin. Psychol. *10*, 393–423.
7. Treadway, M.T., and Zald, D.H. (2013). Parsing anhedonia: translational models of reward-processing deficits in psychopathology. Curr. Dir. Psychol. Sci. *22*, 244–249.
8. Alloy, L.B., and Abramson, L.Y. (2010). The role of the behavioral approach system (BAS) in bipolar spectrum disorders. Curr. Dir. Psychol. Sci. *19*, 189–194.
9. Alloy, L.B., Abramson, L.Y., Urosevic, S., Bender, R.E., and Wagner, C.A. (2009). Longitudinal predictors of bipolar spectrum disorders: A behavioral approach system (BAS) perspective. Clin Psychol (New York) *16*, 206–226.
10. Alloy, L.B., Nusslock, R., and Boland, E.M. (2015). The development and course of bipolar spectrum disorders: an integrated reward and circadian rhythm dysregulation model. Annu. Rev. Clin. Psychol. *11*, 213–250.
11. Depue, R.A., and Iacono, W.G. (1989). Neurobehavioral aspects of affective disorders. Annu. Rev. Psychol. *40*, 457–492.
12. Johnson, S.L. (2005). Mania and dysregulation in goal pursuit: a review. Clin. Psychol. Rev. *25*, 241–262.
13. Johnson, S.L., Edge, M.D., Holmes, M.K., and Carver, C.S. (2012). The behavioral activation system and mania. Annu. Rev. Clin. Psychol. *8*, 243–267.
14. Urosević, S., Abramson, L.Y., Harmon-Jones, E., and Alloy, L.B. (2008). Dysregulation of the behavioral approach system (BAS) in bipolar spectrum disorders: review of theory and evidence. Clin. Psychol. Rev. *28*, 1188–1205.
15. Alloy, L.B., Olino, T., Freed, R.D., and Nusslock, R. (2016). Role of reward sensitivity and processing in major depressive and bipolar spectrum disorders. Behav. Ther. *47*, 600–621.
16. Mason, L., Eldar, E., and Rutledge, R.B. (2017). Mood Instability and Reward Dysregulation-A Neurocomputational Model of Bipolar Disorder. JAMA Psychiatry *74*, 1275–1276.
17. Schultz, W., Dayan, P., and Montague, P.R. (1997). A neural substrate of prediction and reward. Science *275*, 1593–1599.
18. Dayan, P., and Niv, Y. (2008). Reinforcement learning: the good, the bad and the ugly. Curr. Opin. Neurobiol. *18*, 185–196.
19. Wimmer, G.E., and Poldrack, R.A. (2017). Reinforcement learning over time: spaced versus massed training establishes stronger value associations. bioRxiv. https://doi.org/10.1101/158964.
20. Tanaka, S.C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., and Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat. Neurosci. *7*, 887–893.
21. Iigaya, K. (2016). Adaptive learning and decision-making under uncertainty by metaplastic synapses guided by a surprise detection system. eLife *5*, e18073.
22. Karni, A., and Sagi, D. (1993). The time course of learning a visual skill. Nature *365*, 250–252.
23. Rutledge, R.B., Skandali, N., Dayan, P., and Dolan, R.J. (2014). A computational and neural model of momentary subjective well-being. Proc. Natl. Acad. Sci. USA *111*, 12252–12257.
24. Rutledge, R.B., Moutoussis, M., Smittenaar, P., Zeidman, P., Taylor, T., Hrynkiewicz, L., Lam, J., Skandali, N., Siegel, J.Z., Ousdal, O.T., et al. (2017). Association of neural and emotional impacts of reward prediction errors with major depression. JAMA Psychiatry *74*, 790–797.
25. Rutledge, R.B., Skandali, N., Dayan, P., and Dolan, R.J. (2015). Dopaminergic modulation of decision making and subjective well-being. J. Neurosci. *35*, 9811–9822.
26. Mendl, M., Burman, O.H., and Paul, E.S. (2010). An integrative and functional framework for the study of animal emotion and mood. Proc. Biol. Sci. *277*, 2895–2904.
27. Collins, A.G.E., Ciullo, B., Frank, M.J., and Badre, D. (2017). Working memory load strengthens reward prediction errors. J. Neurosci. *37*, 4332–4342.
28. Collins, A.G., and Frank, M.J. (2018). Within-and across-trial dynamics of human EEG reveal cooperative interplay between reinforcement learning and working memory. Proc. Natl. Acad. Sci. USA. *115*, 2502–2507.
29. Iigaya, K., Ahmadian, Y., Sugrue, L., Corrado, G., Loewenstein, Y., Newsome, W.T., and Fusi, S. (2017). Learning Fast And Slow: Deviations From The Matching Law Can Reflect An Optimal Strategy Under Uncertainty. bioRxiv. https://doi.org/10.1101/141309.
30. Stoy, M., Schlagenhauf, F., Sterzer, P., Bermpohl, F., Hägele, C., Suchotzki, K., Schmack, K., Wrase, J., Ricken, R., Knutson, B., et al. (2012). Hyporeactivity of ventral striatum towards incentive stimuli in unmedicated depressed patients normalizes after treatment with escitalopram. J. Psychopharmacol. (Oxford) *26*, 677–688.
31. Burkhouse, K.L., Kujawa, A., Kennedy, A.E., Shankman, S.A., Langenecker, S.A., Phan, K.L., and Klumpp, H. (2016). Neural reactivity to reward as a predictor of cognitive behavioral therapy response in anxiety and depression. Depress. Anxiety *33*, 281–288.
32. Chang, C.C., and Lin, C.J. (2011). LIBSVM: a library for support vector machines. ACM T. Intel. Syst. Tec. *2*, 27.
33. Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. *2011*, 156869.
34. Voon, V., Derbyshire, K., Rück, C., Irvine, M.A., Worbe, Y., Enander, J., Schreiber, L.R.N., Gillan, C., Fineberg, N.A., Sahakian, B.J., et al. (2015). Disorders of compulsivity: a common bias towards learning habits. Mol. Psychiatry *20*, 345–352.

35. Smith, M.A., Ghazizadeh, A., and Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biol. *4*, e179.

36. Bornstein, A.M., Khaw, M.W., Shohamy, D., and Daw, N.D. (2017). Reminders of past choices bias decisions for reward in humans. Nat. Commun. *8*, 15958.

37. Niv, Y., Edlund, J.A., Dayan, P., and O'Doherty, J.P. (2012). Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. J. Neurosci. *32*, 551–562.

38. Giles, D., Draper, N., and Neil, W. (2016). Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur. J. Appl. Physiol. *116*, 563–571.

39. Lim, C.K.A., Chia, W.C., and Chin, S.W. (2014). A mobile driver safety system: Analysis of single-channel EEG on drowsiness detection. In 2014 International Conference on Computational Science and Technology (IEEE), pp. 1–5.

40. Mak, J.N., Chan, R.H., and Wong, S.W. (2013). Evaluation of mental workload in visual-motor task: Spectral analysis of single-channel frontal EEG. In IECON 2013 - 39th Annual Conference of the IEEE Industrial Electronics Society (IEEE), pp. 8426–8430.

41. Crowley, K., Sliney, A., Pitt, I., and Murphy, D. (2010). Evaluating a brain-computer interface to categorise human emotional response. In 2010 10th IEEE International Conference on Advanced Learning Technologies (IEEE), pp. 276–278.

42. Das, R., Chatterjee, D., Das, D., Sinharay, A., and Sinha, A. (2014). Cognitive load measurement-a methodology to compare low cost commercial eeg devices. In 2014 International Conference on Advances in Computing, Communications and Informatics (IEEE), pp. 1188–1194.

43. Vourvopoulos, A., and Liarokapis, F. (2014). Evaluation of commercial brain–computer interfaces in real and virtual world environment: A pilot study. Comput. Electr. Eng. *40*, 714–729.

44. Schölkopf, B., Smola, A.J., Williamson, R.C., and Bartlett, P.L. (2000). New support vector algorithms. Neural Comput. *12*, 1207–1245.

45. Mergl, R., Juckel, G., Rihl, J., Henkel, V., Karner, M., Tigges, P., Schröter, A., and Hegerl, U. (2004). Kinematical analysis of handwriting movements in depressed patients. Acta Psychiatr. Scand. *109*, 383–391.

46. Akiskal, H.S., Mendlowicz, M.V., Jean-Louis, G., Rapaport, M.H., Kelsoe, J.R., Gillin, J.C., and Smith, T.L. (2005). TEMPS-A: validation of a short version of a self-rated instrument designed to measure variations in temperament. J. Affect. Disord. *85*, 45–52.

47. Bishop, C.M. (2006). Pattern Recognition and Machine Learning (Springer).

48. Huys, Q.J.M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., and Roiser, J.P. (2012). Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput. Biol. *8*, e1002410.

49. Eldar, E., Hauser, T.U., Dayan, P., and Dolan, R.J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. Proc. Natl. Acad. Sci. USA *113*, 4812–4817.

50. Kass, R.E., and Raftery, A.E. (1995). Bayes factors. J. Am. Stat. Assoc. *90*, 773–795.

51. Golder, S.A., and Macy, M.W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science *333*, 1878–1881.

52. Storey, J.D. (2002). A direct approach to false discovery rates. J. R. Stat. Soc. B *64*, 479–498.

53. DiCiccio, T.J., and Efron, B. (1996). Bootstrap confidence intervals. Stat. Sci. *11*, 189–212.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
MATLAB 2016a	MathWorks	RRID: SCR_001622
LIBSVM	[32]	RRID: SCR_010243
FieldTrip	[33]	RRID: SCR_004849

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources or raw data should be directed to and will be fulfilled by the Lead Contact, Eran Eldar (e.eldar@ucl.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Subjects

10 human subjects, aged 20 to 29, 8 female, completed the experiment. Subjects were recruited from a subject pool at University College London (UCL). Before being accepted to the study, each subject was asked whether they satisfy any of the study's inclusion or exclusion criteria. Inclusion criteria included fluent English and possession of an Android smartphone that could connect to wearable sensors via Bluetooth Low Energy. Exclusion criteria included age (younger than 18 or older than 30), impaired color discrimination, use of psychoactive substances (e.g., psychiatric medications), and current neurological or psychiatric illness. Subjects received £10 per day for participation and 6 pence for each coin they collected in the experimental task, which together added up to an average sum of £151.04 (±£1.77 SD). The experimental protocol was approved by the University of College London local research ethics committee, and informed consent was obtained from all subjects.

METHOD DETAILS

Experimental design

To study the temporal relationship between reward responsivity and mood, we had subjects regularly report their mood, while also performing a reward learning experimental task, over a period of one week, using a mobile phone platform that we developed for this purpose. Since we aimed to characterize a longitudinal process that manifests in most people, we opted to study a relatively small group of 10 subjects but to collect a very large dataset from each. Thus, each subject made 2316 choices in the experimental task while their physiological responses were being recorded. Due to the novelty of the experimental measures, this sample size was not determined based on a quantitative power analysis. However, the amount of data collected from each subject was an order of magnitude greater than the amount of data that comprise a typical learning study. Due to the size of this dataset, we exercised particular caution in determining whether subjects' physiological responses reflected reward prediction errors. To this end, we separated between training and testing data, we tested statistical significance using permutation tests, and we replicated the finding of physiological prediction error signals separately for each individual subject (see Physiological responses decoding below). In addition, due to the relatively small number of subjects, we only tested main effects across the whole study sample.

Mobile platform

To allow a longitudinal study of reward learning processes, associated physiological responses, and their interaction with mood fluctuations, we developed an app for Android smartphones using the Android Studio programming environment (Google, Mountain View, CA). The app asks users to perform experimental tasks according to a pre-determined schedule, while it records electroencephalographic (EEG) and heart rate signals derived from wearable sensors connected using Bluetooth. Additionally, we equipped the app with additional features designed to probe changes in a person's mental state, including regular mood self-report questionnaires and life events and activities logging. Motor activity was tracked via the phone's accelerometer and global positioning system (GPS). Subjects also completed a temperamental trait questionnaire. All behavioral and physiological data were saved locally on the phone as SQLite databases (The SQLite Consortium), which were regularly uploaded via the phone's data connection to a dedicated cloud storage space.

Daily schedule

Subjects first visited the lab to receive instructions, test the app on their phones, and try out the experimental task (see Initial lab visit section below). Starting from the next day, subjects performed two experimental sessions a day, one in the morning and one in the evening, over a period of 7 days. Each session began with a 5-minute heart rate measurement during which subjects were asked to remain seated, report their mood, and perform a circle drawing task (see details below). Following this, subjects put on the EEG sensor and played two games of the experimental task. The app allowed subjects to perform the morning session from 8AM and the evening session from 4PM. In addition, subjects were asked to report their mood twice more, at 12PM and 8PM. Subjects were allowed to adjust the timing of the sessions according to their daily schedule, but were required to ensure a gap of at least 4 hr between successive sessions. On average, subjects performed the morning session at 9:06AM (mean SD ± 25 min) and the evening session at 5:21PM (mean SD ± 32 min), and provided additional mood self-reports at 12:44PM (mean SD ± 19 min) and 20:23PM (mean SD ± 60 min). One subject was not able to perform the experiment on Day 2, and thus all her subsequent tasks were postponed by one day.

Experimental task

To test for fluctuations in reward processing, we had subjects perform regularly a trial-and-error learning task over a period of one week. On each trial, subjects chose from one of two available images, and then collected a coin reward with a probability associated with the chosen image (Figure 1A). Each game consisted of 48 such trials involving a set of 3 images with reward probabilities of 0.25, 0.5 and 0.75. The probabilities were never revealed to the subjects, though subjects were instructed that each image was associated with a fixed probability of reward. Subjects played four games a day, two during the morning session and two during the evening session.

To examine changes in subjects' learning throughout the week, we had each image appear with reward feedback only in two successive sessions. This way, subjects learned about each given image during a specific part of the week, and this allowed us to probe fluctuations in the effect of learning by later asking subjects to choose between images they had learned about during different parts of the week. To prevent new learning during this probing, outcomes were not revealed on such trials but subjects were informed that they would be rewarded for their choices as before (at the end of the entire experiment). Each game included 24 such no-feedback trials (every 3rd trial), 12 of which involved choosing between images associated with the same reward probability. The no-feedback trials primarily allowed us to measure subjects' rate of forgetting, since they involved familiar images re-appearing following variable lags after subjects had learned about them. In addition, the interleaving of feedback trials involving new images with no-feedback trials involving familiar images allowed us to dissociate fluctuations in how subjects learned from fluctuations in how subjects formed their decisions (see Modeling sections below).

In the first two days, familiar images were taken from the session performed during the initial lab visit. Thereafter, familiar images were those subjects learned about during the week. The app dynamically populated the no-feedback trials of each game to ensure the following criteria: 1. No pair of images appeared more than once in a given game. 2. The app prioritized pairs of images that had previously appeared fewer times. 3. Out of pairs that had appeared a similar number of times, the app prioritized pairs of images about which the subject learned in dissimilar moods. To compute mood during learning about a given image, the app computed the average timing of all revealed outcomes associated with the image and linearly interpolated between the mood self-reports preceding and following this timing. The last 4 games of the experiment consisted solely of no-feedback trials involving familiar images, with 48 such trials per block. Thus, the evening session on Day 7 consisted of 3 such games, and another such game was played prior to that in Day 7's morning session.

Modeling: learning and forgetting

To identify the computations that guided subjects' choices in the experimental task we compared a set of computational models in terms of how well each model fitted subjects' choices. We were first interested in determining how subjects learned from the outcomes associated with each image, and whether the learned information decayed as a function of time.

To this end, we compared the following four models:

Model 1 (fixed learning; Equations 1, 2, and 3) learns the expected value of each image by adjusting its expectation following each outcome as follows:

$$Q_{t+1}(s_{t,c_t}) = Q_t(s_{t,c_t}) + \eta\delta_t, \tag{Equation 1}$$

where s_{t,c_t} is the image chosen at trial t, $Q_t(s)$ is the expected outcome for image s (initialized as $Q_0(s) = 0$), η is a fixed learning rate parameter between 0 and 1, and δ_t is the prediction error at trial t:

$$\delta_t = R_t - Q_t(s_{t,c_t}), \tag{Equation 2}$$

computed as the difference between the outcome and the expected outcome (a reward outcome corresponds to $R = 1$ and no-reward to $R = 0$). On each trial, the model chooses either the left image $(c_t = 1)$ or the right image $(c_t = 2)$, according to the expectations it has learned:

$$p(c_t = i) = \frac{e^{\beta Q_t(s_{t,i})}}{\sum_{j=1}^{2} e^{\beta Q_t(s_{t,j})}}, \tag{Equation 3}$$

where $s_{t,1}$ and $s_{t,2}$ are the left and right images, respectively, the subject can choose on trial t, and β is an inverse temperature parameter.

Model 1 has a fixed learning rate, and thus, it assigns greater weight to more recent outcomes (i.e., ‘leaky integration’). In contrast, Model 2′s (dynamic learning; Equations 2, 3, 4, and 5) learning rate changes as a function of the number of observed outcomes (N_t) for the chosen image:

$$Q_{t+1}(s_{t,c_t}) = Q_t(s_{t,c_t}) + \alpha_t \delta_t, \tag{Equation 4}$$

$$\alpha_t = \frac{1}{\varepsilon + N_t(s_{t,c_t})}, \tag{Equation 5}$$

where ε is a free parameter that determines the initial learning rate. Here, the learning rate gradually decreases asymptotically toward zero so as to compute an average of observed outcomes in which all outcomes are similarly weighted. $\varepsilon > 0$ slows down initial learning, and its impact is similar to that of a prior expectation that all expected outcomes equal Q_0, with the precise value of ε reflecting the strength of this prior.

Model 3 (‘fixed + dynamic learning’; Equations 2, 3, 4, and 6) combines Models 1 and 2 in that its learning rate is composed of fixed and changing components, implying that the learning rate gradually decreases to an asymptote that is larger than zero:

$$\alpha_t = \eta + \frac{1-\eta}{\varepsilon + N_t(s_{t,c_t})}, \tag{Equation 6}$$

Model 4 (‘fixed learning + decay’; Equations 1, 2, 3, and 7), Model 5 (‘fixed & dynamic learning + decay’; Equations 2, 3, 4, 5, and 7), and Model 6 (fixed & dynamic learning + decay’; Equations 2, 3, 4, 6, and 7) are similar to Models 1, 2, and 3, except that expectations decay back to zero as a function of time, both during and in between sessions. To implement this decay, we updated all model expectations at the beginning of every trial as follows:

$$Q_t(s) \leftarrow Q_t(s) e^{-\gamma(T_t - T_{t-1})}, \tag{Equation 7}$$

where T_t is the time at trial t, measured in units of days, and γ determines the rate of decay.

Out of these six models, we found that the model that best fitted subjects’ choices was Model 6 (‘fixed & dynamic learning + decay’), and thus, in the next step we tested variants of this basic model.

Modeling: multiple timescales

Since learning within a single session, over a timescale of minutes, might involve different processes than learning over a whole week, we tested whether subjects’ choices could be better explained by allowing the model to operate over two different timescales. For this purpose, we compared Model 6 with a combination of two such models, each with its own set of expectations (Q and Q') and parameters (η and η', ε and ε, γ and γ', β and β'). This combined model (Model 7, ‘Two dynamic-learning processes’; Equations 2, 4, 6, 7, and 8) simultaneously learns two sets of expectations, updating both in the same manner but with different learning and decay rates. Importantly, in the iterative model fitting procedure described below (Model Fitting subsection), the learning and decay rate parameters of the two processes spontaneously differentiated so as to form one fast process and one slow process. The model forms its decisions by combining the two sets of expectations:

$$\mathrm{p}(c_t = i) = \frac{e^{\beta Q_t(s_{t,i}) + \beta' Q'_t(s_{t,i})}}{\sum_{j=1}^{2} e^{\beta Q_t(s_{t,j}) + \beta' Q'_t(s_{t,j})}}. \tag{Equation 8}$$

Model 8 (’two processes: dynamic + fixed; Equations 1, 2, 4, 6, 7, and 8) is a variant of Model 7, also involving two independent learning processes, except that in this model one of the processes has a fixed learning rate (as in Equation 1).

Since Models 7 and 8 fitted subjects’ choices significantly better than a single-process model, we next tested whether an additional set of expectations was indeed necessary. To this end, we tested whether subjects’ choices can be better fitted with more complex single-process algorithms that allow for multiple timescales but only maintain a single set of expectations. Specifically, we designed the following four models:

Model 9 (‘single process: multiple learning dynamics’; Equations 2, 3, 4, 7, and 9) allows for more complex learning-rate dynamics, since its learning is composed of one fixed component (η) and two separate dynamic components (ε and ε'):

$$\alpha'_t = \eta + \frac{\omega(1-\eta)}{\varepsilon + N_t(s_{t,c_t})} + \frac{(1-\omega)(1-\eta)}{\varepsilon' + N_t(s_{t,c_t})}, \tag{Equation 9}$$

where $0 < \omega < 1$.

In Model 10 (‘single process: multiple forgetting dynamics’; Equations 2, 3, 4, 6, and 7) expectations decay at a different rate within (γ) and between (γ') sessions and, in addition, the expected value of each image is multiplied by a positive factor (γ'') once learning about the image concludes.

Model 11 (‘single process: multiple decision temperatures’; Equations 2, 3, 4, 6, and 7) forms its decisions with different inverse temperature parameters (β and β') depending on whether the trial involves new images (β; ‘feedback’ trials involving images the model is still learning about) or familiar images (β'; ‘no feedback’ trials involving images about which learning has concluded).

Model 12 (‘single process: two full sets of parameters’; Equations 2, 3, 4, 7, and 9) combines all of the enhancements featured by Models 12 to 14.

We found that none of the single-process models fitted subjects’ choices nearly as well as the best two-process model (Model 8) and therefore we used Model 8 as a basis for the last model comparison.

Modeling: session-to-session variability

In the models described so far, all parameters of an individual subject are sampled from a group-level distribution and remain fixed throughout the subject’s sessions. To test whether (and in what way) a subject performed the task differently in different sessions, we compared Model 8 with six variants of this model in which either the learning rate, or the subjective value of reward outcomes during learning (modeled as $R_t' = \psi R_t$), or the inverse decision temperature (β), was allowed to vary across sessions for one of the learning processes. For this purpose, for the value of the variable parameter was determined as before, but was then multiplied by a session-specific scaling parameter. The natural logarithm of this scaling parameter was sampled from a subject-specific normal distribution with a zero mean and a standard deviation that was sampled from a group-level gamma distribution.

We found that the model that best fitted subjects’ choices was the model with variable subjective value of reward for the slow process (Model 18). Since this subjective value is learned, it has a lasting impact in future sessions when the probe images are presented without feedback. We used this model for all results displayed in the main text, and we produce its graphical model in Figure S4.

Additional alternative models

Along with the model comparisons described above, we tested Model 18 against several additional alternative models that did not fit subjects’ choices as well. These included variants of Model 18 with the addition of a fixed choice bias (iBIC = 21085) or a perseveration bias (iBIC = 21084) [34], or where the expectations of the slow and fast processes are combined to form a single prediction (and thus lead to a single prediction error; iBIC = 21155) [35], a model that makes choices based on sampling of previously observed outcomes [36], where the probability of sampling an observation decays with time according to a power law (iBIC = 24310), a model that allows for asymmetry in the rate of learning from positive and negative prediction errors (iBIC = 21083) [37], and a model that uses Bayesian inference to determine which one of the three possible reward probabilities is associated with each stimulus (iBIC = 24022). Equations describing these algorithmic elements are provided elsewhere.

Heart rate data collection

Inter-heart-beat intervals were recorded using a Polar H7 chest strap (Polar Electro, Kempele, Finland). The chest strap senses and analyzes electrocardiographic (ECG) signals, and reports the detected inter-beat (R-R) intervals as well as a derived heart rate measurement once every second via Bluetooth Low Energy (BLE). Its measurements have been shown to be highly reliable in comparison with clinical ECG (error rate lower than 0.01%; intra-class correlation coefficient (ICC) > 0.97) [38]. To ensure that subjects started the experimental task at a relatively similar state of rest, subjects wore the heart rate sensor 5 min prior to each session during which a resting heart rate measurement was taken. Subjects were asked to remain seated throughout this time as well as while performing the task. The app allowed subjects to perform the experimental task only while heart-beat intervals were being received and the sensor’s heart rate measurement was not lower than 30 or higher than 250. Due to communication errors and conflicts between the experimental app and the other apps installed on the subjects’ phones, heart rate data from 5.0% of trials were not saved.

Heart rate preprocessing

All data analysis was carried out in MATLAB (Mathworks, Natick, MA). Since the heart rate sensor sends a message once every second, we first identified and corrected the timing of messages that were received with a delay of more than 100 ms. Correction was applied only when the delay affected a single isolated message and thus there was no ambiguity with respect to its correct timing. The timing of each message indicates a window of one second within which the inter-beat intervals reported in the message have concluded. To time heart beats more precisely, we found the timings that best minimize the discrepancy between the cumulative sum of consecutive inter-beat intervals and the timings of the messages containing these intervals. This procedure narrowed down the timing of each heart beat to a 4.5 ms window on average, the center of which was considered as the heart beat’s precise time. We then converted the sequences of precisely timed intervals into unsmoothed 20 Hz heart-rate signals, where the heart rate at any given moment is estimated as the inverse of the corresponding interval. The heart rate response to an outcome in the experimental task was assessed based on the heart rate signal recorded from 1 s preceding the outcome to 10 s following the outcome. This provided one 221-feature vector per outcome. Features were z-scored across trials and used for the decoding analyses (see Decoding below). Heart rate responses for which the standard deviation of the signal across time was higher than 5 times the median standard deviation (0.6% of responses) were considered noisy and excluded from further analysis.

EEG data collection

EEG was recorded during the experimental task using Brainlink Lite (Neurosky, Hong Kong), a single-channel 512Hz EEG headband. The headband senses electrical signals via 3 dry electrodes placed on the forehead, and reports 512 raw measurements per second as well as several derived measures via Bluetooth. Signals recorded using similar sensors from the same manufacturer have been shown to successfully discriminate subjects’ cognitive and affective states in a range of scenarios [39–43]. The app allowed subjects to perform the experimental task only while EEG data were being received and the sensor's signal-quality assessment was lower than 50 (on a scale of 0 and 100, where lower is better). Due to communication issues and software conflicts, EEG data from 1.0% of the trials were not saved.

EEG preprocessing

The EEG response to an outcome in the experimental task was assessed based on the EEG signal recorded from 500 ms preceding the outcome to 1500 ms following the outcome. EEG responses for which the standard deviation of the signal across time was higher than 5 times the median standard deviation (0.3% of responses) were considered noisy and excluded from further analysis. Time-frequency analysis of the EEG responses was performed using the FieldTrip toolbox [33] multitaper method with 4-cycle-long Hanning windows for the following eight frequencies: 1, 5, 9, 13, 17, 21, 25, and 29 Hz. Frequencies higher than 30 Hz were excluded so as to mitigate the effects of muscle artifacts. The resulting time-series were downsampled to 15 Hz, providing one 353-feature vector per outcome. These vectors were *z*-scored across trials and used for the decoding analyses.

Physiological responses similarity

We tested how consistently outcomes and expectations affected subjects' heart rate and EEG responses by examining the degree to which physiological responses from different sessions were correlated. To isolate the effect of outcomes and expectations on physiological responses, we *z*-scored responses within each session across trials, and then computed the average response for 6 types of outcomes: reward and no-reward outcomes following choices of three types of image (reward probabilities 0.25, 0.50 and 0.75). Consistency of both heart rate and EEG responses were measured within and between subjects as the average pairwise temporal correlation between responses to similar types of outcomes from different sessions.

Physiological responses decoding

To test whether heart rate and EEG responses to outcomes reflected a subject's prediction errors (which were inferred using the model from the subject's choices), we trained and tested support vector machines that decoded these prediction errors from the physiological responses. To avoid over-fitting, training and testing were performed on separate sets of trials following a 5-fold cross validation scheme. Training and testing sets were stratified such that the different sets included similar distributions of prediction errors. This analysis was performed using LIBSVM's implementation [32] of the ν-SVR algorithm [44], whose parameters were fitted to each training set using a nested 5-fold-cross-validated grid search among the following settings: ν = [0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9] and C = [0.25 0.5 1 2 4]. Decoding accuracy was computed as the correlation between actual and decoded prediction errors.

Mood self-reports

The app regularly asked subjects to rate on an analog scale how well they were feeling (Figure S2A). Naturally, subjects did not report their mood at precisely the same times. Consequently, to assess a subject's mood at a particular time of interest, we computed a weighted average of all the subject's mood ratings with weights determined by a Gaussian filter centered on the time of interest with a 4-hour standard deviation (approximating the time between mood reports). In addition, following each mood rating, subjects had to report at least one event or activity that may have affected their mood since the last time they were asked, as well as how strong this effect was and whether it was good or bad. No subject reported that performing the experimental task affected their mood. Finally, subjects were also asked to predict how well they expect to feel over the next several hours.

Movement tracking

The app tracked subjects' movement throughout the week by means of the phone's accelerometer. Movement data were recorded in terms of number of steps and distance with a temporal resolution of 0.2 Hz (except for one subject whose phone did not allow that). Subjects were asked to carry their phones with them at all times unless they were engaged in an activity that precludes that (e.g., swimming). Movement exceeding 20 m or 20 steps was detected during only 6 games out of the 350 games that subjects played in total.

Circle drawing

At the beginning of each session, we asked subjects to trace a circle (15 mm diameter) with their thumb as many times as possible for a period of 30 s. This task was modeled after Mergl et al. [45] who showed that patients with depression differ from healthy controls in the kinematics of their strokes. This raises the possibility that stroke regularity could serve as an implicit measure of a person's mood state. However, we did not analyze performance on this task since subjects reported that the kinematics of their strokes were significantly affected by how moisturized their hands happened to be at the time (this was not an issue in the original study since there a pen was used for this task).

Initial lab visit

Subjects first arrived at the Welcome Trust Centre for Neuroimaging in University College London to receive instructions and have the app installed and tested on their phones. In the lab, subjects played 6 games, each consisting of 48-feedback trials involving a unique 3-image set. In addition, the last three games included 12 no-feedback trials (every 5th trial) involving familiar images from the first three games. Subjects also performed the circle drawing task once in the lab, and filled out one mood self-report and a standardized questionnaire (short version of TEMPS-A) designed to measure five temperamental traits (cyclothymic, dysthymic, irritable, hyperthymic, and anxious) [46]. Before allowing subjects to perform the experiment for a whole week, we verified that subjects succeeded in choosing images associated with higher reward probabilities at above-chance levels, and that the heart-rate and EEG data were recorded and saved to the cloud without significant losses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Model fitting

To fit the parameters of the different models to subjects' decisions, we used an iterative hierarchical expectation-maximization procedure [47]. We first sampled 10^4 random settings of the parameters from predefined group-level prior distributions. Then, we computed the likelihood of observing subjects' choices given each setting, and used the computed likelihoods as importance weights to re-fit the parameters of the group-level prior distributions. These steps were repeated iteratively until model evidence ceased to increase (see Model Comparison below for how model evidence was estimated). This procedure was then repeated with $10^{4½}$ samples per iteration, and finally with 10^5 samples per iteration. To derive the best-fitting parameters for each individual subject, we computed a weighted mean of the final batch of parameter settings, in which each setting was weighted by the likelihood it assigned to the individual subject's decisions. Fractional parameters (η, η', ω) were modeled with Beta distributions (initialized with shape parameters $a = 1$ and $b = 1$). Expectation decay rates (γ, γ') and decision parameters (β, β', λ) were initially modeled with normal distributions (initialized with $\mu = 0$ and $\sigma = 1$) to allow for both positive and negative effects, but were then re-fitted with Gamma distributions if all fitted values were positive. All other parameters were modeled with Gamma distributions (initialized with $k = 1$, $\theta = 1$).

Session-by-session parameter fits

To estimate the best-fitting setting for ψ for each session of a given subject, we sampled 10^5 random settings from its posterior distribution given the fitted group-level prior and all of the subject's choices. We then computed a weighted mean of the 10^5 parameter settings, where the weight of each setting was determined by the likelihood it assigned to the subject's choices on all 'feedback' trials within the session as well as on 'no feedback' trials from subsequent session that involved images about which subjects learned during the session.

Trial-by-trial prediction errors

We derived reward prediction errors for each observed outcome by instantiating the model with the parameter settings that best fitted the individual subject's choices.

Model comparison

We compared between pairs of models in terms of how well each model accounted for subjects' choices by means of the integrated Bayesian Information Criterion (iBIC) [48, 49]. To do this, we estimated the evidence in favor of each model ($\mathcal{L}$) as the mean likelihood of the model given 10^5 random parameter settings drawn from the fitted group-level priors. We then computed the iBIC by penalizing the model evidence to account for model complexity as follows: iBIC $= -2 \ln\mathcal{L} + k \ln n$, where k is the number of fitted parameters and n is the number of subject choices used to compute the likelihood. Lower iBIC values indicate a more parsimonious model fit, and the log Bayes Factor [50] comparing two models can be estimated as their iBIC difference divided in half. We validated this model comparison procedure by generating simulated data using each model, and applying our model comparison procedure to recover the model that generated each dataset (see Table S2).

Physiological responses decoding

Statistical significance of decoding accuracies was measured using a one-tailed permutation test. For this purpose, we generated a null distribution based on 100 random permutations of the data, permuting each subject's behavior-derived prediction errors with respect to that subject's physiological responses. We then applied the full decoding procedure to each permutated dataset and measured the resulting accuracy.

Regression Analyses

We used linear regression to test the relationship between reward-prediction-error decoding from the physiological responses to outcomes during an experimental session and mood change following the sessions. For this purpose, we examined how mood changed 4 hr after each experimental session, when subjects were next asked to report their mood. In addition, to control for diurnal variations in mood [51], we examined how mood changed 24 hr following each session. To account for possible 'regression to the mean' effects, we included the current level of mood (i.e., during the experimental session) as a control regressor. To control for multiple comparisons for the two physiological source (heart rate or EEG) and the two timescales of mood change (4 or 24 hr), results were considered statistically significant below a Bonferroni-corrected threshold of $p = 0.0125$. A complementary analysis similarly assessed the relationship between reward-prediction-error decoding and subsequent mood change following any integer number of hours between 1 and 24. Here we corrected p values for multiple comparisons across all possible lags using false-discovery-rate (FDR) adjustment [52].

Logistic regression was used to test the relationship between the subjective value of reward during a session in which an image appeared with reward feedback and later choices involving the image. Here the number of observed rewards for each image served as a control regressor. For both types of regression, statistical significance of regression coefficients was computed at the group level using a two-tailed bias-corrected and accelerated bootstrap test [53] with default MATLAB options.

DATA AND SOFTWARE AVAILABILITY

All experimental data and analysis scripts are available upon request by contacting the Lead Contact, Eran Eldar (e.eldar@ucl.ac.uk).

Current Biology, Volume *28*

Supplemental Information

Decodability of Reward Learning Signals Predicts Mood Fluctuations

Eran Eldar, Charlotte Roth, Peter Dayan, and Raymond J. Dolan

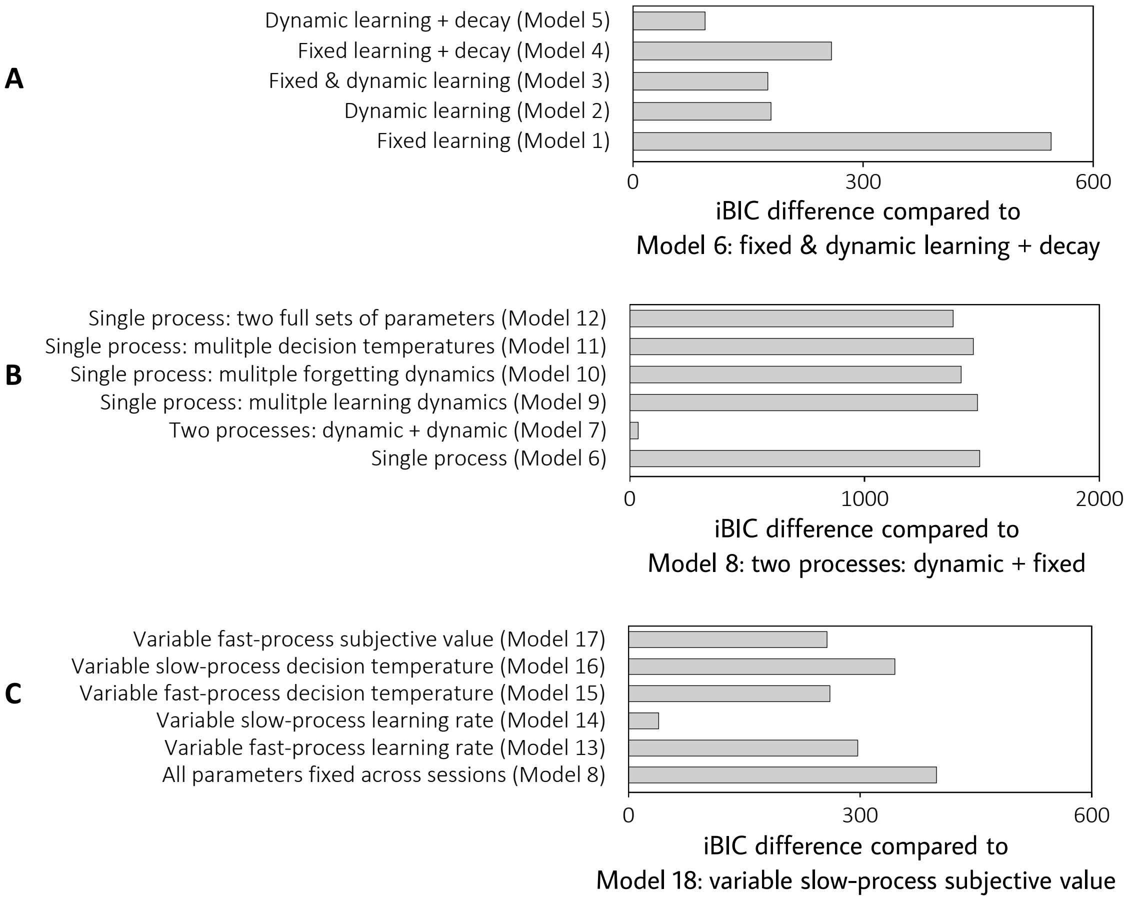

Figure S1. Modeling subjects' choices. Related to Figure 2. *n* = 10 subjects. To gain insight into subjects' learning processes, we compared multiple models in terms of how well they explained subjects' choices in the task. For each set of models, iBIC scores (integrated Bayesian Information Criterion) are shown in comparison with the best-fitting model. indicates better fit with subjects' choices (see **STAR Methods** for details of all models and model comparison procedure). (**A**) We first tested whether subjects updated their expectations similarly following each outcome ('fixed learning'), gradually reduced their learning rate ('dynamic learning'), or a combination of both ('fixed & dynamic'), as well as whether subjects' expected values decayed as a function of time ('decay'). The best-fitting model ('fixed & dynamic + decay'; iBIC = 22918) included all features. (**B**) We then tested whether subjects' decisions were better explained by assuming that learning involved two sets of expectations, each reflecting outcomes on a different timescale ('Two processes'), whether one of these sets was updated with a fixed learning rate ('dynamic + fixed'), and whether the same dynamics could be captured by models with a single set of expectations but where learning, forgetting and choice selection involve additional parameters that allow for dynamics with multiple timescales (Models 12 to 15). The best-fitting model ('dynamic + fixed'; iBIC = 21428) involved two sets of expectations. (**C**) Finally, we tested whether a subject performed the task similarly in different sessions ('parameters fixed across sessions') or whether subjects' behavior indicated that some aspect of learning or decision making varied from session to session. The best-fitting model ('variable slow-process subjective value'; iBIC = 21070) involved variability in the subjective value of reward. We also tested models with two variable parameters (not shown), but these did not achieve a better score than the best model with a single variable parameter (Model 18).

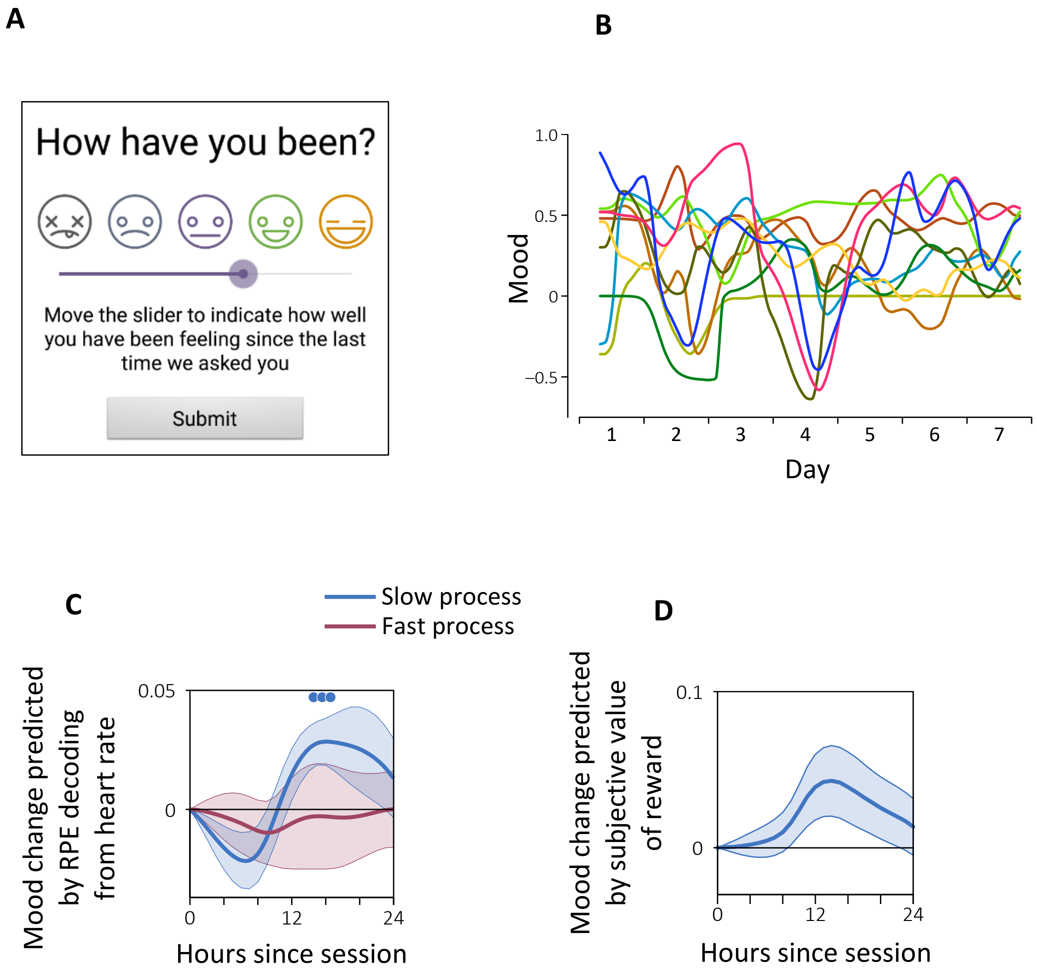

Figure S2. Changes in self-reported mood. Related to Figure 4. (**A**) Mood self-report analog scale. Subjects regularly rated their mood by adjusting the continuous slider. The rightmost end of the slider counted as +1 and the leftmost end as –1. Icons were modeled after the Daylio mobile app [S1]. (**B**) Subjects' self-reported mood over the course of the experiment. 1.0 and –1.0 correspond to best and worst possible mood, respectively. Each line shows one individual subject's mood, integrating all of the subject's self-reports using Gaussian filters (see **STAR Methods**). (**C**) Average change in mood following each experimental session as a function of reward PE decoding from heart rate. (**D**) Average change in mood following each experimental session as a function of reward sensitivity. Sensitivity was inferred from subjects' choices using the computational model via the 'subjective value of reward' parameter ψ. In both panels, magnitude of change is shown per one standard deviation of decoding accuracy / reward sensitivity. ●: difference from zero ($p_{\text{corrected}}$ = 0.04). Shaded areas: SEM.

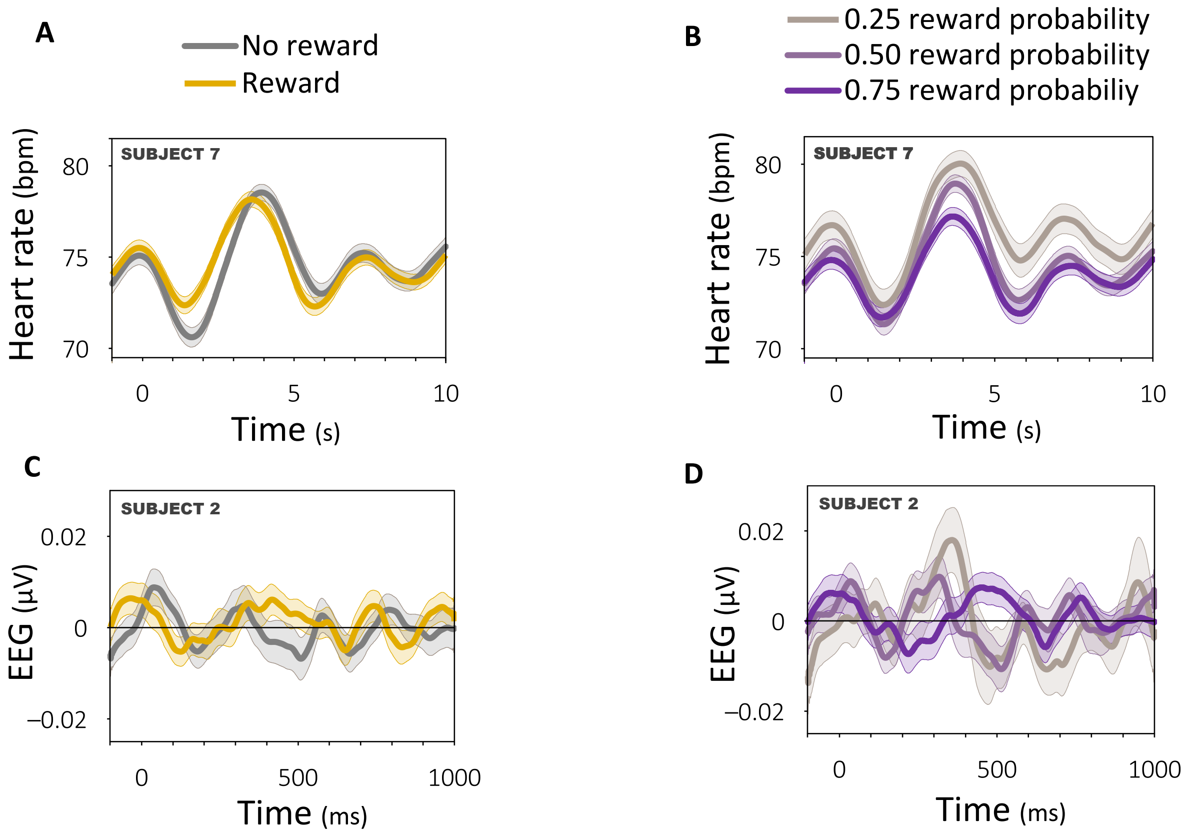

Figure S3. Heart rate and EEG responses to outcomes in the experimental task in two exemplar subjects. Related to Figure 3. Time 0 indicates outcome onset. Shaded areas: SEM (**A**, **C**) Average heart rate (**A**) and EEG (**C**) recorded following reward and no-reward outcomes. (**B**, **D**) Heart rate (**B**) and EEG (**D**) responses to outcomes as a function of the reward probability associated with the chosen image.

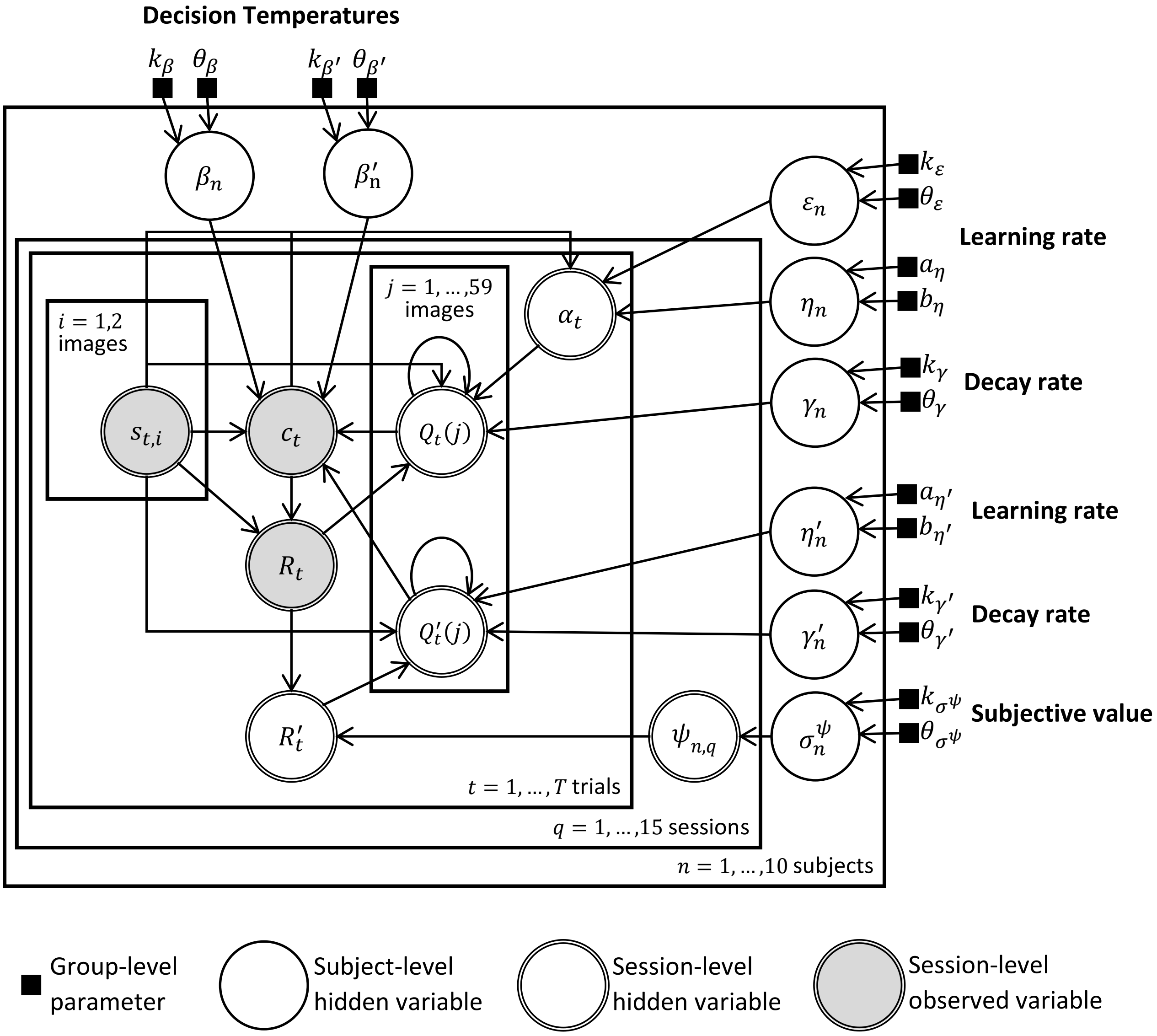

Figure S4. Graphical model of Model 18 (Eqs. 1,2,4,6–8). Related to STAR Methods. Subject n's choice (c) between the images (s) available on trial t of session q, reflect a combination of two sets of expected values (Q and Q'). These two sets of expectations are learned from the same series of choices and outcomes (R) but with different learning and decay dynamics. For the first set (Q), the learning rate (α) decreases as a function of the number of times the chosen image has previously been chosen. For the second set (Q'), the learning rate is fixed (η') and the subjective value of reward (R') varies across sessions. Self-directed arrows indicate that the value of a variable in trial t depends on its value in trial $t - 1$. Session and subject indices are omitted within trials for simplicity.

Model	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Fixed learning rate	+			+				+	+			+	+	+	+	+	+	+
Dynamic learning rate		+			+													
Fixed + dynamic learning			+			+	+	+	+	+	+	+	+	+	+	+	+	+
Expectation decay				+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Two sets of expectations							+	+					+	+	+	+	+	+
Multiple learning rates							+	+	+			+	+	+	+	+	+	+
Multiple decay rates							+	+		+		+	+	+	+	+	+	+
Multiple decision temperatures							+	+			+	+	+	+	+	+	+	+
Variable fast learning rate													+					
Variable slow learning rate														+				
Variable fast dec. temperature															+			
Variable slow dec. temperature																+		
Variable fast subjective value																	+	
Variable slow subjective value																		+

Table S1. Summary of model features. Related to STAR Methods.

Simulated Model

Detected Model

	1	2	3	4	5	6
1	5	0	0	0	0	0
2	0	5	0	0	0	0
3	0	0	5	0	0	0
4	0	0	2	3	0	0
5	0	0	0	0	5	0
6	0	0	0	0	0	5,2

Detected Model

	6	7	8	9	10	11	12
6	5	0	0	0	0	0	0
7	0	0	5	0	0	0	0
8	0	0	5,4	0	0	0	0
9	4	0	0	1	0	0	0
10	1	0	0	0	4	0	0
11	0	0	0	0	0	5,2	0
12	0	0	0	0	2,1	0	3,1

Detected Model

	8	13	14	15	16	17	18
8	5	0	0	0	0	0	0
13	1	4	0	0	0	0	0
14	0	0	4,2	0	0	0	1
15	0	0	0	1	0	4	0
16	1	0	0	0	4	0	0
17	0	0	0	1	0	4	0
18	0	0	0	0	0	0	5,2

Table S2. Validation of the model comparison procedure. Related to STAR Methods. We simulated 5 data sets using each model with its parameters fitted to subjects' real choices, and we applied the model comparison procedure to each data set. Each cell shows how many datasets generated by the model indicated on the vertical axis were detected as reflecting the model indicated on the horizontal axis. In red, we indicate cases where a model was detected with a BIC difference (compared to second best model) that was equal or higher than the BIC difference found for subjects' actual data.

Supplemental References

S1. Chaudhry, B.M. (2016). Daylio: mood-quantification for a less stressful you. *mHealth* 2, 34.