Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward

Kenneth T. Kishida[a,1], Ignacio Saez[a,2], Terry Lohrenz[a], Mark R. Witcher[b], Adrian W. Laxton[b], Stephen B. Tatter[b], Jason P. White[a], Thomas L. Ellis[b,3], Paul E. M. Phillips[c,d], and P. Read Montague[a,e,f,1]

[a]Virginia Tech Carilion Research Institute, Virginia Tech, Roanoke, VA 24016; [b]Department of Neurosurgery, Wake Forest Health Sciences, Winston-Salem, NC 27157; [c]Department of Psychiatry & Behavioral Sciences, University of Washington, Seattle, WA 98195; [d]Department of Pharmacology, University of Washington, Seattle, WA 98195; [e]Department of Physics, Virginia Tech, Blacksburg, VA 24060; and [f]Wellcome Trust Centre for Neuroimaging, University College London, London WC1N 3BG, United Kingdom

Edited by Marcus E. Raichle, Washington University in St. Louis, St. Louis, MO, and approved October 23, 2015 (received for review July 13, 2015)

In the mammalian brain, dopamine is a critical neuromodulator whose actions underlie learning, decision-making, and behavioral control. Degeneration of dopamine neurons causes Parkinson's disease, whereas dysregulation of dopamine signaling is believed to contribute to psychiatric conditions such as schizophrenia, addiction, and depression. Experiments in animal models suggest the hypothesis that dopamine release in human striatum encodes reward prediction errors (RPEs) (the difference between actual and expected outcomes) during ongoing decision-making. Blood oxygen level-dependent (BOLD) imaging experiments in humans support the idea that RPEs are tracked in the striatum; however, BOLD measurements cannot be used to infer the action of any one specific neurotransmitter. We monitored dopamine levels with subsecond temporal resolution in humans (n = 17) with Parkinson's disease while they executed a sequential decision-making task. Participants placed bets and experienced monetary gains or losses. Dopamine fluctuations in the striatum fail to encode RPEs, as anticipated by a large body of work in model organisms. Instead, subsecond dopamine fluctuations encode an integration of RPEs with counterfactual prediction errors, the latter defined by how much better or worse the experienced outcome could have been. How dopamine fluctuations combine the actual and counterfactual is unknown. One possibility is that this process is the normal behavior of reward processing dopamine neurons, which previously had not been tested by experiments in animal models. Alternatively, this superposition of error terms may result from an additional yet-to-be-identified subclass of dopamine neurons.

dopamine | reward prediction error | counterfactual prediction error | decision-making | human fast-scan cyclic voltammetry

Dopamine is an essential neuromodulator whose presence is required for normal learning, decision-making, and behavioral control (1, 2) and whose absence or dysfunction is associated with a variety of disease states including Parkinson's disease, schizophrenia, addiction, and depression (3–7). Experiments in animal models support the hypothesis that changes in dopamine release at target neural structures encode reward prediction errors (RPEs) (the difference between actual and expected outcomes) important for learning and value-based decision-making (1, 8–12). In support of this claim, direct recordings of spike activity in mesencephalic dopaminergic neurons in nonhuman primates demonstrate that these neurons encode prediction errors in future reward delivery (8–10, 13, 14) and they may also encode other computations relevant for reward-guided actions (1, 15–17). However, action potential production in brainstem dopaminergic neurons can only be part of the story because activity in parent axons must be converted to changes in neurotransmitter release at synaptic terminals to have any impact on downstream neural systems (1, 18). There have been no direct measurements of dopamine release in human striatum that tests these ideas directly. In a large cohort of human subjects (n = 17), we tested the hypothesis that subsecond fluctuations in dopamine delivery to the human striatum encode RPEs generated during a sequential choice task.

Our measurements of dopamine release are made in patients undergoing deep brain stimulating (DBS)-electrode implantation for the treatment of Parkinson's disease. This patient population provides a unique and important window of opportunity to investigate dopamine's role in human brain function. Parkinson's disease symptoms are treated with dopamine replacement

Significance

There is an abundance of circumstantial evidence (primarily work in nonhuman animal models) suggesting that dopamine transients serve as experience-dependent learning signals. This report establishes, to our knowledge, the first direct demonstration that subsecond fluctuations in dopamine concentration in the human striatum combine two distinct prediction error signals: (*i*) an experience-dependent reward prediction error term and (*ii*) a counterfactual prediction error term. These data are surprising because there is no prior evidence that fluctuations in dopamine should superpose actual and counterfactual information in humans. The observed compositional encoding of "actual" and "possible" is consistent with how one should "feel" and may be one example of how the human brain translates computations over experience to embodied states of subjective feeling.

Author contributions: K.T.K., T.L., T.L.E., P.E.M.P., and P.R.M. designed research; P.R.M. guided all aspects of this work, including conception of the adaptation of prior rodent microsensor technology for use in humans; T.L. and P.R.M. designed the sequential choice task; M.R.W., A.W.L., S.B.T., and T.L.E. conceived of surgical strategies for safe and effective placement of microsensors for human fast-scan cyclic voltammetry (FSCV) experiments; P.E.M.P. guided microsensor fabrication; I.S. assisted with optimization of microsensor design and engineering of mobile electrochemistry unit; K.T.K., I.S., M.R.W., A.W.L., and S.B.T. performed research; M.R.W., A.W.L., and S.B.T. performed surgical placement of probes; P.E.M.P. guided FSCV experiments; K.T.K. executed FSCV experiments (in vivo and in vitro); I.S. assisted with FSCV experiments (in vivo and in vitro); K.T.K., I.S., J.P.W., and P.E.M.P. contributed new reagents/analytic tools; K.T.K. built and optimized parameters for the extended carbon-fiber microsensors and engineered the integration of mobile electrochemistry unit with game play technology; P.R.M. guided and interpreted signal extraction development and optimization procedures; K.T.K. optimized the signal extraction algorithm using the elastic net; J.P.W. performed temporal alignment of signals collected on electrochemistry unit and integrated game play system (NEMO); K.T.K., I.S., T.L., J.P.W., and P.R.M. analyzed data; P.R.M. guided all analyses; P.R.M. guided and interpreted results from FSCV experiments; K.T.K., I.S., T.L., M.R.W., A.W.L., S.B.T., J.P.W., P.E.M.P., and P.R.M. interpreted results; and K.T.K., T.L., and P.R.M. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

See Commentary on page 22.

[1]To whom correspondence may be addressed. Email: read@vt.edu or kenk@vtc.vt.edu.

[2]Present address: Helen Wills Neuroscience Institute and Haas School of Business, University of California, Berkeley, CA 94720.

[3]Deceased June 30, 2012.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1513619112/-/DCSupplemental.

therapies, and yet we know nothing about how rapid (subsecond) dopamine concentration changes contribute to their symptoms or changes in their decision-making abilities. The opportunity to measure dopamine release with subsecond temporal resolution in the brains of humans with Parkinson's disease is an opportunity to learn about fundamental processes in human brain function as well as an opportunity to assess dopamine signaling in a patient population whose primary treatment is focused on replacing function lost as dopamine neurons degenerate.

Participants ($n = 17$) in these experiments performed a simple, yet engaging, sequential investment game (Fig. 1 and refs. 19–21) while dopamine measurements with subsecond temporal resolution were made in the striatum ($n = 14$ in the caudate and $n = 3$ in the putamen). Participants were offered participation after they were deemed candidates for deep brain-stimulating electrode implantation (22, 23). The research protocol was explained to the participants verbally, and they were provided a written consent form, as required by dual-institutional review board (IRB)-approved protocols at Wake Forest University Health Sciences and Virginia Tech Carilion Research Institute. Patients thus indicated that they understood the research protocol and provided written informed consent to proceed with the research procedure.

The sequential investment game (Fig. 1 and refs. 19–21) consists of 120 investment decisions. On each trial (t), this game requires participants to use button boxes to adjust and submit an investment [bet (b_t), where bet sizes could range from 0% to 100% of the participants portfolio, in 10% increments], after which, participants experience a gain or loss (participant return) equal to the bet size times the fractional change in the market price [market return (r) at time t: $r_t = \frac{\Delta p_t}{p_t}$, where p is the market price and the participant return (i.e., gain or loss) at time t is equal to $b_t r_t$]. Previous work used this task and functional magnetic resonance imaging to demonstrate that RPEs and CPEs over gains are tracked by blood oxygenation level-dependent (BOLD) responses in the striatum (19, 20). These reports also demonstrated at the behavioral level that humans use counterfactual information over choices that "might have been made" and RPE information over choices that were actually made to make their next choice (19, 20).

Results

Cross-Validated Penalized Linear Regression Approach Reliably Estimates Low Dopamine Concentrations. During the execution of the sequential investment game, an adaptation of fast-scan cyclic voltammetry (FSCV) was used to track subsecond dopamine fluctuations in the striatum. Standard approaches [see *SI Methods*, *Principal Components Regression to Estimate Dopamine Concentration* and Figs. S1–S3 and Table S1, which follow recommendations in ref. 24] for estimating dopamine concentration from FSCV measurements produced unreliable predictions for low dopamine concentrations in vitro (Fig. 2*A*). Furthermore, and also under controlled in vitro conditions, we observed that these methods produced predictions of dopamine concentration fluctuations that confused changes in pH for changes in dopamine (Fig. 2*B*). Thus, we sought to develop a novel approach that uses in vitro calibration data to fit a cross-validated penalized linear regression model for estimating dopamine concentrations from non–background-subtracted voltammograms [see *SI Methods* for details on our elastic net (EN)-based approach]. The new approach was sufficiently sensitive and stable to permit dopamine measurements at low levels expected in patients diagnosed with Parkinson's disease (Fig. 2 *C–E* and Fig. S4). Fig. 2*C* shows our approach stably and accurately estimating dopamine levels in out-of-sample test cases from the same electrode and flow cell conditions used in Fig. 2 *A* and *B*. Fig. 2*D* shows that the cross-validated EN-based approach used to accurately track changes in dopamine concentration in Fig. 2*C* does not confuse changes in pH for dopamine fluctuations. Fig. S5 shows our approach achieving signal-to-noise ratios (SNRs) ranging from 2/1 to 5,000/1 for tonic dopamine concentrations ranging from 500 nM to 10 μM, respectively. For the results below, we use our EN-based approach to estimate dopamine levels from non–background-subtracted voltammograms measured in the striatum of humans undergoing DBS-electrode implantation surgery.

Dopamine Transients Fail to Simply Encode RPEs. Dopamine measurements were made in 17 participants; each participant made 20 investment decisions per market in a total of six markets (120 decisions total per subject; one subject did not complete one market). At each decision within a market, an RPE was computed as the difference between the outcome (gain or loss as defined above) and the expected value of the outcome for that market (i.e., the average participant return up to that trial in that market); this difference is normalized to the variability of the preceding outcomes to facilitate comparison across markets and across participants (see Eq. **2** in *Materials and Methods* for equation and description of terms). The distribution of RPEs (Fig. 3*A*) is peaked around 0 but evenly distributed for positive and negative values. We divide these events into positive and negative RPEs and report the mean dopamine responses to positive (green; $n = 17, n = 1{,}022$) and negative (red; $n = 17$, $n = 991$) RPEs in Fig. 3*B*. The measured dopamine fluctuations in human striatum fail to distinguish RPEs categorized by sign [Fig. 3*B*; two-way ANOVA: $F(1,7) = 1.67$, $P = 0.1965$]. This null result holds even at lower sample sizes ($n \cong 200$ per category, randomly sampled). Prior work strongly supports the hypothesis that dopamine fluctuations in striatum should track RPEs (1, 8, 10, 11, 13, 14). Our results contradict this expectation; however, the task we use was designed to also assess the impact of counterfactual feedback (e.g., difference between actual outcomes and what might have happened; Eq. **3**). In this game,

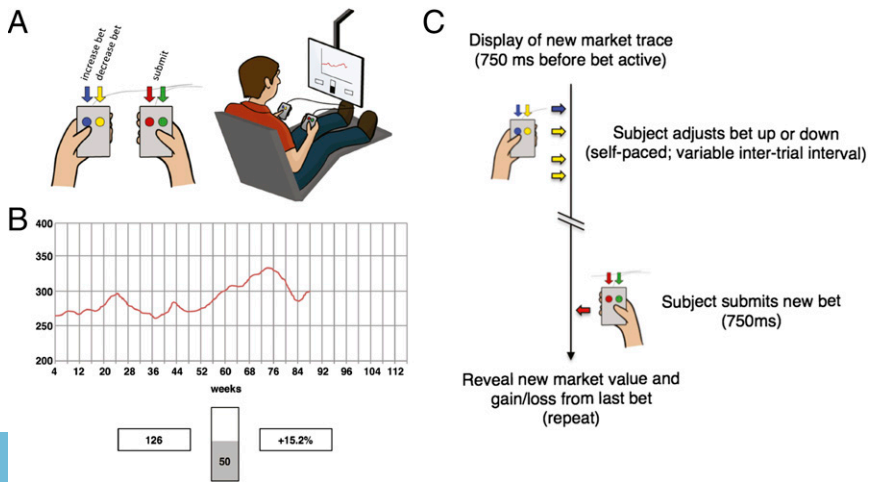

Fig. 1. Investment game. (*A*) Participants played a sequential-choice game during surgery using button boxes (*Left*) and a visual display (*Right*). For each patient, bet size adjustments (e.g., increase bet or decrease bet) and the decision to submit one's answer were performed with button boxes. (*B*) Investment game (19, 21): participants view a graphical depiction of the market price history (red trace), their current portfolio value (bottom left box), and their most recent outcome (bottom right box) to decide and submit investment decisions (bets) using a slider bar in 10% increments (bottom center). Bet sizes were limited to 0–100% (in 10% increments) of the participant's portfolio—no shorting of the market was allowed. During an experiment, a participant played 6 markets with 20 decisions made per market. (*C*) Timeline of events during a single round of the investment game.

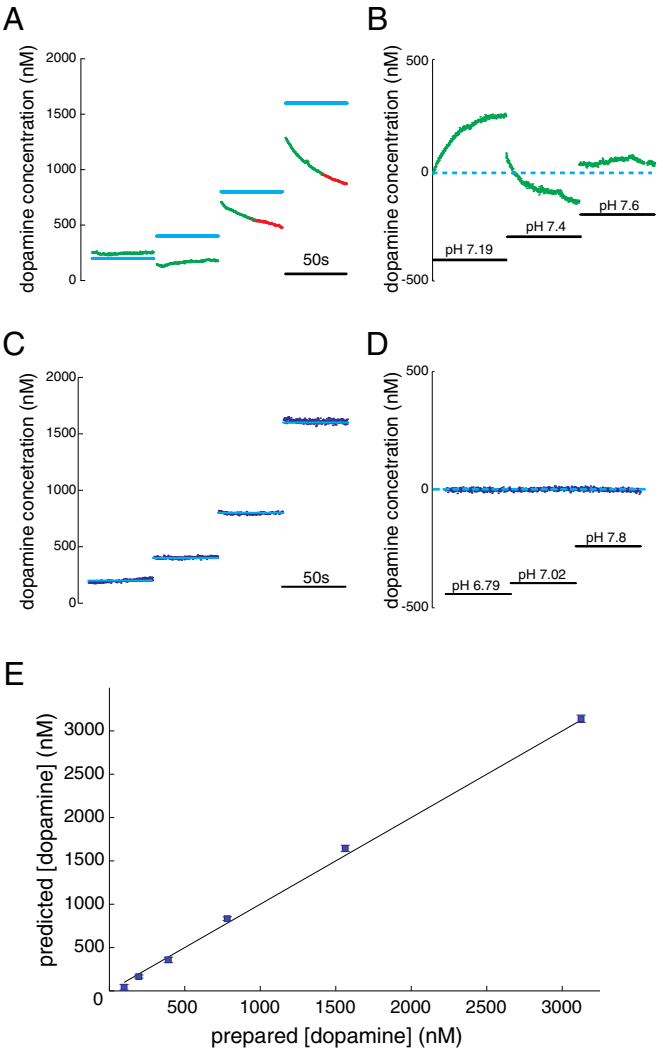

Fig. 2. Performance of EN-based dopamine estimation algorithm. (*A* and *B*) Performance of PC regression-based approach on out-of-sample test cases. (*C*–*E*) Performance of EN-based approach on out-of-sample test cases. For *A* and *B*, light blue lines indicate dopamine concentration of prepared out-of-sample calibration solutions, green points indicate PC regression-based predictions accepted by Q-value analysis, and red points indicate PC regression-based predictions rejected by Q-value analysis. The inset scale bar indicates measurement period of 50 s. (*A*) PC regression-based prediction of changes in dopamine concentration under stable pH (pH 7.4). The prepared dopamine concentration (light blue; 200, 400, 800, and 1,600 nM) compared with the PC regression-based predictions for dopamine concentration (red and green). (*B*) PC regression-based predictions of dopamine concentration when pH is changed, but dopamine concentration is held constant (0 dopamine in solution). The dotted light blue line indicates actual concentration of dopamine is equal to 0. *Insets* indicate pH levels (pH range: 7.19, 7.4, 7.6). (*C*) EN-based predictions of changes in dopamine concentration under stable pH (pH = 7.4). The prepared dopamine concentration (light blue, 200, 400, 800, and 1,600 nM) compared with the EN-based predictions for dopamine concentration (dark blue). The inset scale bar indicates measurement period of 50 s. (*D*) EN-based predictions of changes in dopamine concentration when pH is changed, but dopamine concentration is held constant (0 dopamine in solution). The dotted light blue line indicates that actual concentration of dopamine is equal to 0. *Insets* indicate pH levels (pH range: 6.79, 7.02, 7.8). (*E*) Dopamine concentration predictions from the EN-based procedure gives accurate predictions of dopamine concentration (blue squares). Horizontal axis: concentration of prepared dopamine; vertical axis: predicted dopamine concentration. Plotted are mean predicted values for five measurements at each concentration ± SEM (note: SEM bars are plotted but are consumed by the marker).

counterfactual prediction errors (CPEs) (Eq. **3**) are parameterized by the distribution of participants' bets. Instances where there is no CPE occur when the participants' bet is equal to one (i.e., "all in"). In these specific instances, we observe (Fig. 3*B*, *Inset*) that dopamine transients to positive ($n = 173$) and negative ($n = 164$) RPEs indeed separate. Together, these results suggest that counterfactual information (as bet sizes decrease from 1) disrupts the expected standard response of dopamine release to positive and negative RPEs. We test this hypothesis below (*Results*, *Dopamine Transients Integrate RPEs and CPEs*) by examining the dopamine response to equivalent magnitude RPEs for different bet sizes.

Dopamine Transients Integrate RPEs and CPEs. Behaviorally, CPEs in this task have been shown to combine with RPEs to influence participants' next decision (19, 20). Given the impact of bet size (and thus potentially counterfactual information) on the encoding of RPEs by dopamine fluctuations (Fig. 3*B*), we tested a novel hypothesis: subsecond dopamine transients encode a combination of RPEs and CPEs. To test our hypothesis, we follow the model of CPEs presented by Lohrenz et al. (19) and assume that dopamine encodes a linear combination of two separate computations: RPEs (Eq. **2**) and CPEs (Eq. **3**): dopamine transient $\propto \{\mathrm{RPE}\} - \{\mathrm{CPE}\}$, or

$$\text{dopamine transient} \propto \{\mathrm{RPE}\} - \{r_t(1 - b_t)\}. \qquad [1]$$

Here, b_t is the subject's fractional bet at choice trial t and r_t expresses the relative fractional change in the market price ($r_t = \frac{\Delta p_t}{p_t}$). The difference in what the participant earned and what the participant could have earned is the CPE (second term on the right in Eq. **1**). The guiding intuition for the form of Eq. **1** is twofold: (*i*) that what might have been should adjust overall valuation estimates and encode this adjusted amount in a composite dopamine signal and (*ii*) that the RPE and CPE terms are computed in two separate pathways before being integrated at the level of dopamine release. Thus, the valuation error encoded by dopamine release is consistent with the intuition that "better-than-expected" outcomes (positive RPEs) that might have been better should be reduced in value and "worse-than-expected" outcomes (negative RPEs) that might have been worse should be increased in value. In this form, positive CPE terms (which

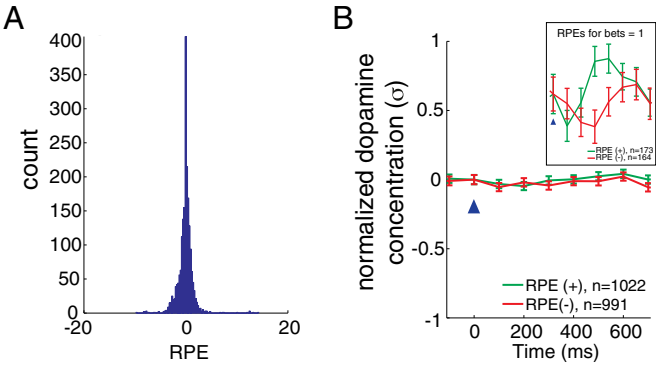

Fig. 3. Dopamine fails to simply track RPEs in the investment game. (*A*) Histogram showing the distribution of events for RPEs ($n = 17$ participants; $n = 2{,}013$ outcome revelations). (*B*) Mean normalized dopamine responses (±SEM) to positive (green; $n = 1{,}022$) and negative (red; $n = 991$) RPEs. Two-way ANOVA (RPE-sign and time: 700 ms following and including outcome reveal) reveals no significant difference comparing dopamine responses for positive and negative RPEs [$F_{\text{RPE-sign}}(1,7) = 1.67$, $P = 0.1965$]. Note, this null result holds even at lower sample sizes (n ≅ 200 per category, randomly sampled) comparable to those in Fig. 4. Horizontal axis: time (ms) from outcome reveal (blue arrow head); vertical axis: mean change in the dopamine response. Before averaging, dopamine traces are normalized to the SD (σ) of the fluctuations measured within patient. Error bars: SEM. *Inset* shows dopamine response to a subset of positive (green) and negative (red) RPE events (i.e., when the participants' bet all in).

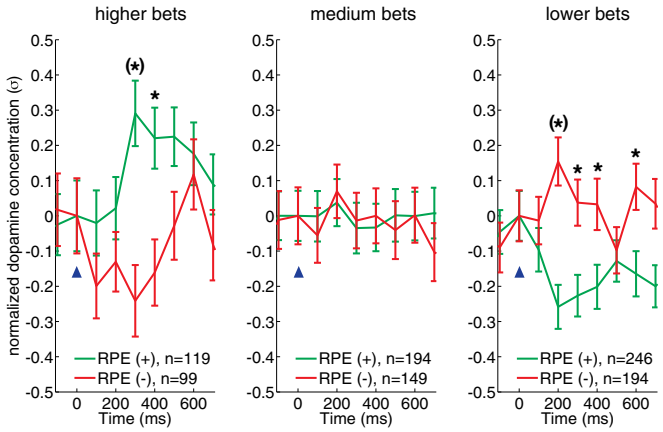

Fig. 4. RPE encoding by dopamine transients invert as a function of bet size. Dopamine responses to equal absolute magnitude positive and negative RPEs (–0.75 > RPE > +0.75) when bets are high (higher bets, 100–90%) (*Left*), medium (medium bets, 80–60%) (*Center*), or low (lower bets, 50–10%) (*Right*). For all three plots, mean normalized dopamine responses (±SEM) to positive RPEs (green traces) and negative RPEs (red traces). Inset legends show sample sizes for event types. Two-way ANOVA (RPE-sign and time: 700 ms following and including outcome reveal) reveals a significant difference comparing dopamine responses for positive and negative RPEs following higher bets [$F_{RPE\text{-}sign}(1,7) = 21.17$, $P = 0.00$] and lower bets [$F_{RPE\text{-}sign}(1,7) = 32.64$, $P = 0.00$] but not medium bet sizes [$F_{RPE\text{-}sign}(1,7) = 0.15$, $P = 0.6957$]. Asterisks indicate significant difference between red and green traces: $P < 0.05$, post hoc, two-sample t test following ANOVA with time and RPE-sign as the two main factors. Asterisks with parentheses indicate Bonferroni correction for multiple comparisons. For low bets (i.e., large CPEs), only those events where the market price change and the RPE-sign are the same are considered. Horizontal axis: time (ms) from outcome reveal (blue arrowhead); vertical axis: mean change in normalized dopamine response.

occur for missed opportunities on positive-going markets) diminish the value of the RPE event, and negative CPE terms (which occur for avoided losses on negative-going markets) increase the value of the RPE event.

This model (Eq. **1**) makes three testable predictions all of which derive from a dependence on the bet size b_t. We test these predictions for events that have the same magnitude of RPEs (positive and negative) but grouped for different size bets. According to the model in Eq. **1**, we predict (and observe) the following.

Prediction 1: Impact of bet size equal to 1 (i.e., no CPE). When the bet (b_t) is set near 1 (all in), the CPE is 0, so positive RPEs will be encoded as positive-going dopamine transients and negative RPEs will be encoded as negative-going dopamine transients [similar to experiments in rodents and nonhuman primates (8, 11, 25) and exactly what is observed in Fig. 3*A*, *Inset*, and for the "higher bets" graph in Fig. 4].

Prediction 2: Impact of decreasing bet size on dopamine transient polarity for positive RPEs. As the bet size decreases, the CPE grows in magnitude; thus, dopamine transients to positive RPEs will diminish (as is observed with the green traces in the "medium bets" graph in Fig. 4) and eventually will be encoded as a negative-going transients as the CPE term dominates (as is observed for green traces in the "lower bets" graph in Fig. 4).

Prediction 3: Impact of decreasing bet size on dopamine transient polarity for negative RPEs. Again, as the bet size decreases the CPE grows in magnitude; thus, dopamine transients to negative RPEs will diminish (as is observed for red traces in the medium bets graph in Fig. 4) and eventually be encoded as a positive-going transients as the CPE term dominates (as is observed for red traces in the lower bets graph in Fig. 4).

Fig. 4 demonstrates that striatal dopamine measurements in humans follow predicted responses of the simple model expressed in Eq. **1**. The three separate predictions are pertinent because in nonhuman primates dopamine neurons show asymmetrical modulation of their activity as a function of the RPE polarity (8, 10, 13, 15–17), and the integration of a CPE term has not previously been shown in experiments measuring changes in dopamine neuron firing rate. One possible interpretation here is that there is a separate class of midbrain dopamine neurons carrying the counterfactual information and these have yet to be recorded from in prior experiments. This hypothesis is particularly important because these subjects are patients with Parkinson's disease, suggesting a class of dopamine neuron possibly preserved in the disease. Alternatively, these results suggest a previously untested mode of operation of reward processing dopamine neurons. Along these lines, different error terms for evaluating behavioral outcomes may be integrated before dopamine release. Both of these possibilities suggest two separate pathways for computing actual and counterfactual outcomes over past decisions. Where these computations may take place is unknown. Further work is needed to distinguish these and other possibilities.

Discussion

We tested the hypothesis that fast fluctuations in dopamine concentration encode RPEs over monetary gains and losses using FSCV (to measure dopamine release) in 17 participants while they played a sequential investment game. The data show that a simple encoding of RPEs by dopamine release is not the case (Fig. 3*B*). Instead, our data are consistent with the idea that dopamine fluctuations integrate a RPE term with a CPE term (Fig. 3*B*, *Inset*, and Fig. 4). A model (Eq. **1**) that subtracts a CPE term from the RPE term is consistent with our data (Figs. 3 and 4) and is consistent with how counterfactual experience should modulate actual experience but in computational terms. This model makes the surprising prediction that counterfactual outcomes can suppress and even invert dopamine responses to positive and negative RPEs.

Notably, our model and the dopamine responses it explains also capture qualitative aspects about how one should "feel" (e.g., good or bad) given one's action, the resulting outcome, and the overall context of that outcome. For example, a better-than-expected outcome should feel good (i.e., rewarding); however, if the exact same outcome occurs when an alternative action could have resulted in an even better outcome, then the positive feelings associated with the better-than-expected experience should be diminished and in extreme cases such an experience should feel bad (i.e., aversive). This is consistent with feelings of "regret" and the negative feelings associated with missed opportunities. Likewise, a worse-than-expected outcome should feel bad (i.e., aversive), but, if that outcome is experienced when the outcome could have been much worse, then the overall experience should be driven toward the positive. These analogous feelings of "relief" are typically positive and rewarding for actions that avoid counterfactually large losses or severe punishment. Our model, and the impact of combining actual and counterfactual information to evaluate decision-making, has connections to regret-based theories of decision-making under uncertainty (26, 27). An interesting point here is that this combination of information in a single physical signal (the dopamine response) could be one way that the human brain translates computations about actual and simulated experience to embodied states of feeling.

These data are collected in humans undergoing DBS-electrode implantation for the treatment of Parkinson's disease. In many respects, decision-making in patients with Parkinson's disease remains largely intact: they make their own financial decisions, they are free to choose to consent in clinical and research procedures, and they make many other life critical decisions. However, prior work suggests that pharmacological agents, DBS therapies, and the Parkinson's disease state have been associated with changes in patient behaviors associated with impulse control, adaptive decision-making, and goal-directed behaviors (28–30). The impact of significant dopamine neuron loss, which characterizes this disease, is important to consider. For example, it is unclear what aspects of the dynamics in the dopamine response we report are attributable to the normal state of dopamine neuron function in humans, to reduced dopaminergic signaling caused by Parkinson's disease pathology, or perhaps downstream adaptive mechanisms resulting from patients' history of pharmacotherapy. Thus, it is unclear whether the integration of

these two error terms is representative of typical dopamine release in a non-Parkinsonian brain. For example (and speculatively), exogenously increased levels of dopamine via l-3,4-dihydroxyphenylalanine (L-DOPA) therapy could cause serotonergic terminals to inappropriately load dopamine through cellular reuptake mechanisms or directly convert L-DOPA into dopamine in terminals that normal release serotonin (31, 32); thus, signals normally encoded by serotonin release might then be misencoded by dopamine release. Although an integration of these terms is consistent with how a decision-making agent might account for these opponent feedback signals, it is not a priori necessary that dopamine release encodes this specific computation. Further experiments are required to determine whether dopamine release encodes the integration of actual and CPE terms in humans without Parkinson's disease or model systems where the dopaminergic system is intact.

These present results are unanticipated by current models and data collected in nonhuman model organisms including nonhuman primate dopamine neuron recordings and dopamine release measurements in rodents. Work in nonhuman primates has demonstrated that neural activity (somatic spikes) in the anterior cingulate cortex (33), orbital frontal cortex and dorsolateral prefrontal cortex (34), and in rodent striatum (35) are able to track counterfactual information. These studies indicate that activity of single neurons in rodents and nonhuman primates are able to track counterfactual information reflected through changes in spike frequencies but do not demonstrate a mechanism by which these signals are integrated to represent modulations in value estimates of outcomes. Our results demonstrate that experience-dependent RPEs and simulated CPEs are combined at the level of extracellular dopamine fluctuations in the striatum within hundreds of milliseconds following the revelation of a decision–outcome.

In humans, it has been shown that lesions to the orbital frontal cortex impair counterfactual information processing as read out through decision-making behavior and subjective reports about feelings of regret and relief (36). Also, BOLD imaging experiments in humans support the idea that counterfactual information is represented by brain responses in the orbital frontal cortex (37) and striatum (19, 20, 38). However, BOLD imaging is unable to provide specific information about the neurotransmitters involved (39), nor do BOLD imaging experiments provide specific information about how the brain encodes this information at the level of neurotransmitter release, modulations in local field potentials, or somatic spike activity (39, 40). One report has demonstrated neural activity in human substantia nigra that was consistent with dopamine neuron activity (41). In that report, dopamine neuron spike rates were demonstrated to track RPEs as in the animal model literature; however, no association between dopamine neuron activity and counterfactual signaling could be made, nor could a direct link be made between dopamine neuron activity in the substantia nigra and dopamine release in downstream targets.

Our results show dopamine fluctuations combine evaluative information about actual outcomes (RPE) and feedback about outcomes that would have occurred had the agent performed a different action (CPE). These computations are related to temporal difference learning and related Q-learning methods (42, 43), both of with are constrained by experience-dependent learning signals, meaning that these approaches only update state–action value estimates on those states and actions actually experienced. This means that an agent must sample all state–action pairs to gain a full representation of the state–space. A more efficient approach would be to update representations independent of immediate state–action experiences as alternative forms of information become available. Thus, the ability to incorporate counterfactual information should speed up the process of learning because the agent could then update value estimates on multiple states in parallel. This kind of learning from fictive experiences could occur with counterfactual information coming from a variety of sources including other agents (social learning) or more complete information becoming available after certain actions have been made. For example, in the current task, the CPE signal is the difference between the best (or worst) possible outcome and the actual outcome (Eq. **3**). This kind of counterfactual information has been shown to be an important signal for driving human choice behavior (19) and is similar to the supervised actor critic framework proposed in ref. 44 and discussed in ref. 45. Together with experience based learning, counterfactual learning signals like this one serve to speed up learning about the optimal strategy in complex and information rich environments.

How RPEs and CPEs are physically combined and contribute to the composite dopamine signal is not known. One possibility is that there are separate sets of dopaminergic neurons with activity modulations that specialize in either the prediction errors or CPEs. Such heterogeneity in dopamine neuron response profiles has been demonstrated (16, 46). This possibility has simply not been tested. Other possibilities are that such signal-dependent coding is multiplexed in a common set of mesostriatal dopamine neurons or that direct modulation of dopamine release and clearance in the terminal regions of the striatum provide direct control over the dynamics of error tracking by dopamine transients. Further work is required to separate these and other possibilities.

Methods

For more detail on all procedures, materials, and analyses presented below, refer to the *SI Methods*.

Informed Consent and Participant Recruitment. Participants (n = 17) gave informed written consent and verbal assent to the dual IRB-approved research protocol. IRB committees at Wake Forest University Health Sciences (IRB00017138) and Virginia Tech (IRB 11-078) approved all procedures involving human experimentation. Once written and informed consent was obtained from the patient, the details of the computer task (i.e., the sequential-choice game) were described, and participants practiced a version of the game to gain familiarity with the game controls and game play.

Investment Game. The investment game (Fig. 1 and refs. 19–21) requires participants to make decisions about how much of their portfolio they will invest in a "stock market" given three pieces of information: (*i*) the history of the market price, (*ii*) the participant's current portfolio value, and (*iii*) the most recent fractional change in the participant's portfolio value. The participant begins the game endowed with 100 points and plays six markets with 20 decisions in each market. The participants' final portfolio value (after all 120 decisions have been made) determines the participants' compensation. Fig. S6 shows the distribution of market returns (Fig. S6*A*), bets (Fig. S6*B*), participant returns (Fig. S6*C*), RPEs (Fig. S6*D*), and CPEs (Fig. S6*E*).

RPE Calculation. The term $b_t r_t$ corresponds to participant outcomes (gains or losses) depending on the sign of r_t. Positive r_t results in a gain if participants' bet size was greater than 0; likewise, negative r_t results in a loss for bets greater than 0. RPEs are calculated as the difference between the actual participant return on that trial and the expected return on that trial. This term is then normalized by the variability in returns experienced up to that trial and within each market:

$$\frac{b_t r_t - E(b_t r_t)}{\sigma_{b_t r_t}}, \qquad [2]$$

where $E(b_t r_t)$ is the expected value of $b_t r_t$, which is calculated as the mean of participant outcomes from the first trial of the game to trial $t-1$, and $\sigma_{b_t r_t}$ is the SD over those same events.

CPE Calculation. Participant outcomes ($b_t r_t$) are a fraction of the maximum possible outcome on any given trial. The "maximum possible" is revealed as the market return (i.e., price change) is revealed—had the participant bet all in or b_t, then the gain or loss (dependent on the sign of r_t) would have been the largest that it could have been on that trial. The difference between this value and the value of the participant's actual return ($b_t r_t$) is the CPE – the difference between what could have been and what actually happened:

$$r_t(1 - b_t). \qquad [3]$$

FSCV Carbon-Fiber Microsensors. We performed FSCV on extended carbon-fiber microsensors in the striatum (n = 14 caudate and n = 3 putamen) of patients (n = 17 total) with Parkinson's disease. The extended carbon-fiber microsensor is

constructed to match the dimensions of the tungsten microelectrodes used for functional mapping during DBS-electrode implantation surgery following ref. 21. Fig. S7 shows the component parts and assembly of the extended carbon-fiber microsensor used in these experiments.

FSCV Protocol. Our FSCV protocol follows previous work in rodents (21, 47, 48). An electrochemical conditioning protocol (see Fig. S8*A* for depiction of applied waveform) is first applied consisting of a 60-Hz application of the measurement waveform for approximately 10 min to allow equilibration. Following this conditioning procedure, a 10-Hz application of the same triangular waveform is applied for the duration of the experiment (Fig. S8*B*). Examples of the resulting voltammograms (for each patient) and their derivatives, which were used for analysis, are shown in Figs. S9 and S10, respectively.

Estimation of Dopamine Concentration. We estimate dopamine concentration, as measured by FSCV using linear regression models trained using in vitro data and the EN algorithm (refer to *SI Methods* for more information). The EN algorithm is an automatic shrinkage and regularization approach to fitting-regression models (49). We use the glmnet package developed for use in Matlab (50) to train and test cross-validated models against prepared solutions of known dopamine concentrations. Solutions of dopamine are prepared in PBS (pH 7.4). Powdered dopamine hydrochloride (Sigma-Aldrich) is dissolved in a 0.1 N solution of HCl to a concentration of 100 mM. Aliquots of this solution are diluted to 10 mM in 1× PBS and further diluted (in 1× PBS) to the desired concentration for in vitro calibration of the carbon-fiber microsensors.

ACKNOWLEDGMENTS. The authors thank the patient volunteers and the research and surgical nursing staff at Wake Forest University Health Sciences Center for invaluable support and cooperation. In particular, the authors thank Wendy Jenkins, Valerie Hughes, and Patti Pepper for coordinating the patients and clinical and research staff in support of the research efforts reported here. The authors thank Nathan Apple for help in digitizing artwork displayed in Fig. 1. The authors also thank Peter Dayan, Sam McClure, Rosalyn Moran, Cathy Price, and Alec Solway for reading and commenting on earlier drafts of this manuscript. During the course of this work, prior to publication, T.L.E. died. His contributions were critical and invaluable in the early stages of this project, including planning the execution of these experiments during surgery and evaluating the safety and applicability of the reported work. T.L.E. recognized the potential of the technology to be developed and the questions to be asked and dedicated significant time and effort leading his staff and collaborators to accomplish this work. This work was funded by the Wellcome Trust (P.R.M.), the Kane Family Foundation (P.R.M.), and Virginia Tech (P.R.M.).

1. Montague PR, Hyman SE, Cohen JD (2004) Computational roles for dopamine in behavioural control. *Nature* 431(7010):760–767.
2. Wise RA (2004) Dopamine, learning and motivation. *Nat Rev Neurosci* 5(6):483–494.
3. Lotharius J, Brundin P (2002) Pathogenesis of Parkinson's disease: Dopamine, vesicles and α-synuclein. *Nat Rev Neurosci* 3(12):932–942.
4. Moore DJ, West AB, Dawson VL, Dawson TM (2005) Molecular pathophysiology of Parkinson's disease. *Annu Rev Neurosci* 28:57–87.
5. Cohen JD, Servan-Schreiber D (1992) Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychol Rev* 99(1):45–77.
6. Hyman SE, Malenka RC (2001) Addiction and the brain: The neurobiology of compulsion and its persistence. *Nat Rev Neurosci* 2(10):695–703.
7. Nestler EJ, Carlezon WA, Jr (2006) The mesolimbic dopamine reward circuit in depression. *Biol Psychiatry* 59(12):1151–1159.
8. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16(5):1936–1947.
9. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593–1599.
10. Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47(1):129–141.
11. Hart AS, Rutledge RB, Glimcher PW, Phillips PE (2014) Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *J Neurosci* 34(3):698–704.
12. Daw ND, Doya K (2006) The computational neurobiology of learning and reward. *Curr Opin Neurobiol* 16(2):199–204.
13. Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. *Neuron* 41(2):269–280.
14. Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* 10(12):1615–1624.
15. Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299(5614):1898–1902.
16. Matsumoto M, Hikosaka O (2009) Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* 459(7248):837–841.
17. Bromberg-Martin ES, Matsumoto M, Hikosaka O (2010) Dopamine in motivational control: Rewarding, aversive, and alerting. *Neuron* 68(5):815–834.
18. Montague PR, et al. (2004) Dynamic gain control of dopamine delivery in freely moving animals. *J Neurosci* 24(7):1754–1759.
19. Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci USA* 104(22):9493–9498.
20. Chiu PH, Lohrenz TM, Montague PR (2008) Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. *Nat Neurosci* 11(4):514–520.
21. Kishida KT, et al. (2011) Sub-second dopamine detection in human striatum. *PLoS One* 6(8):e23291.
22. Limousin P, et al. (1995) Effect of parkinsonian signs and symptoms of bilateral subthalamic nucleus stimulation. *Lancet* 345(8942):91–95.
23. Limousin P, et al. (1998) Electrical stimulation of the subthalamic nucleus in advanced Parkinson's disease. *N Engl J Med* 339(16):1105–1111.
24. Keithley RB, Wightman RM (2011) Assessing principal component regression prediction of neurochemicals detected with fast-scan cyclic voltammetry. *ACS Chem Neurosci* 2(9):514–525.
25. Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* 1(4):304–309.
26. Bell DE (1982) Regret in decision making under uncertainty. *Oper Res* 30(5):961–981.
27. Loomes G, Sugden R (1982) Regret theory: An alternative theory of rational choice under uncertainty. *Econ J* 92(368):805–824.
28. Weintraub D (2008) Dopamine and impulse control disorders in Parkinson's disease. *Ann Neurol* 64(Suppl 2):S93–S100.
29. Witt K, et al. (2008) Neuropsychological and psychiatric changes after deep brain stimulation for Parkinson's disease: A randomised, multicentre study. *Lancet Neurol* 7(7):605–614.
30. Voon V, et al. (2009) Chronic dopaminergic stimulation in Parkinson's disease: From dyskinesias to impulse control disorders. *Lancet Neurol* 8(12):1140–1149.
31. Arai R, Karasawa N, Geffard M, Nagatsu I (1995) L-DOPA is converted to dopamine in serotonergic fibers of the striatum of the rat: A double-labeling immunofluorescence study. *Neurosci Lett* 195(3):195–198.
32. Carta M, Carlsson T, Kirik D, Björklund A (2007) Dopamine released from 5-HT terminals is the cause of L-DOPA-induced dyskinesia in parkinsonian rats. *Brain* 130(Pt 7):1819–1833.
33. Hayden BY, Pearson JM, Platt ML (2009) Fictive reward signals in the anterior cingulate cortex. *Science* 324(5929):948–950.
34. Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* 70(4):731–741.
35. Steiner AP, Redish AD (2014) Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. *Nat Neurosci* 17(7):995–1002.
36. Camille N, et al. (2004) The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304(5674):1167–1170.
37. Coricelli G, et al. (2005) Regret and its avoidance: A neuroimaging study of choice behavior. *Nat Neurosci* 8(9):1255–1262.
38. Tobia M, et al. (2014) Neural systems for choice and valuation with counterfactual learning signals. *Neuroimage* 89:57–69.
39. Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66: 735–769.
40. Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412(6843):150–157.
41. Zaghloul KA, et al. (2009) Human substantia nigra neurons encode unexpected financial rewards. *Science* 323(5920):1496–1499.
42. Watkins CJ, Dayan P (1992) Q-learning. *Mach Learn* 8(3-4):279–292.
43. Watkins CJCH (1989) *Learning from delayed rewards*, PhD dissertation (University of Cambridge, Cambridge, UK).
44. Barto AG, Rosenstein MT (2004) Chapter 14: Supervised actor-critic reinforcement learning. *Handbook of Learning and Approximate Dynamic Programming*, eds Si J, Barto AG, Powell WB, Wunsch D (Wiley-IEEE Press, Piscataway, NJ), pp 359–380.
45. Montague PR, King-Casas B, Cohen JD (2006) Imaging valuation models in human choice. *Annu Rev Neurosci* 29:417–448.
46. Schultz W (2007) Multiple dopamine functions at different time courses. *Annu Rev Neurosci* 30:259–288.
47. Phillips PE, Stuber GD, Heien ML, Wightman RM, Carelli RM (2003) Subsecond dopamine release promotes cocaine seeking. *Nature* 422(6932):614–618.
48. Clark JJ, et al. (2010) Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. *Nat Methods* 7(2):126–129.
49. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol* 67(2):301–320.
50. Qian J, Hastie T, Friedman J, Tibshirani R, Simon N (2013) Glmnet for matlab. Available at: www.stanford.edu/~hastie/glmnet_matlab.
51. D'Haese PF, et al. (2012) CranialVault and its CRAVE tools: A clinical computer assistance system for deep brain stimulation (DBS) therapy. *Med Image Anal* 16(3):744–753.
52. Robinson DL, Venton BJ, Heien ML, Wightman RM (2003) Detecting subsecond dopamine release with fast-scan cyclic voltammetry in vivo. *Clin Chem* 49(10):1763–1773.
53. Ward JH, Jr (1963) Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* 58(301):236–244.
54. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67.
55. Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J R Stat Soc Series B Stat Methodol* 58(1):267–288.
56. Malinowski ER (1989) Statistical f-tests for abstract factor analysis and target testing. *J Chem* 3(1):49–60.
57. Jackson JE, Mudholkar GS (1979) Control procedures for residuals associated with principal component analysis. *Technometrics* 21(3):341–349.