# 

# Dopamine-associated cached values are not sufficient as the basis for action selection

#### Nick G. Hollon<sup>a,b,c</sup>, Monica M. Arnold<sup>a,b</sup>, Jerylin O. Gan<sup>a,b,c</sup>, Mark E. Walton<sup>a,d</sup>, and Paul E. M. Phillips<sup>a,b,c,1</sup>

Departments of <sup>a</sup>Psychiatry and Behavioral Sciences and <sup>b</sup>Pharmacology and <sup>c</sup>Graduate Program in Neurobiology and Behavior, University of Washington, Seattle, WA 98195; and <sup>d</sup>Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, United Kingdom

Edited by Ranulfo Romo, Universidad Nacional Autonóma de México, Mexico City, Distrito Federal, Mexico, and approved November 3, 2014 (received for review October 17, 2014)

Phasic dopamine transmission is posited to act as a critical teaching signal that updates the stored (or "cached") values assigned to reward-predictive stimuli and actions. It is widely hypothesized that these cached values determine the selection among multiple courses of action, a premise that has provided a foundation for contemporary theories of decision making. In the current work we used fast-scan cyclic voltammetry to probe dopamine-associated cached values from cue-evoked dopamine release in the nucleus accumbens of rats performing cost-benefit decision-making paradigms to evaluate critically the relationship between dopamineassociated cached values and preferences. By manipulating the amount of effort required to obtain rewards of different sizes, we were able to bias rats toward preferring an option yielding a high-value reward in some sessions and toward instead preferring an option yielding a low-value reward in others. Therefore, this approach permitted the investigation of dopamine-associated cached values in a context in which reward magnitude and subjective preference were dissociated. We observed greater cueevoked mesolimbic dopamine release to options yielding the high-value reward even when rats preferred the option yielding the low-value reward. This result identifies a clear mismatch between the ordinal utility of the available options and the rank ordering of their cached values, thereby providing robust evidence that dopamine-associated cached values cannot be the sole determinant of choices in simple economic decision making.

decision making | action selection | cached values | dopamine

n contemporary theories of economic decision making, values are assigned to reward-predictive states in which animals can take action to obtain rewards, and these state-action values are stored ("cached") for the purpose of guiding future choices based upon their rank order (1–5). It is believed that these cached values are represented as synaptic weights within corticostriatal circuitry, reflected in the activity of subpopulations of striatal projection neurons (6–9), and are updated by dopaminedependent synaptic plasticity (10-12). Indeed, a wealth of evidence suggests that the phasic activity of dopamine neurons reports instances in which current reward or expectation of future reward differs from current expectations (13-24). This pattern of activity resembles the prediction-error term from temporal-difference reinforcement-learning algorithms, which is considered the critical teaching signal for updating cached values. A notable feature of models that integrate dopamine transmission into this computational framework is that the cached value of an action is explicitly read out by the phasic dopamine response to the unexpected presentation of a cue that designates the transition into a state in which that action yields reward. Therefore, cue-evoked dopamine signaling provides a neural representation of the cached values of available actions, and if these cached values serve as the basis for action selection, then cue-evoked dopamine responses should be rank ordered in a manner that is consistent with animals' behavioral preferences.

Numerous studies that recorded cue-evoked dopamine signaling have reported correlations with the expected utility (subjective value) of actions (24-36). For example, risk-preferring rats demonstrated greater cue-evoked dopamine release for a risky option than for a certain option with equivalent objective expected value (reward magnitude times probability), whereas risk-averse rats showed greater dopamine release for the certain than for the risky option (30). Likewise, the cached values reported by dopamine neurons in macaque monkeys accounted for individual monkeys' subjective flavor and risk preferences, with each attribute weighted according to its influence on behavioral preferences (31, 32). These observations, which are consistent across measures of dopamine neuronal activity and dopamine release, reinforce the prevailing notion that the dopamine-associated cached values could be the primary determinant of decision making (2-5, 17, 28-32) because the cue-evoked dopamine responses were rank ordered according to the animals' subjective preferences. However, there have been some reports that other economic attributes, such as effortful response costs (35-38) or the overt aversiveness of an outcome (39), are represented inconsistently by cue-evoked dopamine responses. For example, Gan et al. (35) showed that independent manipulations of two different dimensions (reward magnitude and effort) that had equivalent effects on behavior did not have equivalent effects on dopamine release. Paralleling these findings, a recent report reached a similar conclusion that dopamine transmission preferentially encodes an appetitive dimension but is relatively insensitive to aversiveness (39).

Because these cue-evoked dopamine signals represent cached values that are purported to determine action selection, their differential encoding of economic dimensions has potentially

#### Significance

Real-world decisions frequently involve tradeoffs between multiple economic dimensions, and integration of benefits and costs into a common currency of subjective value is fundamental to action selection. Phasic dopamine is widely regarded as a critical teaching signal for learning the values assigned to actions, and these stored ("cached") values can be read out from cue-evoked dopamine responses. Here, we observed a significant inversion between animals' behavioral preferences and the rank ordering of dopamine-reported cached values, indicating that these cached values cannot be the sole determinant of choices in simple economic decision making. These data challenge the fundamental tenet of contemporary theories of decision making which posit that dopamine-associated cached values are sufficient to serve as the basis for action selection.

Author contributions: N.G.H., M.M.A., J.O.G., M.E.W., and P.E.M.P. designed research; N.G.H., M.M.A., and J.O.G. performed research; N.G.H. and M.M.A. analyzed data; and N.G.H., M.M.A., and P.E.M.P. wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. Email: pemp@uw.edu.

This article is a PNAS Direct Submission.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10. 1073/pnas.1419770111/-/DCSupplemental.

problematic implications in the context of decision making. Namely, by extrapolating from these studies (35-39), one might infer that when a decision involves the tradeoff between these economic dimensions, the rank order of the dopamine-associated cached values for each of the available options would not consistently reflect the ordinal utility of these options and therefore these cached values could not, on their own, be the basis of choices. However, this counterintuitive prediction was not tested explicitly by any of these previous studies; thus it remains a provocative notion that merits direct examination, because it is contrary to the prevailing hypothesis described above which is fundamental to contemporary theories of decision making. Therefore, we investigated interactions between dimensions that previously have been shown during independent manipulations to be weakly or strongly incorporated into these cached values. Specifically, we increased the amount of effort required to obtain a large reward so that animals instead preferred a low-effort option yielding a smaller reward, and we used fast-scan cyclic voltammetry to record cue-evoked mesolimbic dopamine release as a neurochemical proxy for each option's cached value. These conditions permitted us to test whether the cached values reported via cue-evoked dopamine indeed align with animals' subjective preferences across these mixed cost-benefit attributes.

#### Results

Food-restricted rats performed an instrumental decision-making task with mixed reward and effort contingencies (Fig. 1A). Sessions consisted of repeating blocks of four single-option forced trials in which only one of the two options was available followed by four choice trials in which both options were available concurrently. After a  $45 \pm 15$  s variable intertrial interval (ITI), each trial began with the onset of one or both cue lights and the simultaneous extension of the corresponding lever(s). One lever was a low-value/low-effort (LL) reference option, yielding one food pellet for four lever presses. The alternative option vielded a high-value reward (four pellets) for a medium-effort requirement (eight presses) in the moderate-cost condition (high value/medium effort, HM) or for a high-effort requirement in the high-cost condition (high value/high effort, HH). During initial training before dopamine transmission in the high-cost condition was recorded, this high-effort requirement was determined individually for each rat so that it preferred the LL option. This high response cost ranged from 32-48 lever presses among rats but remained constant within each session for a given rat. A pair of voltammetry recordings (one session per counterbalanced lever side assignments for the high- and low-value options) was conducted for each condition. The behavioral criterion was defined as 75% choice for the HM option in moderate-cost sessions and for the LL option in high-cost sessions within a sliding window of 12 consecutive choice trials. After reaching this criterion, rats performed four additional blocks (32 trials), which provided the data analyzed from each session. Rats' postcriterion choices revealed significant preferences for the HM option in the moderate-cost condition (Fig. 1B;  $t_8 = 7.095$ , P =0.0001) and for the LL option in the high-cost condition (Fig. 1C;  $t_6 = 2.923, P = 0.0265$ ).

To monitor mesolimbic dopamine transmission, we used fastscan cyclic voltammetry at carbon-fiber microelectrodes chronically implanted in the nucleus accumbens core (Fig. S1) (40). Voltammetric recordings from postcriterion trials revealed phasic increases in dopamine concentration following cue onset for all trial types. Cue-evoked dopamine release during forced trials in the moderate-cost condition was greater in HM trials than in LL trials (Fig. 1D). To quantify this selectivity, at each time point, we calculated a dopamine discriminability index (41) based on the area under the receiver operating characteristic (auROC) curve (42), which indicates the probability that an ideal observer could



**Fig. 1.** Task design, behavioral performance, and voltammetry results from forced trials with simultaneous cue and lever onset (first cohort). (*A*) Reward and effort contingencies for moderate-cost and high-cost sessions (the box outlined in green signifies the preferred option). (*B* and *C*) Mean (+ SEM) postcriterion percent choice for moderate-cost sessions (\*\*\**P* = 0.0001, *n* = 9 rats) (*B*) and high-cost sessions (\**P* = 0.0265, *n* = 7 rats) (*C*). (*D* and *E*) Mean (± SEM) cue-evoked dopamine release (*Upper*) and discriminability index time series (*Lower*) in moderate-cost sessions (*P*, *n* = 11 recording sites) and in LL-preferred high-cost sessions (*E*, *n* = 10 recording sites). (*F* and *G*) Mean (± SEM) cue-evoked dopamine release (*G*, *Left*) and discriminability index time series (*G*, *Right*) in HH-preferred high-cost sessions (*n* = 11 recording sites). For each discriminability index time course in *D*, *E*, and *G*, the horizontal bar indicates time points of significant discriminability (\**P* < 0.05, permutation tests).

correctly classify the trial type to which a randomly selected response belongs. The auROC values, which range from 0 to 1, were transformed [discriminability index =  $(auROC - 0.5) \times 2$ ] so that an index approaching 1 indicates that the dopamine response to an HM trial can be discriminated reliably as greater than the response to a LL trial, an index of 0 indicates that the responses from each trial type cannot be discriminated, and an index approaching -1 indicates greater dopamine release in response to the LL than to the HM option. This discriminability index time course confirmed that the cue-evoked dopamine response to the HM option was significantly greater than the response to the LL option, an effect we observed regardless of the side to which each option was assigned in the operant chamber (i.e., across the counterbalanced pairs of recorded sessions within this group of rats; Fig. S24). The greater dopamine response to the preferred HM option is consistent with numerous previous observations

that options with greater subjective value are associated with greater dopamine-reported cached values than less-preferred options (24–36). However, in these moderate-cost sessions, animals' preferences are driven predominantly by the HM option's greater reward value, a dimension that is robustly incorporated into dopamine-associated cached values (24, 27, 32–36). Thus, this condition demonstrated that these signals encode value-related information but did not allow us to determine whether dopamine-reported cached values still correlate with subjective value when preferences are driven by a weakly encoded economic dimension.

The critical test of this hypothesis was provided by the highcost sessions in which animals reached the behavioral criterion for preferring the LL option, because the option yielding a larger reward was rendered the nonpreferred option by its high effort requirement, an attribute to which cue-evoked dopamine was relatively insensitive in previous studies (35–38). If the cached value signaled by cue-evoked dopamine reliably reflects subjective value, we would expect greater dopamine release in response to the preferred LL option. On the other hand, if the dopamine-reported cached value is more sensitive to expected reward value than to anticipated effort, we would expect a greater response to the HH option that yields a larger reward, despite the animals' subjective preferences for the LL alternative. In these high-cost sessions, cue-evoked dopamine release in the nucleus accumbens core was discriminable between HH and LL forced trials, but only after the peak response (Fig. 1E and Fig. S2B). Remarkably, the greater of these responses was for the HH option, even though the LL option was significantly preferred. Therefore, in these high-cost sessions, the relative ordering of the cached values reported by cue-evoked dopamine release was inconsistent with the subjective value of each option. In other high-cost sessions in which rats failed to reach criterion for the LL option but instead preferred the HH option (Fig. 1F;  $t_7 = 5.380, P = 0.001$ ), dopamine release again was greater for the HH option (Fig. 1G). Thus, across all session types there was a greater dopamine-associated cached value for the high-value

option (HM or HH) than for the LL option, regardless of whether the cost to obtain this high-value option was only moderately higher or was much higher and regardless of whether or not it was preferred over the LL option.

The cue-evoked mesolimbic dopamine response consistently was more sustained for the high-value than for the low-value option. However, during this sustained response, the remaining cost to obtain the reward becomes incrementally smaller with each lever press, and so this encoding pattern may reflect the dynamically increasing subjective value. To rule out this possibility, we conducted another set of experiments that allowed us to assess the cached values reported via cue-evoked dopamine release in a 5-s period before the lever was available (Fig. 2 A-C; significant postcriterion behavioral preferences in moderate-cost sessions:  $t_7 = 12.98$ ,  $P = 3.74 \times 10^{-6}$ ; HH-preferred high-cost sessions:  $t_9 = 5.267$ , P = 0.0005; and LL-preferred high-cost sessions:  $t_{11} = 4.319$ , P = 0.0012). In this interval between cue and lever presentation, the pattern of dopamine responses for each of the sessions was comparable to the previous results, in that mesolimbic dopamine release was greater for the option yielding the high-value reward in all session types regardless of whether or not it was preferred (Fig. 2 D-F). Importantly, there was significant discriminability between the high- and low-value options before lever presentation, demonstrating that the greater sustained response observed in high-value trials was not caused by its increase in subjective value as the remaining response cost was reduced with each lever press.

To examine further the relationship between these behavioral and neurochemical data, we pooled the data for each session across all the conditions in all the rats. We plotted the dopamine discriminability index (the ability to discern the dopamine signal over the 5 s following cue onset as being greater for the largerreward option than the smaller-reward option) as a function of behavioral preferences in each session, using a choice index from postcriterion behavior: choice index =  $[p(H) - 0.5] \times 2$ , where p(H) is the proportion of choices for the high-value option (HM or HH). Thus, a choice index of 1 corresponds to 100% choice



**Fig. 2.** Behavioral performance and voltammetry results from forced trials with a 5-s cue-to-lever delay (second cohort). (A–C) Mean (+ SEM) postcriterion percent choice for moderate-cost sessions (\*\*\*P = 3.74 × 10<sup>-6</sup>, n = 8 rats) (A), HH-preferred high-cost sessions (\*\*\*P = 0.0005, n = 10 rats) (B), and LL-preferred high-cost sessions (\*\*P = 0.0012, n = 12 rats) (C). (D–F) Mean ( $\pm$  SEM) cue-evoked dopamine release (Upper) and discriminability index time series (Lower) in moderate-cost sessions (n = 9 recording sites) (D), in HH-preferred high-cost sessions (n = 11 recording sites) (E), and in LL-preferred high-cost sessions (n = 15 recording sites) (F). For each discriminability index time course in D–F, the horizontal bar indicates time points of significant discriminability (\*P < 0.05, permutation tests).

for the high-value option, -1 indicates 100% choice for the LL option, and 0 indicates indifference between the two options. This analysis allowed us to test whether the dopamine-reported cached values reflect subjective value. According to this prevailing hypothesis, the data should occupy the upper-right and lower-left quadrants of the graph exclusively. That is, the dopamine discriminability index should be positive when rats preferred the high-value option (positive choice index) and negative when rats preferred the LL option (negative choice index). Indeed the majority of points with a positive choice index were in the upper-right quadrant (43 of 44 sessions), significantly higher than expected by chance ( $P = 5.16 \times 10^{-12}$ , binomial test). However, the data diverged from this model in sessions with a negative choice index, because the majority of these sessions did not have a negative dopamine discriminability index. In fact, a significant majority of the sessions with a negative choice index had a positive dopamine discriminability index (19 of 23 sessions, P = 0.0026, binomial test; Fig. 3A), favoring an alternative model in which the dopamine discriminability index is positive regardless of the animal's preference (62 of 67 sessions,  $P = 1.42 \times 10^{-13}$ , binomial test; Fig. 3B).

We next carried out a more detailed analysis of a utility model relating the dopamine-reported cached values to choice and compared this model with a model in which utility has no influence. We constructed the utility model as a regression line (y = $\beta_1 \times x + \beta_0$  constrained through the origin ( $\beta_0 = 0$ ; Fig. 3C) and the alternative model as a constant with no slope ( $\beta_1 = 0$ ; Fig. 3D). Comparing the goodness-of-fit using the second-order biascorrected Akaike Information Criterion (AICc) (43, 44) based on the residual sum of squares, the utility model provided an inferior fit to the data than did the alternative model, which explicitly did not account for utility (AICc = -97.30 vs. -154.27, respectively; weight of evidence favoring the origin-constrained slope =  $4.26 \times$  $10^{-13}$ , vs. >0.999 favoring the constant with no slope). Although the positive bias in the dopamine discriminability index, independent of preference, observed in the alternative model was evident in all session types (Fig. 3E; one-sample t tests vs. zero: moderate-cost sessions,  $t_{19} = 11.90$ ,  $P = 2.98 \times 10^{-10}$ ; HH-preferred high-cost sessions,  $t_{21} = 7.78$ ,  $P = 1.28 \times 10^{-7}$ ; LL-pre-ferred high-cost sessions,  $t_{24} = 4.67$ ,  $P = 9.61 \times 10^{-5}$ ), its magnitude differed among these conditions (one-way ANOVA:  $F_{2,64} = 4.28$ , P = 0.018; Bonferroni-corrected post hoc tests: moderate-cost vs. LL-preferred high-cost sessions, P = 0.015; LLpreferred vs. HH-preferred high-cost sessions, P = 0.312; moderate-cost vs. HH-preferred high-cost sessions, P = 0.638). Indeed, a standard linear regression model including both a slope and an intercept term as free parameters provided an improved fit (Fig. 3F; AICc = -163.45) over the origin-constrained slope (Fig. 3C), the constant line without slope (Fig. 3D), and discontinuous lines (Fig. S3 and Table S1). In this unconstrained model, the slope of the linear term differed significantly from 0 ( $\beta_1$  =  $0.199 \pm 0.057$ , t = 3.468,  $P = 9.34 \times 10^{-4}$ ) and explained a small proportion of the variance in the dopamine discriminability index  $(r^2 = 0.156)$ . Nonetheless, the y-intercept of this regression model also was significantly greater than 0 ( $\beta_0 = 0.403 \pm 0.038$ , t =10.737,  $P = 4.89 \times 10^{-16}$ ), suggesting that behaviorally indifferent rats show greater cue-evoked dopamine release to a high-value option than to a low-value option despite their lack of preference. Moreover, this regression line remained positive for all possible choice indices, meaning that dopamine release is greater for the high-value option regardless of preference. Importantly, the type of model that provided the best fit was not changed if we used the data only from the cohort of animals with the 5-s cue-to-lever delay (Fig. S4 and Table S2), if the counterbalanced session pairs were treated as independent data points (Fig. S5 and Table S3), or if we analyzed the peak dopamine release rather than the auROC-based discriminability index (Fig. S6 and Table S4).



**Fig. 3.** Models testing the relationship between dopamine-associated cached values and subjective preferences. In all panels, blue points are moderatecost sessions (n = 20), red points are LL-preferred high-cost sessions (n = 25), and purple points are HH-preferred high-cost sessions (n = 22). (A and B) The x's designate sessions violating the categorical models' predictions. (A) The expected utility categorical model predicts that all data should fall within the upper-right and lower-left quadrants (green). (B) The expected benefits categorical model predicts that all data should fall within the upper quadrants (green). (C) The expected utility regression model: linear regression constrained through origin ( $y = \beta_1 \times x$ ;  $\beta_0 = 0$ ). (D) The expected benefits regression model: constant line, no slope ( $y = \beta_0$ ;  $\beta_1 = 0$ ). (E) The average dopamine discriminability index was significantly greater than 0 for each session type (\*\*\*P < 0.001) but was lower in the LL-preferred high-cost sessions than in the moderate-cost sessions (\*P = 0.015). (F) The standard linear regression model ( $y = \beta_1 \times x + \beta_0$ ). Both the slope and intercept differ significantly from 0 (P < 0.001).

Collectively, these data indicate that, although the dopaminereported cached values showed a modest correlation with utility in these experimental paradigms, this relationship is not sufficient to make them a reliable instrument for determining choices.

#### Discussion

A preponderance of evidence supports the notion that phasic dopamine transmission functions as a neural instantiation of the temporal-difference prediction errors that drive reinforcement learning (13–24). Accordingly, changes in dopamine transmission are evoked whenever there is an unexpected reward-related event, both when reward delivery differs from expectations and when reward-predictive cues drive changes in expectation of available reward. The latter exemplifies how the dopamine response to unexpected cue presentation provides a readout of the cached value assigned to that cue through temporal-difference learning. These cached values are theorized to be used to determine action selection, so that the preferred action is the one associated with the cue with the greatest cached value (1-5). To subserve this role in decision making, the cached values need to incorporate any and all economic attributes insofar as those attributes influence subjective preferences; that is, by definition, the cached values must reliably reflect the ordinal utility of the available actions as revealed by animals' behavioral preferences. Consistent with this premise, there have been numerous reports

in which dopamine-associated cached values incorporate many economic attributes affecting animals' behavioral preferences, such as objective expected value (24-27, 35, 36), and some subjective attributes, including risk preference (30-33), temporal discounting (24, 28, 29), flavor or reward-type preference (31), perceptual uncertainty (45, 46), and even preference for advanced information (34). Even though this system is based on a cached-value ("model-free") architecture, there have been suggestions in the literature that it has access to "model-based" information derived from inferential online computation (47), further supporting the notion that the dopamine-associated cached values are a common currency for economic decision making (17). However, no matter how many positive correlations between dopamine-associated cached values and subjective preferences are observed, the existence of counterexamples in which this relationship is reversed is sufficient to demonstrate that the fundamental claim of decision-making theories-that cached values are all that is required to determine action selection-simply cannot hold as a general principle. Accordingly, the current work identifies circumstances in which there was a significant inversion between the options' ordinal utility and the rank ordering of their dopamine-reported cached values. This breakdown in the relationship between cached values and subjective value arose in situations where the animals' preferences were guided predominantly by effortful response cost, an economic dimension which previous voltammetry (35, 38) and electrophysiology (36, 37) studies have found to be weakly incorporated into the dopamineassociated cached values. Therefore, despite robust evidence that the dopamine-associated cached values do incorporate several subjective attributes such as risk preference (30-33) and temporal discounting (24, 28, 29), the current results by necessity imply that these cached values alone are insufficient for determining economic choices.

We have demonstrated that dopamine-associated cached values are positively correlated with behavioral preferences in some circumstances-those in which the benefits overshadow the costs-but are diametrically opposed to preference order in others in which the costs predominate in guiding animals' choices. Therefore, valuations used in decision making cannot be based on this simple cached-value-based system alone, because they must account for costs that influence action selection. Theorists have proposed that coexisting valuation systems compete for the control of behavioral resources (1-5, 48). Based on this competition framework, one could speculate that choices were determined by an alternative valuation system under circumstances in which costs outweigh benefits (i.e., when the dopamine-associated cached values did not correlate with preference) and that the cached values were used as the determinant of choice only when benefits overshadow costs. However, this marginalized use of the cached-value system would be quite limited, because real-world choices often are strongly influenced by aversive or energetic costs. Moreover, even when benefits do overshadow costs, there still are gradations in preferences for incremental changes in response cost (49), and so even these preferences cannot be based on a valuation system that is relatively insensitive to costs. Therefore, additional information on costs is required to perform these decision-making computations. Representations of costs indeed have been observed in areas such as the anterior cingulate cortex (50-54), insular cortex (55, 56), and basolateral amygdala (49, 57), all of which provide glutamatergic inputs that converge on striatal projection neurons. Therefore, a more integrative valuation system could arise from the downstream combination of benefit information from the dopamine-associated cached values and cost information from other neural sources. However, this concept of an incomplete valuation system requiring additional information is not accommodated in current theories (1-5, 48), whether they describe the

dopamine-associated valuation system alone or used in parallel with alternative systems. Alternatively, it remains possible that dopamine-associated cached values do not contribute to the selection process at all but rather play a more nuanced role in decision making that pertains to the performance or execution of the selected action. This scenario would place the burden of action selection on other reward-related structures. Indeed, representations of subjective value have been observed in multiple cortical regions (53, 58–60). Perturbation of the cached values (9, 21, 23) to test the extent of their contribution to action selection could distinguish between these possibilities in future experiments. Under any of these scenarios, however, the current findings demonstrate that the dopamine-associated cached values alone are not sufficient to serve as the basis of simple economic choices.

In the current work, we used cue-evoked mesolimbic dopamine transmission as a means to examine cached values assigned to reward options and identified circumstances in which animals preferentially selected the option that did not have the greatest cached value. This situation arose in the cost–benefit decisions of the present experiment when the differences in response costs overshadowed the differences in benefits. These findings demonstrate a direct violation of the fundamental principle that these cached values reflect animals' subjective preferences and are sufficient for determining choices. Therefore, we conclude that dopamine-associated cached values cannot be used as the sole determinant of cost–benefit decision making.

#### **Materials and Methods**

See SI Materials and Methods for additional details.

**Subjects and Surgery.** All procedures were approved by the University of Washington Institutional Animal Care and Use Committee. Male Sprague–Dawley rats (Charles River Laboratories), 250–300 g upon arrival, were anesthetized with isoflurane for bilateral implantation of carbon-fiber microelectrodes (40) targeting the nucleus accumbens core (1.3 mm anterior, 1.3 mm lateral, 6.8–7.0 mm ventral to bregma) and an Ag/AgCl reference electrode. After at least 1 wk of postsurgery recovery, rats were food-restricted to 90% of their ad libitum body weight with free access to water in the animals' home cages.

Mixed-Contingency Decision-Making Task. All sessions consisted of blocks of four single-option forced trials in which only one of the two options was available followed by four choice trials in which both options were concurrently available. A 45  $\pm$  15 s variable ITI separated each discrete trial, with a maximum of 120 trials per session. As in initial training (SI Materials and Methods), for the first cohort of rats each trial began immediately after the ITI with the onset of one or both cue lights and the simultaneous extension of the corresponding lever(s); for the second cohort the lever(s) extended 5 s after the onset of the cue light(s). The mixed-contingency decision-making task consisted of two types of sessions: moderate-cost and high-cost conditions (Fig. 1A). In both conditions, one lever served as a low value/low effort (LL) reference option, yielding one food pellet for four lever presses. The alternative option yielded a high-value reward (four pellets) for a medium effort requirement (eight presses) in the moderate-cost condition (high value/ medium effort, HM) or for a high effort requirement in the high-cost condition (high value/high effort, HH). Before voltammetric recordings were conducted (SI Materials and Methods), this high effort requirement was determined individually for each rat such that it preferred the LL option. The final high effort requirements used in recording sessions ranged from 32-48 lever presses for rats in the first cohort and 32-128 presses for rats in the second cohort. Behavioral criterion was defined as 75% choice for the HM option in the moderate-cost condition and for the LL option in the highcost condition within a sliding 12-choice window. After reaching this criterion, rats performed four additional blocks (32 trials), which provided the data analyzed from each recording session. We also obtained recordings from high-cost sessions in which rats did not reach the intended criterion for the LL option and instead reached the opposite criterion, preferring the HH option. The high effort requirements from these HHpreferred high-cost sessions ranged from 32-64 presses for the first cohort and 32–128 presses for the second cohort.

Fast-Scan Cyclic Voltammetry Recording Sessions. Chronically implanted carbon-fiber microelectrodes were connected to a head-mounted voltammetric amplifier for dopamine detection by fast-scan cyclic voltammetry as previously described (40). During recording sessions, a potential of -0.4 V (versus the Ag/AgCl reference electrode) was applied to the carbon fiber and ramped to +1.3 V and back at a rate of 400 V/s at 10 Hz. To confirm that electrodes were capable of detecting chemically verified dopamine, a series of unexpected food pellets was delivered before and after each recording session. The voltammetry data from a recording session were included in the analysis only if the pre- and postsession pellet deliveries elicited dopamine

- 1. Sutton RS, Barto AG (1998) Reinforcement Learning: An Introduction (MIT Press, Cambridge, MA).
- Daw ND, Doya K (2006) The computational neurobiology of learning and reward. Curr Opin Neurobiol 16(2):199–204.
- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9(7):545–556.
- Kable JW, Glimcher PW (2009) The neurobiology of decision: Consensus and controversy. Neuron 63(6):733–745.
- Lee D, Seo H, Jung MW (2012) Neural basis of reinforcement learning and decision making. Annu Rev Neurosci 35:287–308.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310(5752):1337–1340.
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58(3):451–463.
- Cai X, Kim S, Lee D (2011) Heterogeneous coding of temporally discounted values in the dorsal and ventral striatum during intertemporal choice. *Neuron* 69(1):170–182.
- Tai LH, Lee AM, Benavidez N, Bonci A, Wilbrecht L (2012) Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. *Nat Neurosci* 15(9):1281–1289.
- Reynolds JN, Hyland BI, Wickens JR (2001) A cellular mechanism of reward-related learning. Nature 413(6851):67–70.
- Shen W, Flajolet M, Greengard P, Surmeier DJ (2008) Dichotomous dopaminergic control of striatal synaptic plasticity. *Science* 321(5890):848–851.
- Pawlak V, Kerr JN (2008) Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. J Neurosci 28(10):2435–2446.
- Houk JC, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. *Models of Information Processing in the Basal Ganglia*, eds Houk JC, Davis JL, Beiser DG (MIT Press, Cambridge, MA), pp 249–270.
- Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16(5):1936–1947.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275(5306):1593–1599.
- Glimcher PW (2011) Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. Proc Natl Acad Sci USA 108(Suppl 3): 15647–15654.
- Schultz W (2013) Updating dopamine reward signals. Curr Opin Neurobiol 23(2): 229–238.
- Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412(6842):43–48.
- Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47(1):129–141.
- Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482(7383):85–88.
- Steinberg EE, et al. (2013) A causal link between prediction errors, dopamine neurons and learning. Nat Neurosci 16(7):966–973.
- Hart AS, Rutledge RB, Glimcher PW, Phillips PEM (2014) Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. J Neurosci 34(3):698–704.
- Stopper CM, Tse MT, Montes DR, Wiedman CR, Floresco SB (2014) Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron* 84(1):177–189.
- 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.
- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299(5614):1898–1902.
- Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H (2004) Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43(1): 133–143.
- Tobler PN, Fiorillo CD, Schultz W (2005) Adaptive coding of reward value by dopamine neurons. *Science* 307(5715):1642–1645.
- Kobayashi S, Schultz W (2008) Influence of reward delays on responses of dopamine neurons. J Neurosci 28(31):7837–7846.
- Day JJ, Jones JL, Wightman RM, Carelli RM (2010) Phasic nucleus accumbens dopamine release encodes effort- and delay-related costs. *Biol Psychiatry* 68(3):306–309.
- Sugam JA, Day JJ, Wightman RM, Carelli RM (2012) Phasic nucleus accumbens dopamine encodes risk-based decision-making behavior. *Biol Psychiatry* 71(3):199–205.

release whose cyclic voltammogram (electrochemical signature) achieved a high correlation ( $r^2 \ge 0.75$  by linear regression) with that of a dopamine standard.

ACKNOWLEDGMENTS. We thank S. Ng-Evans, K. Reinelt, and K. Dofredo for technical assistance and A. Hart, G. Boynton, and G. Horwitz for advice on data analysis. This work was supported by a National Science Foundation Graduate Research Fellowship (to N.G.H.), a Wellcome Trust Research Career Development Fellowship (to M.E.W.), and National Institutes of Health Grants GM007108 and DA036278 (to N.G.H.), DA007278 (to M.M.A.), and MH079292, DA027858, and AG044839 (to P.E.M.P.).

- Lak A, Stauffer WR, Schultz W (2014) Dopamine prediction error responses integrate subjective value from different reward dimensions. *Proc Natl Acad Sci USA* 111(6): 2343–2348.
- Stauffer WR, Lak A, Schultz W (2014) Dopamine reward prediction error responses reflect marginal utility. *Curr Biol* 24(21):2491–2500.
- Nasrallah NA, et al. (2011) Risk preference following adolescent alcohol use is associated with corrupted encoding of costs but not rewards by mesolimbic dopamine. *Proc Natl Acad Sci USA* 108(13):5466–5471.
- Bromberg-Martin ES, Hikosaka O (2009) Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* 63(1):119–126.
- Gan JO, Walton ME, Phillips PEM (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13(1):25–27.
- Pasquereau B, Turner RS (2013) Limited encoding of effort by dopamine neurons in a cost-benefit trade-off task. J Neurosci 33(19):8288–8300.
- 37. Ravel S, Richmond BJ (2006) Dopamine neuronal responses in monkeys performing visually cued reward schedules. *Eur J Neurosci* 24(1):277–290.
- Wanat MJ, Kuhnen CM, Phillips PEM (2010) Delays conferred by escalating costs modulate dopamine release to rewards but not their predictors. J Neurosci 30(36): 12020–12027.
- Fiorillo CD (2013) Two dimensions of value: Dopamine neurons represent reward but not aversiveness. Science 341(6145):546–549.
- Clark JJ, et al. (2010) Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. Nat Methods 7(2):126–129.
- Kepecs A, Uchida N, Zariwala HA, Mainen ZF (2008) Neural correlates, computation and behavioural impact of decision confidence. *Nature* 455(7210):227–231.
- Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics (Wiley, New York).
- Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19(6):716–723.
- Burnham KP, Anderson DR (2002) Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach (Springer, New York).
- Nomoto K, Schultz W, Watanabe T, Sakagami M (2010) Temporally extended dopamine responses to perceptually demanding reward-predictive stimuli. J Neurosci 30(32):10692–10702.
- de Lafuente V, Romo R (2011) Dopamine neurons code subjective sensory experience and uncertainty of perceptual decisions. Proc Natl Acad Sci USA 108(49):19767–19771.
- Bromberg-Martin ES, Matsumoto M, Hong S, Hikosaka O (2010) A pallidus-habenuladopamine pathway signals inferred stimulus values. J Neurophysiol 104(2):1068–1076.
- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8(12):1704–1711.
- Ghods-Sharifi S, St Onge JR, Floresco SB (2009) Fundamental contribution by the basolateral amygdala to different forms of decision making. J Neurosci 29(16): 5251–5259.
- Walton ME, Bannerman DM, Alterescu K, Rushworth MF (2003) Functional specialization within medial frontal cortex of the anterior cingulate for evaluating effortrelated decisions. J Neurosci 23(16):6475–6479.
- Rudebeck PH, Walton ME, Smyth AN, Bannerman DM, Rushworth MF (2006) Separate neural pathways process different decision costs. *Nat Neurosci* 9(9):1161–1168.
- Hillman KL, Bilkey DK (2010) Neurons in the rat anterior cingulate cortex dynamically encode cost-benefit in a spatial decision-making task. J Neurosci 30(22):7705–7713.
- Kennerley SW, Behrens TE, Wallis JD (2011) Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat Neurosci 14(12):1581–1589.
- Amemori K, Graybiel AM (2012) Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. *Nat Neurosci* 15(5):776–785.
- Prévost C, Pessiglione M, Météreau E, Cléry-Melin ML, Dreher JC (2010) Separate valuation subsystems for delay and effort decision costs. J Neurosci 30(42):14080–14090.
- Palminteri S, et al. (2012) Critical roles for anterior insula and dorsal striatum in punishment-based avoidance learning. *Neuron* 76(5):998–1009.
- McHugh SB, et al. (2014) Aversive prediction error signals in the amygdala. J Neurosci 34(27):9024–9033.
- Padoa-Schioppa C (2011) Neurobiology of economic choice: A good-based model. Annu Rev Neurosci 34:333–359.
- Kim S, Hwang J, Lee D (2008) Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron* 59(1):161–172.
- Louie K, Glimcher PW (2010) Separating value from choice: Delay discounting activity in the lateral intraparietal area. J Neurosci 30(16):5498–5507.

# **Supporting Information**

### Hollon et al. 10.1073/pnas.1419770111

### SI Materials and Methods

Subjects and Surgery. All procedures were approved by the University of Washington Institutional Animal Care and Use Committee. A total of 41 male Sprague-Dawley rats (Charles River Laboratories), 250-300 g upon arrival, were used for this study. Fourteen rats in the first cohort and 14 in the second cohort completed recording sessions included in this study; four additional rats were excluded because of electrode misplacement, three because of failure of electrodes to satisfy criteria for dopamine detection, and six because of postsurgical complications (e.g., headcap loss). Rats were maintained on a 12-h light/dark cycle (lights on at 0700), with all behavioral testing occurring during the light phase. Rats were pair-housed until surgery, after which they were housed individually. Rats were anesthetized with isoflurane for bilateral implantation of carbon-fiber microelectrodes (1) targeting the nucleus accumbens core (1.3 mm anterior, 1.3 mm lateral, 6.8-7.0 mm ventral to bregma) and a Ag/AgCl reference electrode. After at least 1 wk of recovery postsurgery, rats were food-restricted to 90% their ad libitum body weight; for all subsequent behavioral procedures, each rat received a total of ~15 g of food/d consisting of pellets earned as reward during behavioral sessions plus standard laboratory chow after these sessions. Water was available ad libitum in the animals' home cages.

**Initial Behavioral Training.** In their home cages before the first session of training, the food-restricted rats were exposed to the 45-mg food pellets (Bio-Serv dustless precision pellets) that served as rewards for all subsequent sessions. All training sessions took place between 0800 and 1800 in one of four standard operant chambers (Med Associates). Each chamber was equipped with a central food magazine and magazine light, a retractable lever on either side of the magazine (6 cm above the grid floor), a cue light above each lever, a house light at the top back left corner of the chamber, and a ventilation fan on the back wall of the sound-attenuating cabinet around the operant chamber.

Rats underwent one session of magazine training in which a total of 60 food pellets were delivered noncontingently, one at a time with a variable time interval ( $60 \pm 20$  s). Training to press levers for food pellets began the next day. One of the two cue lights was illuminated (side counterbalanced between rats), the corresponding lever was extended continuously for the duration of the session, and each press was reinforced on a fixed ratio (FR)-1 continuous reinforcement schedule until rats received 100 pellets in a 2-h session. If rats did not press the lever within 15–20 min, a pellet was placed behind the lever to encourage the rat to interact with the lever. In the next session, the other cue light was illuminated, and the lever was extended for another 100-pellet session reinforced on a continuous FR-1 schedule.

All subsequent sessions consisted of training on discrete trials: After the rat completed the response requirement, the cue light turned off, the lever retracted, a pellet was delivered into the food magazine, and the magazine light was illuminated for 6 s, followed by a variable ITI. At this stage of training, only one lever was available on each trial (all trials were forced), and each session consisted of 80 total trials (40 for each lever). For the first cohort of rats, each trial began at the end of the ITI with the simultaneous onset of a cue light and extension of the corresponding lever. For the second cohort, the cue light came on 5 s before lever extension. Rats completed one session of FR-1 training with a  $20 \pm 5$  s ITI and unlimited time to initiate responding on each trial. For all subsequent behavioral sessions, rats were connected to a head-stage containing a voltammetric amplifier to habituate them to the equipment used for eventual recording sessions. While tethered, subsequent training consisted of one session each of FR-1 with a 20  $\pm$  5 s ITI, FR-4 with a 30  $\pm$  10 s ITI, and FR-8 and FR-16 with a  $45 \pm 15$  s ITI for all sessions thereafter. Starting in the FR-16 session, failure to initiate responding within 10 s of lever presentation resulted in an unrewarded "Miss." Training on FR-16 sessions continued until rats completed more than 90% of the trials in a session. Rats then performed behavioral decisionmaking sessions that included blocks of four single-option forced and four dual-option choice trials, during which either the reward magnitude or effort requirement differed between the two options, as in our previous study that included independent manipulations of reward and effort (2). For the reward-manipulation sessions, each option required four lever presses, with one lever yielding four pellets and the other yielding one pellet; for the effort-manipulation sessions, each option yielded one food pellet, with one requiring four lever presses and the other requiring 32 presses. The contingencies assigned to each lever side were reversed between each session, and each rat performed daily sessions of either the reward or effort manipulation (order counterbalanced) until it reached criterion (75% choice in a sliding window of 12 choice trials) in fewer than 80 trials for both lever side assignments. After completing both the reward- and effort-manipulation stages, rats then advanced to the mixedcontingency decision-making task described below.

**Mixed-Contingency Decision-Making Task.** All sessions consisted of blocks of four single-option forced trials in which only one of the two options was available followed by four choice trials in which both options were available concurrently. In choice trials, the unchosen lever retracted and the cue light turned off once the rat made an initial press on the chosen lever. A  $45 \pm 15$  s variable ITI separated each discrete trial, with a maximum of 120 trials per session. As in prior training, for the first cohort of rats each trial began immediately after the ITI with the onset of one or both cue lights and the simultaneous extension of the corresponding lever(s), and for the second cohort the lever(s) extended 5 s after the onset of the cue light(s).

The mixed-contingency decision-making task consisted of two types of sessions: moderate-cost and high-cost conditions (Fig. 1A). In both conditions, one lever served as a low value/low effort (LL) reference option, yielding one food pellet for four lever presses. The alternative option yielded a high-value reward (four pellets) for a medium-effort requirement (eight presses) in the moderate-cost condition (high value/medium effort: HM) or for a high-effort requirement in the high-cost condition (high value/ high effort: HH). Before voltammetric recordings were conducted, this high-effort requirement was determined individually for each rat so that the rat preferred the LL option. For the first cohort, the lever side assigned to the low- vs. high-value options was reversed every two behavioral sessions; if a rat did not reliably prefer the LL option in both side configurations, the effort requirement for the HH option was increased by eight presses for the subsequent high-cost session, but the effort requirement always remained constant within a given session. The same procedure was used to determine the high-effort requirement for the second cohort, except the lever side assignments were reversed pseudorandomly every one to three sessions. The final high-effort requirements used in recording sessions ranged from 32-48 lever presses for rats in the first cohort and from 32-128 presses for rats in the second cohort. For both cohorts, recordings were conducted after rats had performed at least eight behavioral

sessions of a given condition (2) and always were conducted on the second session with a given lever side assignment. The behavioral criterion was defined as 75% choice for the HM option in the moderate-cost condition and for the LL option in the highcost condition within a sliding 12-choice window. After reaching this criterion, rats performed four additional blocks (32 trials) that provided the primary data analyzed from each recording session. We also obtained recordings from high-cost sessions in which rats did not reach the intended criterion for the LL option and instead reached the opposite criterion, preferring the HH option. The high-effort requirements from these HH-preferred high-cost sessions ranged from 32–64 presses for the first cohort and 32–128 presses for the second cohort.

Fast-Scan Cyclic Voltammetry Recording Sessions. The chronically implanted carbon-fiber microelectrodes were connected to a head-mounted voltammetric amplifier for dopamine detection by fast-scan cyclic voltammetry as previously described (1). A potential of -0.4 V (versus the Ag/AgCl reference electrode) was applied to the carbon-fiber microelectrode and ramped to +1.3 V and back at a rate of 400 V/s. This voltammetric scan was applied at a frequency of 60 Hz for ~40 min before the behavioral sessions were recorded and then at 10 Hz for ~20 min before and throughout the recording session. To confirm that electrodes were capable of detecting chemically verified dopamine, a series of unexpected food pellets was delivered before and after each recording session. The voltammetry data from a recording session were included in the analysis only if the pre- and postsession pellet deliveries elicited dopamine release whose cyclic voltammogram (electrochemical signature) achieved a high correlation ( $r^2 \ge 0.75$ by linear regression) with that of a dopamine standard.

**Statistical Analyses.** Postcriterion choice proportions were normalized with the arcsine transformation and compared with indifference using two-tailed, one-sample *t* tests in SPSS (IBM). Voltammetry data analysis was carried out using software written in LabView and Matlab. Following 2000-Hz low-pass filtering, dopamine was isolated from the background (1 s before cue onset)-subtracted voltammetric signal using chemometric analysis (3) using a standard training set based on stimulated dopamine release detected by chronically implanted electrodes (1). Dopamine concentration was estimated based on the average postimplantation electrode sensitivity. Noise spikes >1.5 nA versus the immediately preceding and following time points were removed (2), and the data were smoothed using a 0.5-s moving average.

The discriminability of cue-evoked dopamine responses in the different forced trial types was analyzed at each time point using the auROC, an approach from signal-detection theory (4). The high-value option (HM or HH) always was coded as the positive case in comparisons with the LL option, and auROC values were not rectified around 0.5 (i.e., if the LL option had evoked a greater dopamine response, the auROC values would have been less than 0.5). Significant discriminability at each time point was determined using a random permutation test, shuffling the trial types and recomputing the auROC and repeating this process for 2,000 permutations to generate a null distribution. After correcting for multiple comparisons across time using a supra-

- Clark JJ, et al. (2010) Chronic microsensors for longitudinal, subsecond dopamine detection in behaving animals. Nat Methods 7(2):126–129.
- Gan JO, Walton ME, Phillips PEM (2010) Dissociable cost and benefit encoding of future rewards by mesolimbic dopamine. Nat Neurosci 13(1):25–27.
- Heien ML, et al. (2005) Real-time measurement of dopamine fluctuations after cocaine in the brain of behaving rats. Proc Natl Acad Sci USA 102(29):10023–10028.
- Green DM, Swets JA (1966) Signal Detection Theory and Psychophysics (Wiley, New York).

threshold cluster-correction technique (5, 6), all time points outside the 95% confidence interval were considered statistically significant. For graphical display purposes, all auROC values were transformed to a dopamine discriminability index ranging from -1 to 1: discriminability index = 2 × (auROC - 0.5).

To test the relationship between dopamine-associated cached values and subjective preference, we computed a dopamine discriminability index and a choice index to summarize each recorded session. The mean change in dopamine concentration over 5 s following cue onset for each postcriterion trial was used to calculate the auROC for a given session (HM or HH trials as positive cases, LL trials as negative cases), and each session's auROC was transformed to a dopamine discriminability index as above: discriminability index =  $2 \times (auROC - 0.5)$ . Thus, a dopamine discriminability index approaching 1 indicates a greater dopamine response to the high-value option (HM or HH), whereas an index of -1 indicates a greater response to the LL option, and an index of 0 indicates equivalent dopamine release to either option. Likewise, the choice index was calculated by transforming the postcriterion choice behavior in each session: choice index =  $2 \times [p(H) - 0.5]$ , where p(H) is the proportion of choices for the high-value option (HM or HH), so that a choice index of 1 corresponds to 100% choice for the high-value option, -1 indicates 100% choice for the LL option, and 0 indicates indifference between the two options. Categorical models of expected utility and expected benefits were evaluated with binomial tests of the number of sessions violating or satisfying the models' predictions, and regression models were evaluated by comparing the goodness-of-fit using the second-order AICc (7, 8) based on the residual sum of squares.

Finally, to test for the possibility of direction-selective encoding by mesolimbic dopamine, we examined the counterbalanced pairs of recorded sessions from the first cohort of rats to compare the cue-evoked dopamine response during forced trials for each option when it was assigned to the lever side ipsilateral versus contralateral to the hemisphere of the recording electrode. Within each trial type, the dopamine responses were indistinguishable between the two lever side assignments (Fig. S2). Moreover, we observed the same pattern of greater dopamine transmission for the high-value option regardless of the lever assignment configuration. Because these results do not reflect direction encoding by mesolimbic dopamine, for the second cohort we included all recorded sessions meeting the electrochemical and behavioral criteria regardless of whether the counterbalanced pair was obtained.

Histological Verification of Recording Site. Animals were anesthetized with ketamine (100 mg/kg) and xylazine (20 mg/kg), and the recording site was marked by passing a current (~70  $\mu$ A) through the carbon-fiber microelectrode for 20 s to make a small electrolytic lesion. Animals were perfused transcardially with physiological saline and then with 4% (wt/vol) paraformaldehyde in PBS, in which brains also were postfixed after removal from the skull. Brains were sunk in 15% (wt/vol) sucrose solution in PBS for 24 h, in 30% (wt/vol) sucrose for at least 72 h, flash frozen in dry ice, sectioned coronally (30–60  $\mu$ m) on a cryostat, mounted on slides, and stained with a 0.5% cresyl violet solution.

Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum Brain Mapp* 15(1):1–25.

Buschman TJ, Denovellis EL, Diogo C, Bullock D, Miller EK (2012) Synchronous oscillatory neural ensembles for rules in the prefrontal cortex. *Neuron* 76(4):838–846.

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19(6):716–723.

<sup>8.</sup> Burnham KP, Anderson DR (2002) Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach (Springer, New York).



Fig. S1. Recording locations in the nucleus accumbens core in the first (A) and the second (B) cohorts of rats. The numbers next to each section indicate distance in millimeters anterior to bregma. Adapted from Paxinos G, Watson C (2005) Rat Brain in Stereotaxic Coordinates (Elsevier Academic, Burlingame, MA) 5th Ed.

**DNAS** 

S A Z



**Fig. 52.** Lack of direction-selective encoding by mesolimbic dopamine. (*A*, *Upper*) Mean ( $\pm$  SEM) cue-evoked dopamine release during forced trials from the side-counterbalanced pairs of moderate-cost sessions recorded in rats in the first cohort, separated by lever side assignment. (*Upper Left*) Sessions in which the HM option (blue) was assigned to the lever ipsilateral to the hemisphere of the carbon-fiber microelectrode and the LL option (gray) was contralateral. (*Upper Center*) Sessions in which the HM option (blue) was assigned to the lever contralateral to the hemisphere of the electrode and the LL option (gray) was contralateral. (*Upper Right*) Mean ( $\pm$  SEM) discriminability index time series comparing forced trials from the side-counterbalanced pairs. Neither the HM contralateral vs. HM ipsilateral comparison (blue) nor the LL contralateral vs. LL ipsilateral comparison (gray) ever reached significance. (*Lower*) Mean ( $\pm$  SEM) discriminability index time series comparing HM vs. LL forced trials within each session, with lever assignments defined as above. Horizontal bars indicate time points of significant discriminability (\**P* < 0.05, permutation tests). (*B*, *Upper*) Mean ( $\pm$  SEM) cue-evoked dopamine release during forced trials from the side-counterbalanced pairs of high-cost sessions recorded in rats in the first cohort reaching behavioral criterion for preferring the LL option, separated by lever side assignment. (*Upper Left*) Sessions in which the HH option (red) was assigned to the lever ipsilateral to the hemisphere of the electrode and the LL option (gray) was contralateral. (*Upper Right*) Mean ( $\pm$  SEM) discriminability index time series comparing forced trials from the side-counterbalanced pairs of high-cost sessions recorded in rats in the first cohort reaching behavioral criterion for preferring the LL option, separated by lever side assignment. (*Upper Left*) Sessions in which the HH option (red) was assigned to the lever contralateral to the hemisphere of the electrode and



**Fig. S3.** Additional models testing the relationship between dopamine-associated cached values and subjective preferences. (A) Modeling the discriminability index with two constants, depending on whether the low- or high-value option was preferred (AICc = -158.86). (B) Separate linear regressions for cases in which the low- vs. high-value option was preferred (AICc = -159.53). Each provided a better fit than either the origin-constrained utility model (AICc = -97.30, Fig. 3C) or the single-constant without slope (AICc = -154.27, Fig. 3D) but was not better than the standard linear regression model (AICc = -163.45, Fig. 3F). As in Fig. 3, blue points represent moderate-cost sessions, red points are LL-preferred high-cost sessions, and purple points are HH-preferred high-cost sessions.



**Fig. 54.** Models from Fig. 3 and Fig. S3 including only the data from the second cohort (5-s cue-to-lever delay). (*A*) The expected utility regression model: constrained through origin (AlCc = -54.26). (*B*) The constant line without slope (AlCc = -87.25). (*C*) The standard linear regression (AlCc = -90.80;  $r^2 = 0.1564$ ). Both the slope and intercept differ significantly from 0 ( $\beta_1 = 0.162 \pm 0.066$ , t = 2.473, P = 0.019;  $\beta_0 = 0.368 \pm 0.045$ , t = 8.210,  $P = 1.76 \times 10^{-9}$ ). (*D*) Two constants, depending on whether the low- or high-value option was preferred (AlCc = -89.67). (*E*) Separate linear regressions for cases in which the low- vs. high-value option was preferred (AlCc = -85.69). In all panels, blue points are moderate-cost sessions, red points are LL-preferred high-cost sessions, and purple points are HH-preferred high-cost sessions.



**Fig. 55.** Models from Fig. 3 and Fig. S3 splitting the pairs of counterbalanced sessions and treating each as an independent data point. (*A*) Expected utility regression model: constrained through origin (AICc = -162.50). (*B*) Constant line without slope (AICc = -248.43). (*C*) Standard linear regression (AICc = -257.66;  $r^2 = 0.0822$ ). Both the slope and intercept differ significantly from zero ( $\beta_1 = 0.172 \pm 0.050$ , t = 3.413,  $P = 8.56 \times 10 - 4$ ;  $\beta_0 = 0.411 \pm 0.035$ , t = 11.900,  $P = 1.52 \times 10 - 22$ ). (*D*) Two constants, depending on whether the low- or high-value option was preferred (AICc = -253.61). (*E*) Separate linear regressions for cases in which the low- vs. high-value option was preferred (AICc = -253.69). In *A*–*E*, blue points are moderate-cost sessions, red points are LL-preferred high-cost sessions, and purple points are HH-preferred high-cost sessions.



**Fig. S6.** Models from Fig. 3 and Fig. S3 using a peak dopamine index [(H - L)/(H + L)] instead of the auROC-based discriminability index. (*A*) Expected utility regression model: constrained through origin (AICc = -172.77). (*B*) Constant line without slope (AICc = -180.93). (*C*) Standard linear regression (AICc = -190.74;  $r^2 = 0.1639$ ). Both the slope and intercept differ significantly from zero ( $\beta_1 = 0.167 \pm 0.047$ , t = 3.570,  $P = 6.77 \times 10 - 4$ ;  $\beta_0 = 0.146 \pm 0.031$ , t = 4.776,  $P = 1.06 \times 10 - 5$ ). (*D*) Two constants, depending on whether the low- or high-value option was preferred (AICc = -185.73). (*E*) Separate linear regressions for cases in which the low- vs. high-value option was preferred (AICc = -188.27). In *A*–*E*, blue points are moderate-cost sessions, red points are LL-preferred high-cost sessions, and purple points are HH-preferred high-cost sessions.

Table S1. AICc and weights of evidence for each model fromFig. 3 and Fig. S3

Model	Free parameters	AICc	Weight for model
Origin-constrained slope	2	-97.30	$3.45  imes 10^{-15}$
Constant with no slope	2	-154.27	0.0081
Standard linear regression	3	-163.45	0.7991
Two constants	3	-158.86	0.0802
Two linear regressions	5	-159.53	0.1125

### Table S2.AICc and weights of evidence for each model fromFig. S4

Model	Free parameters	AICc	Weight for model
Origin-constrained slope	2	-54.26	$6.39  imes 10^{-9}$
Constant with no slope	2	-87.25	0.0932
Standard linear regression	3	-90.80	0.5511
Two constants	3	-89.67	0.3130
Two linear regressions	5	-85.69	0.0427

## Table S3. AICc and weights of evidence for each model from Fig. S5

Model	Free parameters	AICc	Weight for model
Origin-constrained slope	2	-162.50	$1.69 \times 10^{-21}$
Constant with no slope	2	-248.43	0.0077
Standard linear regression	3	-57.66	0.7821
Two constants	3	-253.61	0.1029
Two linear regressions	5	-253.69	0.1073

## Table S4.AICc and weights of evidence for each model fromFig. S6

Model	Free parameters	AICc	Weight for mode
Origin-constrained slope	2	-172.77	$9.11 \times 10^{-5}$
Constant with no slope	2	-180.93	0.0054
Standard linear regression	3	-190.74	0.7244
Two constants	3	-185.73	0.0593
Two linear regressions	5	-188.27	0.2108

PNAS PNAS