Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence

Joseph T McGuire & Joseph W Kable

Deciding how long to keep waiting for future rewards is a nontrivial problem, especially when the timing of rewards is uncertain. We carried out an experiment in which human decision makers waited for rewards in two environments in which reward-timing statistics favored either a greater or lesser degree of behavioral persistence. We found that decision makers adaptively calibrated their level of persistence for each environment. Functional neuroimaging revealed signals that evolved differently during physically identical delays in the two environments, consistent with a dynamic and context-sensitive reappraisal of subjective value. This effect was observed in a region of ventromedial prefrontal cortex that is sensitive to subjective value in other contexts, demonstrating continuity between valuation mechanisms involved in discrete choice and in temporally extended decisions analogous to foraging. Our findings support a model in which voluntary persistence emerges from dynamic cost/benefit evaluation rather than from a control process that overrides valuation mechanisms.

Pursuing long-run rewards often requires persistence in the face of delay and short-run costs. The capacity to delay gratification is central to the notion of self-control in human decision-making, and failures of persistence can appear to reflect impulsivity, inconsistency or self-control failure. We used functional magnetic resonance imaging (fMRI) to examine brain activity associated with sustaining or curtailing persistence toward delayed rewards.

Much is known about neural systems involved in value-based decision-making, but it is unknown what role these mechanisms have in temporally extended persistence. Most intertemporal choice research has focused on discrete choices among outcomes that differ in delay. Delay-of-gratification scenarios, in contrast, involve a prolonged delay period with a continuously available opportunity to give up. These two types of future-oriented behavior are widely thought to involve different mental processes. A previous account argued that the initial selection of a delayed reward depends on a rational cost/benefit assessment, but that the subsequent ability to wait for it depends on self-regulatory dynamics (competition between hot and cool motivational systems).

We previously hypothesized that both successes and apparent failures of persistence emerge from dynamic value maximization. Because the exact timing of future events is usually uncertain, there is no guarantee that a decision maker who was willing to begin waiting for a delayed reward should necessarily be willing to keep waiting indefinitely. In some situations, including many that seem to challenge self-control, a long delay is often predictive of a shorter delay than expected delay. One way to navigate such situations would be to reassess the subjective value of the awaited reward as time passes, based on a continuously updated estimate of the remaining delay time. Such a reassessment might be encoded in the same neural valuation system, comprised of ventromedial prefrontal cortex (VMPFC), ventral striatum (VS) and posterior cingulate cortex (PCC), that encodes subjective value in a highly general manner across many other kinds of decisions. The subjective value representations encoded in VMPFC are known to be sensitive to both immediate and delayed outcomes, primary and secondary forms of reward, goal-related and temptation-related factors, and high-level task contingencies.

Other theoretical perspectives make different predictions. One alternative possibility is that successful persistence depends principally on cognitive control mechanisms external to the valuation system. Although some accounts hold that the VMPFC valuation system mediates cognitive control, other accounts posit a form of control that overrides or competes with valuation. If the latter control mechanism is paramount, successful persistence might be better understood as rule adherence than as value maximization, and curtailing persistence might reflect a lapse in control-related brain activity. A second alternative possibility is based on the structural parallel between delay of gratification and certain kinds of foraging scenarios. It has recently been hypothesized that single-alternative foraging decisions—for example, whether to exploit one’s current food patch or depart to forage elsewhere—might depend, not on the VMPFC valuation system, but on a representation in dorsal anterior cingulate cortex (dACC) of the value of departing.

To examine valuation signals during temporally extended persistence, we conducted an fMRI experiment in which participants repeatedly decided how long to keep waiting for future monetary rewards. On each trial, the participant viewed a token that had no initial value, but matured to a value of $0.30 (30¢) after a random delay. The participant could sell the token anytime and initiate a new trial, aiming to maximize total earnings in a fixed time period. Unlike some previous studies, no small reward was delivered if the participant quit early; instead, the main incentive to quit was the possibility that the next trial might mature with a shorter delay.

Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA. Correspondence should be addressed to J.T.M. (mcguire@psych.upenn.edu) or J.W.K. (kable@psych.upenn.edu).

Received 12 January; accepted 10 March; published online 6 April 2015; doi:10.1038/nn.3994
The ideal strategy depended on the distribution of delay times, which differed between two environments (Fig. 1b,c). In a high-persistence (HP) environment, the most productive strategy was to wait for every reward (up to 40 s). In a limited-persistence (LP) environment, the best strategy was to wait 20 s and then quit if the reward had not arrived. Participants learned about the timing statistics through direct experience during preliminary training. The environments were presented in alternating 10-min runs, marked by different-colored tokens.

We predicted that participants would quit earlier in the LP environment than in the HP environment. In addition, our theoretical model predicted that participants’ subjective valuation of the awaited token would evolve differently in the two environments, increasing more rapidly with elapsed time in the HP environment than in the LP environment. Using neuroimaging, we tested whether canonically value-responsive brain regions would reflect this dynamic reassessment. We would also have been able to detect alternative possibilities such as representations of subjective value elsewhere in the brain, a lapse in control-related activity associated with quitting or a representation of the value of quitting in dACC.

RESULTS

Behavioral results

Participants (n = 20) quit before receiving the reward more often in the LP environment (median = 50.0% of trials, interquartile range (IQR) = 46.6–57.6%) than in the HP environment (median = 3.1%, IQR = 0–15.6%). In the LP environment, the time waited before quitting (median of medians) was 29.3 s (IQR = 17.6–36.6). Within-subject (across-trial) variability in quit timing was comparatively small: the median size of the within-subject interquartile range was 9.1 s.

Theoretical modeling

The passage of time can drive a dynamic reassessment of awaited rewards by furnishing information about the remaining delay. Persistence in the two environments was modestly correlated (Spearman ρn = 20 = 0.37, P = 0.11; Fig. 2b) and behavior was stable across the fMRI experiment (Supplementary Fig. 1).

Participants were willing to wait longer in the HP environment than in the LP environment. We used survival analysis to estimate each participant’s probability of ‘surviving’ various lengths of time without quitting. Figure 2a shows averaged subject-wise empirical survival curves against ideal performance. The area under the curve (AUC) estimates how much of the first 40 s a participant was willing to wait on average (Fig. 2b). Median AUC was 38.9 in the HP environment (IQR = 35.4–40; ideal = 40 s) and 30.2 s in the LP environment (IQR = 22.3–34.9; ideal = 20 s). All 20 participants persisted longer in the HP environment (median difference = 7.6 s, IQR = 3.0–14.2, signed-rank P < 0.001). Persistence in the two environments was modestly correlated (Spearman ρn = 20 = 0.37, P = 0.11; Fig. 2b) and behavior was stable across the fMRI experiment (Supplementary Fig. 1).

Reaction time (RT) to sell rewarded tokens tracked time-varying reward expectancy. When an event’s latency is uniformly distributed, expectancy theoretically increases with elapsed time. Accordingly, subject-wise Spearman correlations between delay and RT were reliably negative in the HP environment (median single-subject ρ = −0.27, IQR = −0.36 to −0.16, signed-rank P < 0.001), indicating faster responses to rewards that were preceded by longer delays (Fig. 2d) and implying that participants successfully encoded the task’s timing statistics.

Figure 1 Experimental task and timing conditions. (a) Schematic of the willingness-to-wait task. (b) Discrete probability distributions governing the scheduled delay times in each environment. (c) Expected monetary rates of return under various waiting policies, where each policy is defined by a giving-up time. The reward-maximizing policy was to wait up to 40 s in the HP environment (that is, never to quit), but only up to 20 s in the LP environment. These rates of return were contingent on the fixed 2-s inter-trial interval (ITI).

Figure 2 Behavioral results. (a) Survival curves reflecting the probability that a participant was still waiting at each elapsed time, provided that the reward had not yet been delivered. Empirical survival curves were averaged across subjects at 1-s intervals (± s.e.m.). Ideal performance is plotted for reference (dashed lines). (b) Area under the curve (AUC) values calculated from individual participants’ survival curves. The maximum possible value was 40 s. The red point marks ideal performance. All 20 participants persisted more in the HP environment. (c) Stem plots show the ground-truth hazard rate for reward in each environment; that is, the probability of the reward arriving at each time, conditional on not having arrived already. Faded lines illustrate hypothetical continuous hazard functions incorporating endogenous temporal uncertainty (Online Methods). (d) Reward RT at each delay (median and IQR of subject-wise medians). RTs are expressed as deviations from each subject’s grand-median RT (median = 475 ms, IQR = 450–506 ms) to display within-subject effects. RTs for 5–20-s delays did not differ between the environments (HP median = 472 ms, IQR = 454–538; LP median = 494 ms, IQR = 443–522).

Participants were willing to wait longer in the HP environment than in the LP environment. We used survival analysis to estimate each participant’s probability of ‘surviving’ various lengths of time without quitting. Figure 2a shows averaged subject-wise empirical survival curves against ideal performance. The area under the curve (AUC) estimates how much of the first 40 s a participant was willing to wait on average (Fig. 2b). Median AUC was 38.9 in the HP environment (IQR = 35.4–40; ideal = 40 s) and 30.2 s in the LP environment (IQR = 22.3–34.9; ideal = 20 s). All 20 participants persisted longer in the HP environment (median difference = 7.6 s, IQR = 3.0–14.2, signed-rank P < 0.001). Persistence in the two environments was modestly correlated (Spearman ρn = 20 = 0.37, P = 0.11; Fig. 2b) and behavior was stable across the fMRI experiment (Supplementary Fig. 1).

Reaction time (RT) to sell rewarded tokens tracked time-varying reward expectancy. When an event’s latency is uniformly distributed, expectancy theoretically increases with elapsed time. Accordingly, subject-wise Spearman correlations between delay and RT were reliably negative in the HP environment (median single-subject ρ = −0.27, IQR = −0.36 to −0.16, signed-rank P < 0.001), indicating faster responses to rewards that were preceded by longer delays (Fig. 2d) and implying that participants successfully encoded the task’s timing statistics.

Theoretical modeling

The passage of time can drive a dynamic reassessment of awaited rewards by furnishing information about the remaining delay. Intuitively, rewards in the HP environment grew nearer and more
subjectively valuable as time passed, but rewards in the LP environment became progressively less likely to be delivered before the participant quit.

We formalized this intuition in a theoretical model of subjective valuation. The model estimates the awaited token’s subjective value at each point in the delay interval, accounting for the changing probability distribution over remaining delay durations. Our model extends a formalism from the optimal foraging literature known as the potential function25. The expected remaining delay is multiplied by the opportunity cost of time and subtracted from the expected reward. Subjective value at a given elapsed time equals the expected net return in the remainder of the current trial, maximized over all possible giving-up times. Its minimum is zero; as the agent can always quit immediately. If the subjective value exceeds zero, this signifies that the decision maker could do better by waiting than by quitting immediately. The level of subjective value at each time reflects the margin of preference for waiting over quitting (Online Methods).

In the HP environment, the token’s theoretical subjective value increased with elapsed time, reflecting the progressive shortening of the expected remaining delay (Fig. 3a). In the LP environment, the token’s subjective value remained positive until 20 s, but then fell to zero, reflecting that the best strategy was to quit if the reward had not arrived by then. Differences between the subjective value trajectories in the two environments were primarily driven by the evolving probability that the reward would arrive before the optimal giving-up time (Supplementary Fig. 2).

We modeled the empirical behavioral data as a stochastic function of theoretical subjective value using logistic regression (Fig. 3b and Online Methods). Greater subjective value was associated with higher odds of waiting (median coefficient $= 0.26$, IQR $= 0.05–0.78$, signed-rank $P < 0.001$). The subjective value model significantly outperformed an intercept-only model (subject-wise likelihood-ratio tests: median $z = 4.26$, IQR 1.79–7.97, signed-rank $P < 0.001$) and an alternative model that directly fit different overall rates of quitting in the HP and LP conditions (subject-wise difference of model deviances: median $= 0.65$, IQR $= −1.85–32.02$, signed-rank $P = 0.033$).

Neuroimaging results

Our fMRI analyses tested for brain signals that evolved differently during physically identical delay intervals in the two environments. Trial onset–locked BOLD time courses were flexibly estimated in each environment using a finite impulse response (FIR) model consisting of a series of single time-point basis functions in a general linear model (GLM). Each trial was modeled from onset up to 1 s before the outcome (reward cue or quit response). Because trials had different durations, earlier time points were observed on more trials than later time points (Supplementary Fig. 3). Group analyses focused on the interval from 2.5–30 s, for which 19 of 20 participants contributed complete data. Because the HP and LP conditions were presented in separate scanning runs with independent baselines, our analyses focused on differential change over time, not the overall offset between the two conditions. Significance was assessed using whole-brain permutation tests to control for multiple comparisons (Online Methods).

A model-based fMRI contrast tested directly for effects of theoretical subjective value on BOLD activity. For each subject and voxel, the empirical difference time course (HP minus LP) was regressed on the predicted difference (Fig. 3c, d and Online Methods) along with a constant intercept. The resulting contrast coefficient reflected the degree to which BOLD signal increased more steeply with elapsed time in the HP environment than the LP environment. Coefficients were submitted to a whole-brain, two-tailed, group-level test ($n = 20$). This identified a single significant cluster, located in VMPFC (Fig. 4a and Table 1), in which BOLD signal was positively related to theoretical subjective value. No negative effects of subjective value on BOLD were identified, even in follow-up analyses tailored to detect signals reflecting the difficulty of persistence (Supplementary Fig. 4).
Table 1  Peak foci of BOLD effects

| Region                        | x    | y    | z    | Cluster extent | Peak value | Cluster P value |
|-------------------------------|------|------|------|----------------|------------|-----------------|
| Trial onset–locked time courses: model-based contrast (t statistic) VMPFC | 0    | 60   | 3    | 311            | 5.38       | 0.014           |
| L posterior parietal          | –42  | –72  | 45   | 42            | 4.79       | 0.003           |
| L VMPFC                       | –3   | 60   | 3    | 84            | 5.54       | 0.001           |
| R posterior parietal          | 12   | 42   | 3    | 38            | 4.23       | 0.005           |
| L superior temporal gyrus     | –54  | 6    | –15  | 16            | 4.09       | 0.034           |

Anticipatory activity in quit-related time courses: main effect of time (F statistic), −12.5 s to −2.5 s

| Region                        | x    | y    | z    | Cluster extent | Peak value | Cluster P value |
|-------------------------------|------|------|------|----------------|------------|-----------------|
| R posterior parietal          | 21   | –72  | 51   | 393            | 9.32       | 0.001           |
| L posterior parietal          | –21  | –72  | 48   | 59            | 5.20       | 0.042           |
| R anterior insula             | 33   | 27   | 6    | 140           | 7.45       | 0.007           |
| DMFC                          | 3    | 12   | 45   | 127           | 7.45       | 0.007           |
| R anterior PFC                | 33   | 54   | 27   | 70            | 7.65       | 0.024           |

The observed effect in VMPFC echoes the effects of subjective value that are seen in a broad range of other contexts3–6. We formally juxtaposed our results with previous findings by quantifying the spatial overlap between our empirical results and canonically valuation-related brain regions derived from a 206-study meta-analysis5 (Fig. 4a). The meta-analysis had identified clusters showing preferentially positive effects of value in VMPFC (9.67 cm³), striatum (21.41 cm³) and PCC (2.62 cm³). There was a 100-voxel (2.70 cm³) region of overlap in VMPFC (27.9% of the canonical region and 32.2% of the empirical cluster).

As an alternative test of the same question, the three canonical valuation areas were tested as regions of interest (ROIs). Model-based contrast coefficients were spatially averaged in each ROI for each participant. The effect of subjective value was significant in VMPFC (signed-rank P = 0.002), but non-significant, albeit with a positive trend, in striatum (P = 0.079) and PCC (P = 0.062; Fig. 4b). Paired-samples comparisons identified a greater effect in VMPFC than striatum (signed-rank P = 0.012) and no significant differences between the other two pairs of ROIs (P > 0.11).

In summary, results suggest that the region of VMPFC previously found to encode subjective value during discrete choices and outcomes also reflects a dynamic reassessment of subjective value during voluntarypersistences. This was true to a greater degree for VMPFC than striatum.

We also conducted a less-constrained analysis that could detect BOLD time course differences predicted by either our model or alternative frameworks. Trial onset–locked time courses were analyzed at the group level in a whole-brain voxel-wise repeated-measures ANOVA (n = 19), with factors for condition (HP versus LP) and time point. We focused on the condition x time point interaction, seeking to identify signals that exhibited different patterns of change over time in the two environments. This analysis avoids a priori assumptions about either the form of the difference or the location of effects in the brain. A significant interaction was observed in left and right VMPFC, left posterior parietal cortex, and a small region of left superior temporal gyrus (Table 1 and Fig. 5a). Time course plots (Fig. 5b–e) suggested that, in VMPFC and parietal regions, the effect took the form of a greater signal increase with elapsed time in the HP environment, consistent with theoretical subjective value.

Further analyses tested for evidence of reward prediction error (RPE) signals30. When a reward occurs, RPE is the difference between the obtained and expected outcome. Because reward expectancy theoretically rose over time in the HP environment (Fig. 2c; see also RT data above and heart rate data below), rewards at short delays should have been more surprising and evoked larger RPEs than rewards at longer delays. We tested whether the amplitude of the phasic BOLD response to reward was modulated by the delay duration that preceded it. A negative effect would reflect an RPE-like pattern (smaller reward responses after longer delays, a pattern seen previously in the firing rates of dopaminergic midbrain neurons31). To focus on phasic reward responses while controlling for nonspecific effects of elapsed time, we compared the modulatory effect in the post-reward epoch against the same effect in a pre-reward epoch.

We observed no significant negative modulatory effect of elapsed time on the reward-related BOLD response in any location. We did, however, identify an occipitoparietal cluster with an effect in the opposite direction: a higher amplitude BOLD response to rewards at longer delays, which theoretically were more strongly anticipated (Supplementary Fig. 5). Expectancy-driven amplification of brain responses has been seen before32, including in visual cortex33; these effects bear a familiar resemblance to the facilitatory effects of spatial attention34.

Numerous brain areas responded differentially to reward and quit key presses, including some that exhibited a ramp-up in activity before quit responses. We used a GLM to estimate subject-wise peri-event time courses for the two event types separately (using all key presses across all four runs), and submitted the difference between reward-related and quit-related time courses to a group-level ANOVA. Effects occurred diffusely across dorsomedial frontal cortex (DMFC), lateral PFC, anterior insula, precentral sulcus, and occipital and posterior parietal cortex (Fig. 6). In DMFC, anterior insula, posterior parietal cortex and anterior PFC, the difference consisted of an earlier elevated response for quit responses than reward responses. Other regions, including occipital cortex and left inferior frontal gyrus (IFG), responded more strongly to rewards. Broadly, these effects reflect...
that rewards involved a visual cue, whereas quitting was freely timed and volitional.

To test directly for signal changes that preceded decisions to quit, we performed a group-level ANOVA on only the first five points in the quit-related time course (−12.5 to −2.5 s). A significant effect of time point in this anticipatory interval was observed in posterior parietal cortex, DMFC, anterior insula and anterior PFC (Fig. 6 and Table 1). VMPFC showed no effects in either of the above analyses; that is, there was no evidence that subjective value effects in VMPFC could be alternatively explained in terms of a role in response preparation.

Somatic arousal

To test whether subjective value effects in BOLD activity were accompanied by changes in general physiological arousal, we performed exploratory analyses of heart rate (inter-beat interval measured via pulse oximetry, n = 17) as a function of task events. Heart rate transiently accelerated after key presses, but did not differ between the two conditions as a function of delay time (Fig. 7a). In the HP condition, there was greater transient cardiac acceleration for rewards preceded by longer delays (Fig. 7b), bolstering our conclusion—also supported by RTs and occipitoparietal BOLD effects—that subjective reward expectancy increased with elapsed time in the HP condition. Comparing heart-rate time courses for reward and quit events revealed cardiac deceleration, a well-known correlate of motor preparation35, before quit responses (Fig. 7c). In summary, pre- and post-key press brain responses (Fig. 6a–f) co-occurred with changes in general somatic arousal, but there was no evidence that arousal effects (indexed by heart rate) accompanied the trial onset–locked BOLD effects of theoretical subjective value (Figs. 4 and 5).

DISCUSSION

Decision makers faced with uncertain delay should reappraise awaited rewards as time passes. Depending on the statistics of the environment, the passage of time may either decrease or increase one's estimate of how long a delay remains. This type of dynamic reassessment offers a rationale for sustaining or curtailling persistence.

We elicited either greater or lesser willingness to persist in laboratory environments by manipulating the timing statistics that governed reward delivery. Decision makers calibrated their level of persistence adaptively: this extends previous demonstrations of environment-specific calibration of intertemporal choice behavior13,36. Convergent RT, BOLD and heart-rate data suggest that participants encode the relevant timing statistics, responding more vigorously to more strongly expected rewards38. Behavior still fell short of optimality, and an important goal for future work is to determine whether this was a result of inexact statistical learning, strong prior expectations, stochastic noise, unmodeled sources of value (for example, anticipation37) or other causes. Future work should also test whether
performance would differ if immediate or viscerally tempting rewards were at stake (for example, appetizing foods instead of money)\(^3\).

The success of the behavioral manipulation enabled us to examine time-dependent brain signals associated with either high or limited behavioral persistence. We observed signals in VMPFC consistent with a dynamic and context-sensitive reassessment of the awaited outcome’s subjective value. This effect was identified using both model-guided and exploratory fMRI time course analyses, both at the whole-brain level and in ROIs previously implicated in subjective evaluation.

### VMPFC and persistence

Persistence toward future rewards has been classically understood to depend on self-regulatory psychological processes that compete with and override more impulsive, reward-sensitive processes\(^1,2,11\). Dual-system psychological models have given rise to the neuroscientific hypothesis that competitive dynamics exist between brain regions subserving cognitive control and reward processing\(^21-23\).

In contrast with this standard view, we have proposed that delay-of-gratification decisions depend on a dynamic reappraisal of the awaited future reward\(^12,13\). This account attributes differences in waiting behavior across individuals and situations to factors such as temporal beliefs, perceived outcome values and the perceived cost of time, not merely to differences in the capacity to exert self-control\(^15,16\). Here we elicited differences in waiting by manipulating temporal beliefs and obtained evidence for a time-varying representation of subjective value. The hypothesized signal is context dependent, evolving over time in a manner that depends on the timing statistics of the current environment. A corresponding BOLD trajectory was identified in VMPFC, a cortical region regarded as part of a final common pathway in the prospective evaluation of choice alternatives\(^6\). These results are consistent with the view that persistence depends on the same neural and cognitive processes that guide other forms of reward evaluation and economic choice. This view implies that adaptive persistence depends on accurately representing the value of waiting, and need not depend on the engagement of effortful inhibitory control processes\(^3,8\). Our results add to the large body of evidence that VMPFC valuation processes utilize a detailed representation of higher order task structure\(^18,19\). Our findings also extend current conceptions of VMPFC function; although VMPFC activity is known to encode phasic subjective value during discrete choices\(^3-8\), we found that it also tracked subjective value in a temporally extended manner (see ref. 39 for a related finding).

Our neuroimaging results suggest that there is no need to posit antagonistic dynamics between neural reward systems and control systems to explain voluntary persistence (though we cannot, of course, rule out such dynamics in other situations). Our analyses could have detected patterns suggestive of dual-system competition. For example, the analyses in Figures 4 and 5 and Supplementary Figure 4 could have detected activity scaling with the difficulty of persistence, but no such effects were found. The analyses in Figure 6 could have detected a lapse in control-related activity before decisions to quit, but instead the opposite occurred: an ensemble of regions previously implicated in cognitive control, lateral PFC, DMFC, insula and parietal cortex, increased activity before quits, consistent with brain responses found to precede shifts of strategy in other task paradigms\(^26,40\).

Our findings are more compatible with the hypothesis that cognitive control operates via value modulation\(^9,17,20\). The value modulation hypothesis stipulates that control mechanisms in lateral PFC operate by modulating subjective value representations in VMPFC. The hypothesis therefore posits a VMPFC signal that incorporates all relevant information and suffices as a final common pathway to guide decisions, consistent with our findings. It also posits that this signal depends on lateral PFC inputs. On this point our data are mostly silent. We found no evidence for condition-dependent activation trajectories in lateral PFC; nevertheless, we assume that value computation involves interactions among multiple brain regions, and we cannot exclude the possibility that lateral PFC participates.

### Value representation during foraging

The problem of calibrating persistence in our willingness-to-wait task is closely analogous to the patch departure problem in foraging\(^24-26\). It has recently been hypothesized that foraging, which typically involves a succession of accept-or-reject decisions, imposes fundamentally different information-processing demands from standard multi-alternative economic choice\(^41\). Recent work has implicated dACC in signaling the value of exiting foraging patches\(^26\) or of shifting away from default alternatives\(^41\), although other findings have questioned this idea\(^42\).

We did not find evidence for continuous, prospective encoding of the value of quitting (analogous to patch departure) in dACC. Such a signal would theoretically have followed an inverted version of the value of waiting (Fig. 3c,d), and could have been detected in either our model-based analysis (as a negative effect) or our exploratory time course analyses. We did, however, observe a response in dACC and other frontal and parietal regions in anticipation of quit decisions. This pattern is consistent with general motor preparation as well as with the possibility that decision-related signals in dACC manifest predominantly during overt choice execution\(^26,43\).

Our results point to a role for VMPFC valuation signals even in a foraging-like situation in which decision makers encounter one opportunity at a time and seek to maximize their overall rate of return. VMPFC activity correlated with the value of the current opportunity (waiting for the current token). This finding agrees with the idea that VMPFC encodes a ‘best minus next best’ comparative value signal\(^44\) even when the next best is the constant background option of moving on to a new opportunity. This parallels previous demonstrations that VMPFC reflects the subjective value of individual options that are evaluated in turn against a fixed reference alternative\(^7\). Our findings suggest continuity between the valuation mechanisms involved in temporally extended foraging scenarios and multi-option economic choice.

### Reward prediction error

The willingness-to-wait task theoretically involves both positive and negative RPE. Positive RPE should accompany reward delivery, as, given temporal uncertainty, rewards are not fully predicted at the specific time they are delivered\(^31\). Conversely, the pre-reward interval (when the reward could have occurred, but did not) presumably involves negative RPE\(^44,45\). Long delays in the HP environment highlight the dissociability of value and RPE signals. Reward expectancy ramped up over time (Fig. 2c), so non-reward should be associated with progressively larger negative RPE, even as the awaited reward’s subjective value steadily increases (Fig. 3a). Even though decision makers may be increasingly surprised that the reward did not come now, they are also increasingly confident that it will arrive soon. One potential explanation for the lack of clear RPE signals in our neuro-imaging data might be that, at least at the resolution of fMRI, RPE and subjective value signals were superimposed.

Subjective value is canonically associated with BOLD activity in VMPFC, PCC and striatum\(^3-5\), and a broad standing question is how these regions might differ in their computational contributions to decision behavior. One possibility is that striatum preferentially encodes RPE\(^46,47\), whereas VMPFC preferentially encodes prospective...
decision values.

We find that the individual components of subjective value (Supplementary Fig. 2), to isolate variation from related factors such as moment-by-moment reward probability, and to test the generality of these effects across different magnitudes and types of rewards. Research on these topics will yield an enriched picture of how the brain’s valuation mechanisms contend with the complexity of real-world decision environments.

METHODS

Methods and any associated references are available in the online version of the paper.

ACKNOWLEDGMENTS

This research was supported by NIH grants DA038070 to J.T.M. and DA029149 to J.W.K.

AUTHOR CONTRIBUTIONS

J.T.M. and J.W.K. designed the experiment, developed the analysis procedures, and wrote the paper. J.T.M. collected and analyzed the data and developed the theoretical model.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
ONLINE METHODS

Participants. The participants were 20 members of the University of Pennsylvania community (age 18–30, mean = 22, 11 female). Two additional participants were excluded for head movement (shifts of at least 0.5 mm between >5% of adjacent time points). Participants were paid a show-up fee ($15 per h) plus rewards earned in the task (median = $19.80). All participants provided informed consent. The procedures were approved by the University of Pennsylvania Internal Review Board. No statistical methods were used to predetermine sample size, but our sample size was similar to those reported in previous publications16,18,19,32,42,47.

Task. The task was programmed using Matlab (The MathWorks) with Psychophysics Toolbox extensions51,52. A circular token, colored green or purple, appeared in the center of the screen, labeled ‘0¢’. After a random delay, the token turned blue and its value changed to 30¢. Participants could sell the token anytime by pressing a key with their right hand. The word ‘SOLD’ appeared in red over the token for 1 s. After a 1-s blank screen, a new token appeared. The previous token’s value was added to the participant’s total earnings, which were displayed only at the end of each scanning run. Setting the token’s initial value to 0¢ meant that, unlike earlier work using this procedure13, participants received no immediate reward upon quitting. This served to simplify the task without substantially altering either its incentive structure or the resulting pattern of behavior.

A white progress bar marked the amount of time the current token had been on the screen. The bar’s full length corresponded to 100 s. It grew continuously from the left and reset when a new token appeared. The progress bar was included to reduce interval-timing demands and discourage a strategy of covertly counting time.

The scanning session was divided into four 10-min runs. New tokens were presented until time was up. Each run presented one timing environment (corresponding to one token color). The two environments alternated in successive runs. The order of environments and the mapping of token color to environment were counterbalanced across participants.

Each participant completed a preliminary behavioral training session consisting of four 10-min runs alternating between the HP and LP environments. Participants were explicitly instructed that the green and purple tokens might differ in their timing, but that they had to learn the nature of the differences from direct experience and were free to adopt any behavioral strategy they preferred. During behavioral training (but not during scanning) the screen displayed the time left in the 10-min run and the amount earned so far, to help ensure that participants understood the structure of the task. Each token during behavioral training was worth 10¢. Participants explored the task environments during training, waiting through full 90-s delays in the LP condition on a median of 3.5 trials (IQR = 1–5.5, 30 for 18/20 subjects). Participants completed two additional 5-min runs (one per condition) outside the scanner just before the fMRI session. Waiting behavior over time across training, practice, and fMRI sessions is plotted in Supplementary Figure 1.

Participants would have faced fundamentally the same trial-by-trial decision problem if they had received explicit information about the probabilistic contingencies53 in lieu of experience-based training. However, there is evidence that probabilistic information is encoded differently when learned from description versus direct experience54–56. Our training procedure was designed to involve the type of experience-based, implicit statistical learning that is thought to guide the intended future quitting time. The rate-maximizing strategy was to wait through T equal the expected delay if the reward is received, pT = Pr(reward ≤ T), and let Tp equal the expected delay if the reward is received, T = E(reward | t ≤ T). Each trial’s expected rate of return, in ¢ s⁻¹, is

\[ R_T = \frac{30p_T}{T(1 - p_T) + 2} \]  

The numerator is a trial’s expected gain in cents and the denominator is a trial’s expected cost in seconds, assuming a 30¢ reward and a 2-s inter-trial interval. The goal is to find the value of T that maximizes R_T. We use R^* to denote the best available rate of return. Figure 1c plots R_T as a function of T. The best policy in the HP environment was to wait 40 s (R^* = 1.22¢ s⁻¹), whereas the best policy in the LP environment was to wait 20 s (R^* = 0.82¢ s⁻¹).

Modeling subjective value as a function of elapsed time. At each point in a trial, the token’s subjective value depended on three factors: (1) the expected earnings from that token, (2) the expected additional time to be spent on that token, and (3) the monetary value of time, which corresponds to R^* from above. We denote the expected earnings as g_T(t) and the expected time as h_T(t). Each of these depends jointly on the current elapsed time t and the intended future quitting time T. For given values of t and T, the expected return is

\[ g_T(t) = d_T(t) - (R^*)(h_T(t)) \]  

The current trial’s subjective value (denoted potential in the model’s original formulation23) equals the maximum value of g_T across all possible quitting times

\[ g(t) = \max_T g_T(t) \]

Put differently, g(t) is the expected net return in the remainder of the current trial, accounting for the cost of time, under the best available waiting policy. Its minimum is zero because there is always an option to quit immediately (we treat the ITI as part of the subsequent trial). The decision maker should continue waiting if g(t) > 0.

Figure 3a shows g(t) as a function of t in each environment (see Supplementary Fig. 2 for decomposition of g(t) into its components). The function approaches 30¢ at the last possible reward time, when a 30¢ reward is expected with no further delay. The best strategy is to wait up to 40 s in the HP environment but quit at 20 s in the LP environment. If a decision maker in the LP environment were to have waited 53.5 s already it would be better at that point to continue waiting.
for the reward that was sure to arrive at 90 s. We obtained very similar results if we used each participant's actual environment-specific reward rate in place of the theoretical maximum, $A^\prime$ (Supplementary Fig. 6).

Behavior could be well characterized as a stochastic function of theoretical subjective value. To evaluate this we represented each subject's behavior as a series of pseudo-choices between waiting and quitting, placed every 1 s throughout all delay intervals in the experiment. We then modeled pseudo-choice outcomes ($1 = \text{wait, } 0 = \text{quit}$) as a function of subjective value and a constant intercept in subject-wise logistic regressions. Subject-wise maximum-likelihood coefficients were tested at the group level using a Wilcoxon signed-rank test. We additionally used likelihood-ratio tests at the single-subject level to compare the full model to the (nested) intercept-only model, and tested the resulting z statistics at the group level. Finally, we tested an alternative model that, in place of subjective value, coked the HP and LP conditions categorically. This model had the same number of parameters as the subjective value model and could represent the possibility that participants merely quit more often in the LP than HP environment. Subject-wise differences in model deviance were tested against zero using a group-level Wilcoxon signed-ranks test.

Allowing for endogenous temporal uncertainty did not substantially alter the theoretical results described above. Figure 2c displays hypothetical continuous hazard functions allowing for subjective uncertainty in time-interval perception. For an interval of true duration $t$, subjective uncertainty is typically well characterized by a Gaussian distribution with mean $ \mu = t$ and $ \sigma^2 = t^2 \times CV$, where $CV$ is a fixed coefficient of variation. We modeled temporal uncertainty by convoluting each discrete distribution in Figure 1b to a Gaussian mixture distribution. A Gaussian component was placed at each possible reward time, with $ \mu = t$, $ \sigma = t \times CV$, and weight equal to $Pr(t_{\text{reward}} = t)$. We set $CV = 0.16$ on the basis of human behavioral findings (the median $CV$ from Table 2 of ref. 59 converting Gaussian kernel (FSL's FLIRT)) using boundary-based registration, undistorted and warped to MNI space (see Supplementary Fig. 3c), constituted our subject-by-subject theoretical predictions.

The model-based analysis was performed voxel-wise on all 20 subjects across 2.5–30 s from trial onset. Each subject's empirically estimated difference time course (HP minus LP) was regressed on the theoretical difference time course (Fig. 3d) together with a constant intercept. Using the simplified HRF-convolved theoretical time courses in Figure 3c yielded equivalent results. Time points lacking data in either environment for a given subject were omitted (this resulted in the omission of 3 time points for one subject; see Supplementary Fig. 3). We adopted a two-step approach (first estimating the time courses and then submitting them to the model-based contrast) so that included time points were weighted uniformly. Otherwise, early time points, which were sampled more frequently (Supplementary Fig. 3), would have received greater weight, and the pattern of time point weighting could have differed between environments for individual subjects. Contrast coefficients were tested against zero at the group level in two-tailed voxel-wise $t$ tests.

An additional open-ended analysis tested for condition-by-time point interactions in the trial onset–locked BOLD time courses (using $n = 19$ participants with complete data; Supplementary Fig. 3). The main effect of time point is of limited interest because it captures nonspecific effects of time-from-keypress; similarly, the main effect of condition is uninformative because the two conditions were presented in separate runs with independent baselines. The condition-by-time point interaction tests for a difference in BOLD trajectories between the two environments without constraining the form of the difference. In a repeated-measures framework this is equivalent to testing the main effect of time point on the difference in signal between the two environments. Accordingly, we performed a voxel-wise one-way repeated-measures ANOVA on the difference time courses (HP minus LP) at the group level. An equivalent procedure was used to compare BOLD time courses aligned to reward-related and quit-related key presses.

The RPE analysis was limited to the HP environment, in which the sustained time in reward expectancy supported clear predictions. Within a GLM we estimated FIR coefficients for the peri-reward time course (from 7.5 s before to 10 s after each reward). Eight terms modeled the mean time course, and another eight terms modeled amplitude modulation at each time point as a function of the preceding delay duration. We then computed a contrast of the modulatory effect for three post-reward time points (2.5 to 7.5 s) minus three earlier time points ($−5$ to 0 s). The value of this contrast reflected modulation of the phasic reward response as a function of preceding delay time, over and above any nonspecific effect of elapsed time on the pre-reward baseline.

All whole-brain, group-level analyses assessed statistical significance on the basis of cluster mass, with the cluster-defining threshold set to the nominal $P < 0.01$ level. Corrected $P$ values were determined using permutation testing (FSLs randomise; 5,000 iterations), and results were thresholded at corrected $P < 0.05$. For $F$ tests, each random iteration shuffled time points within subject. For one-sample $t$ tests, each iteration randomly sign-flipped individual subjects' coefficient maps.

Heart rate data acquisition and analysis. Pulse oximetry data were recorded at 50 Hz using the MRI system's built-in oximeter, which also performed automatic
heartbeat detection. Time-stamped data were successfully recorded for 17 of the 20 participants. Heartbeat times were converted to IBI. IBI values farther than 30% above or below the grand median were treated as missing (median = 1.6% of points removed; IQR = 0.5–3.6%). Since IBI varied across individuals (median = 820 ms; IQR = 760–950 ms), IBI values were converted to a percentage of the individual’s grand median. Mean peri-event time courses were calculated on a 0.25-s grid for each subject and event type. For comparisons, time courses for two event types were subtracted to yield single-subject difference time courses, which were then tested at the group level for significant excursions from zero. Entire time courses were tested using cluster-based control for multiple comparisons. Cluster size was defined as the number of adjacent time points with nominal $P < 0.05$ in single-time point Wilcoxon signed-rank tests. A cluster was assigned a corrected $P$ value based on its percentile in the empirical null distribution for cluster size, which was obtained via permutation testing (10,000 iterations with randomized sign-flipping of individual subjects’ difference time courses).

A Supplementary Methods Checklist is available.

51. Brainard, D.H. The psychophysics toolbox. Spat. Vis. 10, 433–436 (1997).
52. Pelli, D.G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. 10, 437–442 (1997).
53. Luhmann, C.C., Chun, M.M., Yi, D.-J., Lee, D. & Wang, X.-J. Neural dissociation of delay and uncertainty in intertemporal choice. J. Neurosci. 28, 14459–14466 (2008).

54. Ungemach, C., Chater, N. & Stewart, N. Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? Psychol. Sci. 20, 473–479 (2009).
55. Fitzgerald, T.H., Seymour, B., Bach, D.R. & Dolan, R.J. Differentiable neural substrates for learned and described value and risk. Curr. Biol. 20, 1823–1829 (2010).
56. Hetwig, R., Barron, G., Weber, E.U. & Erev, I. Decisions from experience and the effect of rare events in risky choice. Psychol. Sci. 15, 534–539 (2004).
57. Kaplan, E.L. & Meier, P. Nonparametric estimation from incomplete observations. J. Am. Stat. Assoc. 53, 457–481 (1958).
58. Gibbon, J. Scalar expectancy theory and Weber's law in animal timing. Psychol. Rev. 84, 279–325 (1977).
59. Rakitin, B.C. et al. Scalar expectancy theory and peak-interval timing in humans. J. Exp. Psychol. Anim. Behav. Process. 24, 15–33 (1998).
60. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. Med. Image Anal. 5, 143–156 (2001).
61. Smith, S.M. et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208–S219 (2004).
62. Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W. & Smith, S.M. FSL. Neuroimage 62, 782–790 (2012).
63. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841 (2002).
64. Cox, R.W. AFNI: What a long strange trip it’s been. Neuroimage 62, 743–747 (2012).
65. Cox, R.W. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173 (1996).
66. Greve, D.N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. Neuroimage 48, 63–72 (2009).
67. Nichols, T.E. & Holmes, A.P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25 (2002).