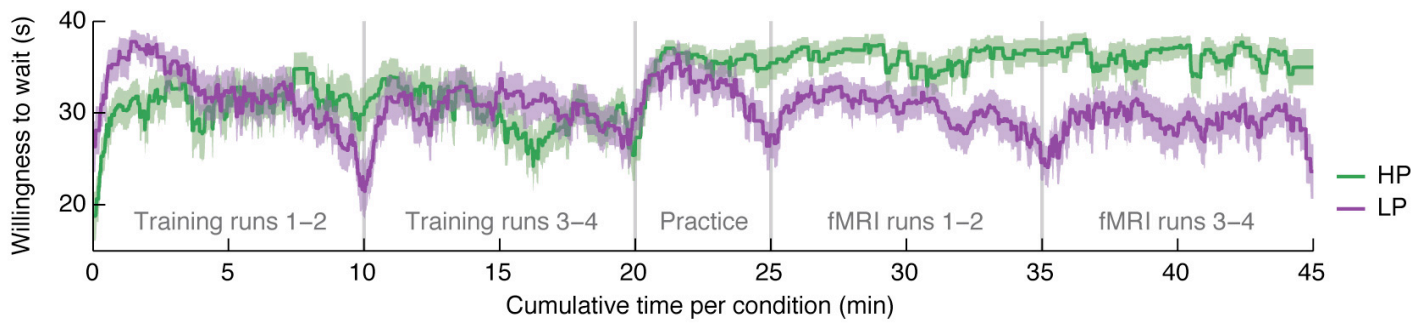


Supplementary Figure 1

Behavior over time across training and fMRI sessions.

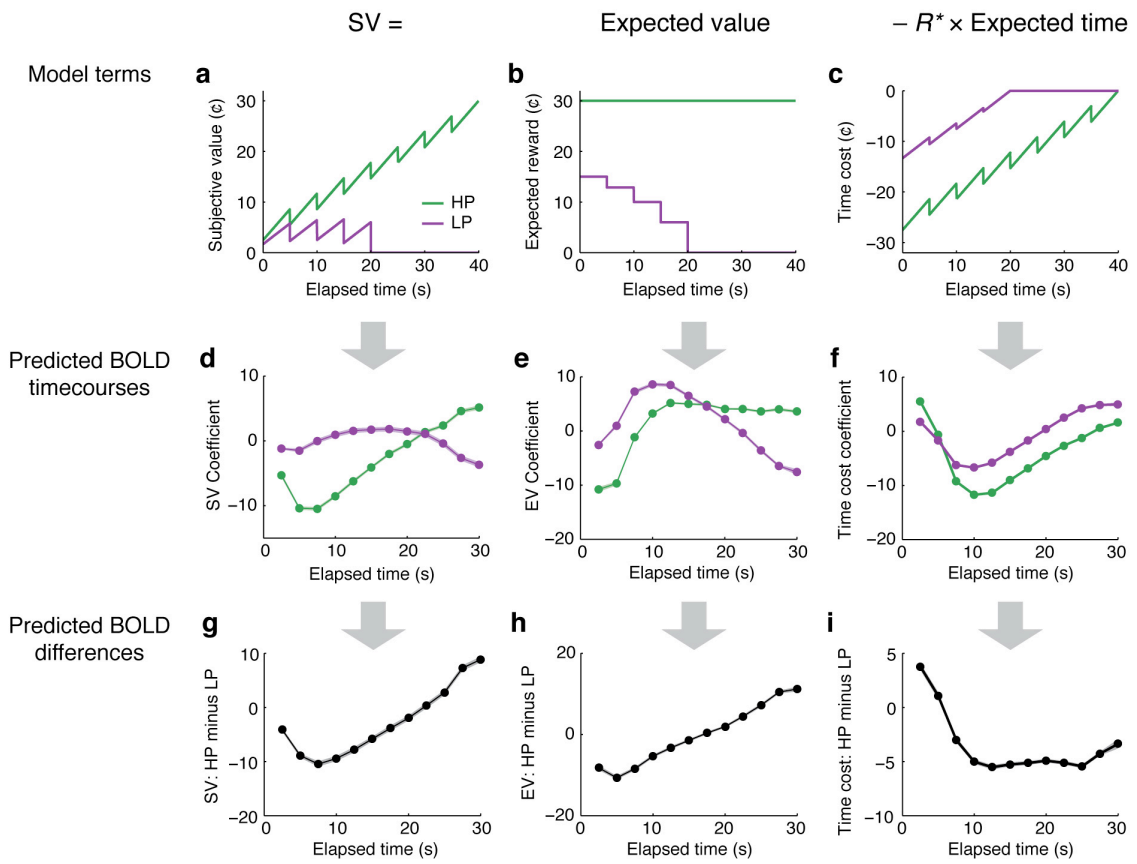
A secondary behavioral analysis assessed the trajectory of persistence behavior over the course of the experiment in each condition. We estimated the timecourse of willingness to wait (WTW) in each condition across preliminary training, pre-scan practice, and fMRI runs (plot shows mean +/- SEM). Within each pair of adjacent runs (e.g., training runs 1–2), participants differed in whether the HP or LP condition was presented first. Cumulative minutes 0–20 are the preliminary training session, minutes 20–45 are from the day of scanning, and minutes 25–45 represent data collected in the scanner (the data used in all other analyses in the paper).

WTW timecourses were estimated using a nonparametric procedure described previously (McGuire & Kable, 2012). WTW at each point in the experiment was estimated as the longest time waited since the last quit trial. The estimate is necessarily only an approximation; we lack full moment-by-moment information about WTW because reward delivery events censor our observation of participants' waiting times. This means there can be a lag before increases in WTW are reflected in the estimated timecourse (in particular, the gradual rise at the beginning of individual runs may be an artifact of the estimator).

Mean behavior was stable during scanning. We estimated subject-wise linear slopes in each condition across the fMRI runs (25–45 min) and neither differed significantly from zero (HP: median slope=0.00, IQR -0.22 to 0.13, signed-rank p=0.970; LP: median slope=-0.07, IQR -0.62 to 0.19, signed-rank p=0.391). We also estimated slopes in each run and condition individually (10 tests); there was a significant negative slope in the first training run of the LP condition (signed-rank p<0.001) but not in the other 9 condition $\times$ runs ($0.06<p<0.97$).

We further confirmed the stability of behavior within the fMRI experiment by calculating AUC for each run individually (cf. Fig. 2b). We observed strong Spearman correlations between an individual's two HP runs ($\rho_{n=20}$=0.81, p<0.001), and LP runs ($\rho_{n=20}$=0.80, p<0.001).

The plot suggests a possible discontinuity between the training session and the fMRI session (at the 20-min point). One possible interpretation is that participants adopted a strategy of exploratory information-gathering during training, when tokens were less valuable, and shifted to a more exploitive strategy when token values increased from 10¢ to 30¢ on the day of the fMRI session. A second possibility is that it took several exposures for participants to learn to disambiguate the two environments.



Supplementary Figure 2

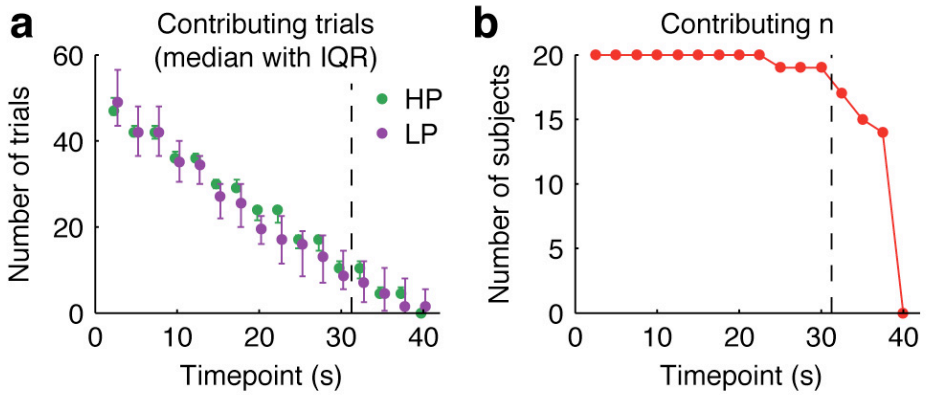
Individual terms in the theoretical model.

Our theoretical model computes an awaited reward's subjective value (SV) as the sum of two terms: expected reward value (EV) and expected time cost (Panels A–C; see *Methods*, Eq. 2). The EV term factors in the subjective probability of obtaining the reward on the current trial (rather than quitting). The time cost term factors in the number of additional seconds the agent expects to spend waiting on the current trial (before either giving up or receiving the reward). Both terms vary as a function of elapsed delay time, and depend on the agent's temporal expectations and intended giving-up time. The values plotted in Panels A–C assume the ideal giving-up time of 20s in the LP condition. This ideal strategy can be identified by maximizing total SV (*Methods,* Eq. 3), but cannot be identified by maximizing the terms in B or C individually. (A strategy of always waiting would maximize EV, whereas a strategy of never waiting would minimize the time cost.)

In our paradigm, the different SV trajectories between the HP and LP environments were mainly driven by the EV term. As time passed in the LP environment, it became more likely that the scheduled delay would exceed the agent's giving-up time and the reward would not be obtained. The dynamic estimate of time cost alone did not follow markedly different trajectories in the two environments.

Accordingly, the SV-related brain responses that we observed in VMPFC could be alternatively described as encoding an EV estimate that evolved dynamically on the basis of temporal expectations. To verify this, we extracted trial-onset-locked timecourses for HRF-convolved EV using the same methods as our main analysis (Panels D–E; see *Methods*; equivalent results for time cost are shown for reference in Panel F). Although the SV and EV timecourses look dissimilar, the predicted *difference* between the HP and LP conditions was nearly identical between SV and EV (Panels G–H; median r^2 = 0.85, IQR 0.83 to 0.88; time cost results shown for reference in Panel I). A whole-brain analysis of EV effects identified a significant VMPFC cluster similar to the SV-related cluster shown in Fig. 4a (257 voxels, corrected p = 0.020).

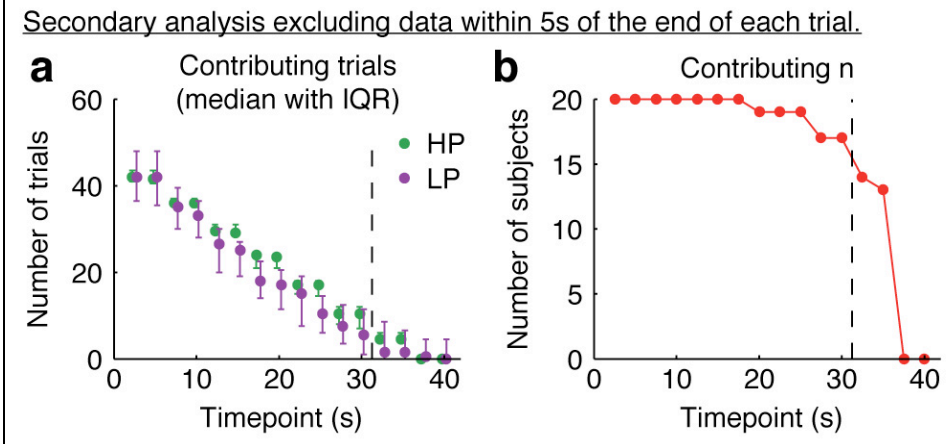
Therefore, although the responsiveness of VMPFC to delay-related costs is well established in other tasks, it remains for future work to establish definitively whether VMPFC encodes time costs *per se* in the willingness-to-wait paradigm. However, an effect either of total SV or of EV alone would support our main conclusion that VMPFC encodes a dynamic and context-dependent value signal in a foraging-like decision context.



Supplementary Figure 3

Amount of data at various delays.

Amount of data available at various lags from trial onset. A: Median number of trials available per subject for each timepoint (with IQR). Dashed line marks the 30s window analyzed. B: Number of subjects with data in both environments for each timepoint.



Supplementary Figure 4

Absence of BOLD effects varying inversely with subjective value.

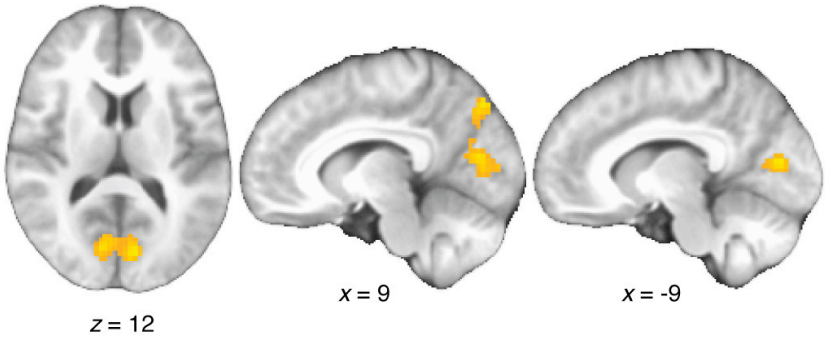
Although our behavioral results demonstrated a direct relationship between persistence and the subjective value of the awaited reward, alternative theoretical frameworks might posit that persistence depends on control processes whose engagement varies *inversely* with value. For example, it might be necessary to engage control processes in order to sustain persistence when an awaited reward's value is in doubt.

Our two-tailed model-based contrast (Fig. 4) could in principle have detected BOLD effects negatively related to subjective value, but no such effects were found. The strongest sub-threshold negative cluster was in left superior occipital cortex (37 voxels; corrected p = 0.226).

It is possible that such an analysis would be more sensitive if it were restricted to timepoints when participants went on to continue waiting. For example, control processes might have become more strongly engaged as the awaited reward's subjective value decreased, *provided* that persistence was indeed sustained further. To investigate this possibility we conducted a secondary analysis in which BOLD timecourses were estimated only from data that preceded the end of a trial by a margin 5s or more (in contrast to the margin of 1s used in our main analyses). For example, the timepoint coefficient at 30s was estimated only from trials that lasted at least 35s. If control processes were engaged to sustain persistence as subjective value decreased, then cognitive-control-related activations might have emerged as negative BOLD effects of subjective value in this analysis.

A disadvantage of this analysis strategy was that it severely curtailed the amount of available data. The number of available trials at later timepoints was reduced by more than 30% for the median participant (compare Panel A to Supplementary Fig. 3a), and some participants were eliminated entirely (compare Panel B to Supplementary Fig. 3b). Accordingly, neither positive nor negative effects were significant when timecourses estimated in this manner were submitted to a two-tailed model-based contrast (cf. Fig. 4). As in our primary analysis, the strongest sub-threshold negative effect was in a left superior occipital cluster (27 voxels, corrected p = 0.29).

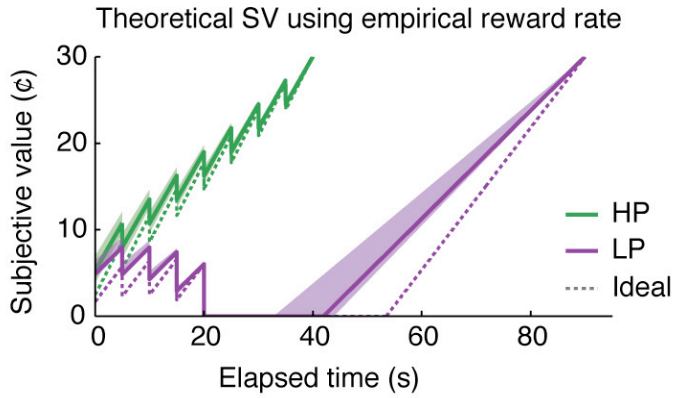
The paucity of data for this analysis was a direct consequence of participants' overall pattern of successful, value-sensitive behavioral calibration. Participants tended not to persist very long in the absence of a valuable future prospect. An internal control process that enforced persistence in such circumstances would not have been beneficial, at least in the present experimental task.



Supplementary Figure 5

Expectancy effects on reward-related brain response.

Occipitoparietal cluster in which the amplitude of the reward-related brain response was positively modulated by the duration of the preceding delay in the HP condition (398 voxels; local peaks in left [-9,-81,12] and right [12,-78,9] calcarine sulci and posterior parietal cortex [12,-78,45]). This region showed a higher-amplitude response to rewards that arrived after longer delays. Rewards at longer delays theoretically involved higher levels of expectancy (Fig. 2c), and were also associated with faster reaction times (Fig. 2d) and larger changes in heart rate (Fig. 7b). The analysis did not find any regions in which reward-related BOLD amplitude decreased as a function of expectancy, a pattern that would be characteristic of a reward prediction error signal.



Supplementary Figure 6

Theoretical subjective value using subject-specific rates of reward.

We recalculated the predictions of our theoretical model using subject-specific empirical estimates of the richness of the environment. Environmental richness is used in our model to define the opportunity cost of time (i.e., the gains one might expect to attain by quitting, akin to abandoning a food patch to forage elsewhere). Because participants' behavior fell short of optimality, it is reasonable to suppose they had a lower-than-optimal estimate of the richness of the environment.

Actual rates of reward were calculated from the fMRI sessions for each subject in each condition. Median reward rate was 1.10¢/s in the HP environment (range 0.80 to 1.20; optimal=1.22) and 0.63¢/s in the LP environment (range 0.40 to 0.73; optimal=0.82). We calculated performance-based theoretical subjective value trajectories for each subject and condition by using these observed reward rates to define the opportunity cost of time.

Plotted is the median performance-based subjective value trajectory in each condition (with IQR; dotted lines represent optimal trajectories from Fig. 3a). Incorporating the lower-than-optimal empirical reward rates tended to increase the subjective value of waiting, especially early in the delay, because delay time was treated as incurring a smaller opportunity cost. Individual subjects' performance-based subjective value trajectories across 0–30s were nonetheless highly correlated with the optimal trajectories (HP: median r^2=1.00, IQR 1.00 to 1.00; LP: median r^2=0.91, IQR 0.84 to 0.92).

The modified procedure for calculating subjective value did not change the results of our fMRI analyses. For each subject we generated synthetic BOLD timecourses encoding performance-based subjective value and passed these through our fMRI analysis to obtain predicted trial-onset-locked BOLD trajectories (analogous to Fig. 3d; see *Methods* for details). The resulting performance-based difference timecourses (HP minus LP) were highly correlated with the original difference timecourses (median r^2=1.00, IQR 0.99 to 1.00), and using this version of the model-derived regressor yielded the same pattern of whole-brain and ROI results described in the main text.

Using performance-based subjective value also did not alter the results for the stochastic behavioral choice model (Fig. 3b). The two variants of the model yielded equivalent fits to the data (difference of model deviances: median=-3.50, IQR -8.22 to 10.72, signed-rank p=0.852).