Frontal–Parietal and Limbic-Striatal Activity Underlies Information Sampling in the Best Choice Problem

Vincent D. Costa and Bruno B. Averbeck

Laboratory of Neuropsychology, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892-4415, USA

Address correspondence to Bruno B. Averbeck, Laboratory of Neuropsychology, NIMH/NIH, Building 49 Room 1B80, 49 Convent Drive MSC 4415, Bethesda, MD 20892-4415, USA. Email: bruno.averbeck@nih.gov

Best choice problems have a long mathematical history, but their neural underpinnings remain unknown. Best choice tasks are optimal stopping problem that require subjects to view a list of options one at a time and decide whether to take or decline each option. The goal is to find a high ranking option in the list, under the restriction that declined options cannot be chosen in the future. Conceptually, the decision to take or decline an option is related to threshold crossing in drift diffusion models, when this process is thought of as a value comparison. We studied this task in healthy volunteers using fMRI, and used a Markov decision process to quantify the value of continuing to search versus committing to the current option. Decisions to take versus decline an option engaged parietal and dorsolateral prefrontal cortices, as well ventral striatum, anterior insula, and anterior cingulate. Therefore, brain regions previously implicated in evidence integration and reward representation encode threshold crossings that trigger decisions to commit to a choice.

Keywords: Bayesian, decision-making, fMRI

Introduction

In many decision problems, one must sample items from a list to find a good option. Often one cannot return to previous items from the list even if they are better than the current offer. For example, one may track interest rates on home mortgages trying to find a low rate. When rates increase, you cannot return to older, lower rates. Stochastic fluctuations in economic variables may drive the rates lower in the future, but they may not. In other cases, previously declined choice options may no longer be available because of changes in choice preferences of another agent, whose offer had been previously declined, as in the fiancée problem. Johannes Kepler famously struggled with this problem when interviewing candidates for his second wife (Koestler 1960; Ferguson 1989). After interviewing the fifth candidate he decided to propose to candidate 4, who turned him down because he had waited too long, or perhaps because he had interviewed candidate 5. While this problem has a long mathematical history in the form of optimal stopping problems (Hill 2009), the neural underpinnings of this choice process have received less attention. Here, we present the results of an fMRI study of this stopping problem. Our task is similar to the secretary problem, which has also been called the marriage problem, the fussy suitor problem, and the best choice problem.

Threshold crossing problems are a special case of optimal stopping problems (DeGroot 1970). In threshold crossing problems, one gathers evidence which supports 2 or more hypotheses. When the evidence in favor of one of the hypotheses crosses a threshold, a decision is executed. This framework has been applied to decision-making tasks, particularly perceptual inference tasks (Gold and Shadlen 2001; Ditterich 2006). Recently, however, this problem has been recast as an optimal stopping problem (Drugowitsch et al. 2012). In optimal stopping problems, at each point in time, one considers the relative values of making a decision versus continuing to sample information. Continuing to gather information is similar to being "below the bound" in a threshold crossing framework and committing to a decision is similar to being "above the bound." However, the values of stopping versus continuing are calculated dynamically after each new piece of information arrives, explicitly considering the number of remaining samples. Therefore, a value which would lead to making a decision at one point in the search process would not necessarily lead to making a decision at another point in the search process. Thus, optimal stopping is a form of value comparison. However, the process differs somewhat from value comparisons in, for example, 2-armed bandit tasks, as outcomes are immediate in these tasks. In stopping problems, when one decides to continue to sample information because the value of sampling exceeds the value of stopping, one is always choosing an expected future reward.

The decision problem in the current study cannot be modeled as a drift diffusion process. Each choice option from the list improves one's estimate of the distribution from which samples are being drawn, and therefore improves one's ability to estimate whether each option is a good candidate. However, declining the current option and sampling again to improve one's estimate of the underlying distribution may mean foregoing the best choice. The trade-off between declining sufficient options to make an informed choice and not missing a good option can be modeled formally using a Markov decision process (MDP).

Materials and Methods

Participants and Task

Thirty-two healthy volunteers were enrolled in the experiment, which was approved by the National Institute of Mental Health Institutional Review Board. All participants were right handed, had normal or corrected-to-normal vision, and were given a physical and neurological examination by a licensed clinician to verify that they were free of psychiatric and neurological disease.

Participants carried out a sequential sampling task, in which they had to try to find a high ranking item from a list of items. Before each list was presented, information was provided to the participants about the nature of the problem to be solved in the list (Fig. 1). For example, participants could be told that they had to buy a used car for $10 000. They would see a list of 8 or 12 cars, one at a time. They were told ahead of time how long the list would be. Each car would be characterized by the number of miles on it. After seeing each item, they would be asked whether they wanted to take that item, or decline that item and see the next option. Once they declined an option they could not

Figure 1. Task events. Participants were given an instruction screen at the beginning of each sequence of options that framed the problem. They were then shown a series of options one at a time. They did not know what the other options in the list were. After each option, they had to decide whether to take or decline the current option. Declined options were shown on the bottom of subsequent screens. If an option was chosen, the participant was given reward feedback.

return to it, and they had to take one of the subsequent options. All declined options were shown on the screen to eliminate working memory load. Thus, each new option could be compared directly to the options which had already been presented. The participants knew that if they took the best option from the entire list they would get $5. If they took the second best option they would get $3 and if they took the third best they would get $1. If the option they chose ranked <3 in the list they received nothing. The ranks for payout were always calculated relative to the entire list of options, including seen and unseen options. Therefore selecting an option that was highly ranked relative to the options that had been seen did not guarantee a reward, as there could be more highly ranked options in the list that had not yet been seen.

The option lists were drawn from 14 categories which include: buying an airplane ticket, a house, a subway ticket, a diamond ring, a digital camera, a used car, bed linens, a used truck, a television, a printer and a motorcycle, or renting an apartment. The other 2 categories were maximizing salary on a job offer, and finding a credit card with a low interest rate. As detailed above, each of these was characterized by a single parameter, and the participant was told how to optimize that parameter (e.g., find a car with the lowest mileage). There were also 3 cost levels for each category so that, for example, subjects could buy a used car for $5000, $10 000, or $20 000. We attempted to make the values in the categories ecologically valid. Thus, the values used for each category were set by examining costs in the region from where subjects were drawn, for each of the items. Behavioral pilot testing suggested that this was relatively successful. Further, the participants performed relatively well in the task, which suggests that the values were reasonable. We ran 7 trials in each of 6 blocks for each participant. Category and cost level per category were chosen randomly without replacement for each participant, for each trial.

Model Fitting

The task was modeled as an infinite state, discrete time, finite-horizon MDP. The MDP framework models the utility, u, of a state,

s, at time n as

$$u_n(s_n) = \max_{a \in A_{s_n}} \left\{ r_n(s_n, a) + \int_S p_n(j|s_n, a) u_{n+1}(j) dj \right\} \quad (1)$$

where A_{S_n} is the set of available actions in state s at time n, and $r_n(S_n, a)$ is the reward that will be obtained in state s at time n if action a is taken. The integral is taken over the set of possible subsequent states, S at time $n+1$ weighted by the transition probability, or the probability of transitioning into each of those states from the current state, s_n if one takes action a, given by $p_n(j|s_n, a)$. The term inside the curly brackets is the action value, $Q(s_n, a) = r_n(s_n, a) + \int_S p_n(j|s_n, a) u_{n+1}(j) dj$. This is the expected reward plotted in Figure 2, for either take or decline actions.

Utility estimates were calculated using backward induction. In the final state one can only take the final option and there is no transition to a subsequent state. Therefore, if we start by defining the utilities of the final states, we can work backwards and define the utilities of all previous states. Specifically, the algorithm proceeds as follows (Puterman 1994), where N is the final state.

1. Set $n = N$

$$u_N(s_N) = r_N(s_N) \text{ for all } s_N \in N.$$

2. Substitute $n − 1$ for n and compute

$$u_n(s_n) = \max_{a \in A_{s_n}} \left\{ r_n(s_n, a) + \int_S p_n(j|s_n, a) u_{n+1}(j) dj \right\}$$

3. If $n = 1$ stop, otherwise return to 2.

To implement the analysis, we modeled the options as samples from a Gaussian distribution with a normal-inverse-χ^2 (N-Inv-χ^2) prior (Gelman et al. 2004). The N-Inv-χ^2 distribution has four parameters, $\mu_0, \kappa_0, \sigma_0^2, v_0$ which specify the prior mean, the degrees of freedom of
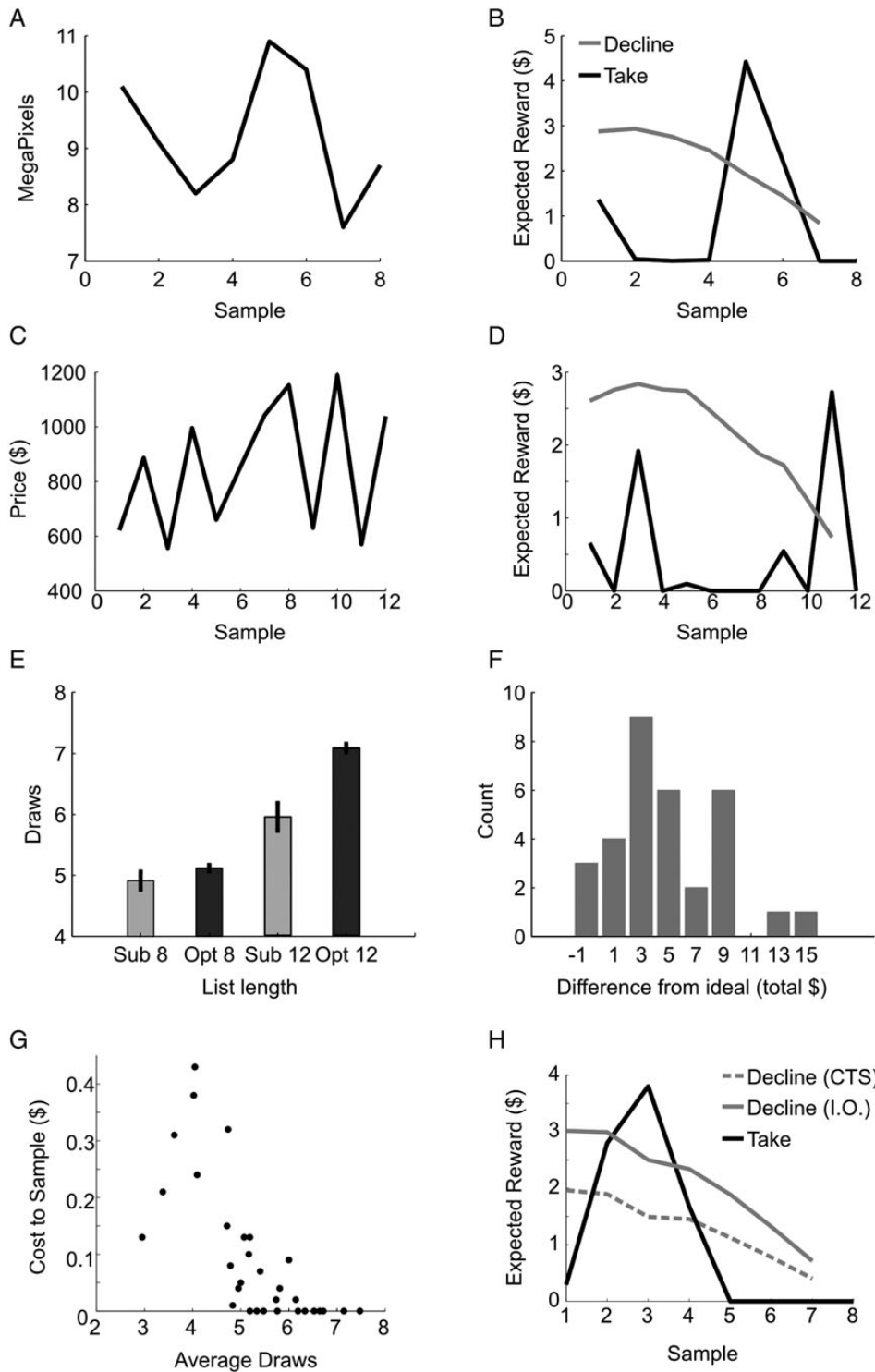
Figure 2. Model output and behavioral results. (*A*) Sequence of options that varied in mega-pixels for the scenario of purchasing a camera. (*B*) Expected reward for either decline or take choices on the current option, for the camera scenario shown at left. (*C*) Sequence of options for the scenario of purchasing an airplane ticket. (*D*) Expected reward for decline and take choices, for the airplane tickets shown at left. (*E*) Participant number of draws (Sub) and optimal number of draws (Opt) at list lengths of 8 and 12. (*F*) Summed difference between actual rewards earned by the MDP algorithm, and rewards earned by individual participants. Negative values indicate subjects that slightly outperformed model. (*G*) Cost-to-sample and average number of draws, for each subject. (*H*) Example sequence for a subject whose cost-to-sample was $0.31. In this case, the ideal observer would not stop at the second sample, but the subject did. CTS, cost-to-sample model; IO, ideal observer model.

the prior mean, the prior variance and the degrees of freedom of the prior variance, respectively. The values of the prior mean and variance (μ_0, σ_0^2) were set to the mean and variance of the complete list of

items. The degrees of freedom were kept small at $\kappa_0 = 2$, $v_0 = 1$. The analysis was not very sensitive to these values over the range 1–4 that we tested. The posterior distribution is then also $N\text{-}Inv\text{-}\chi^2$ with

parameters $\mu_n, \kappa_n, \sigma_n^2, v_n$. These parameters are given by

$$\mu_n = \frac{\kappa_0 \mu_0 + n\bar{x}}{\kappa_0 + n}$$

$$\kappa_n = \kappa_0 + n$$

$$v_n \sigma_n^2 = v_0 \sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{x} - \mu_0)^2$$

$$v_n = v_0 + n$$

where n is the number of data points that have been sampled, $\bar{x}$ is the mean of the sampled data and s^2 is the variance of the sampled data. The sampling distribution, which is the estimate of the distribution from which values will be drawn for the next sample, is given by a t distribution:

$$p(x|D) = t_{v_n}\left(\mu_n, \frac{1 + \kappa_n \sigma_n^2}{\kappa_n}\right).$$

As the state, s_n is the current sample, the sampling distribution is the transition distribution $p_n(j|s_n, a = \text{decline})$, when the subject decides to decline the current option. The variable D indicates the data samples collected up to the current point. The expected reward for decline choices is zero: $r_n(s_n, a = \text{decline}) = 0$. When the subject takes the current option there is no transition and the sequence is finished. We then have to compute the expected reward, $r_n(S_n, a = \text{take})$. To do this, we have to calculate the probability that the current option, s_n will be ranked first, second, or third, among the unseen options. To compute the $p(\text{rank}(x) = i)$ quantities, we have to calculate the probability that we will not sample any options that are better than the current option, that we will sample one that is better, and that we will sample two that are better, from among the unseen samples. The first step is to compute the distribution function for the sampling distribution:

$$F(x) = \int_{-\infty}^{x} p(x|D)dx.$$

If we are currently on sample n and we have N total samples, we have $q = N - n$ remaining draws. We can then calculate the probability that the current sample will be the best, second best, or third best, in the remaining samples (ignoring the declined samples, for the moment). These probabilities are given by

$$p(\text{rank}(x) = 1) = F(x)^q$$

$$p(\text{rank}(x) = 2) = qF(x)^{q-1}(1 - F(x))$$

$$p(\text{rank}(x) = 3) = \left(q * \frac{q-1}{2}\right)F(x)^{q-2}(1 - F(x))^2$$

If the rank of the current sample, relative to the declined samples is b, we can write the expected reward as

$$r_n(s_n, a = \text{take}) = \sum_{i=1}^{3} p(\text{rank} = i) * R(i + (b - 1)).$$

where $R(1) = \$5$, $R(2) = \$3$ and $R(3) = \$1$, and $R(i > 3)) = 0$.

Another modeling approach is possible. In this approach, the best choice problem is solved under the assumption that participants only calculate ranks, and do not have a prior on the sampling distribution. In this case, the probability of sample n having true rank S given that its rank among the already sampled options is s is (Lindley 1961)

$$p(\text{rank} = S) = \binom{S-1}{s-1}\binom{N-S}{n-s}\bigg/\binom{N}{n}$$

In this case, the expected reward is given as

$$r_n(s_n, a = \text{take}) = \sum_{S=1}^{3} p(\text{rank} = S) * R(S)$$

Further, the transition probability, which is the probability that the next sample will have rank s' is $(n+1)^{-1}$. We implemented this model and ran it against ours on simulated data. The models generate highly similar value estimates, although ours slightly outperformed this model. However, our model had accurate prior information about the distribution it was about to sample from, so it would be expected to slightly outperform this rank model, which had no distribution information. All fMRI results were analyzed with respect to the first, Bayesian model, not the second model from Lindley, although results on the fMRI data would be highly similar with either model because value estimates generated by the 2 models were highly similar.

Modeling Individual Subject Behavior

Individual subject behavior often deviates from optimality. To characterize this, we added a parameter to the model to estimate a cost-to-sample for individual subjects. This is implemented by setting

$$r_n(s_n, a = \text{decline}) = C_s$$

where C_s was fit to individual subjects, by minimizing the difference between the sample on which the algorithm stopped, and the sample on which the subject stopped. That is the specific cost function minimized for individual subjects was given by the sum, across lists given to that subject, of the absolute value of the difference, of the sample on which the subject stopped and the sample on which the algorithm stopped. The algorithm stopped on the first sample where the expected value of taking the current option exceeded the expected value of declining the current option.

fMRI Scanning

We carried out whole-brain T_2^*-weighted echo-planar imaging on a 3T GE Signa scanner (GE Medical systems) with an 8-channel head coil. Volumes were collected as 35 slices, 3.5 mm thick, matrix: 64×64 voxels, in-plane resolution 3.75 mm, TR 2 s, TE 30 ms, flip angle 90°. Most participants underwent 6 scanning sessions, except for 2 who underwent 4 and 1 who underwent 5 due to time constraints. The first 5 "dummy scans" of each session were discarded to allow for magnetization equilibration effects.

fMRI Data Preprocessing and Statistical Analysis

Data were preprocessed and analyzed using MATLAB and SPM8 software (Wellcome Trust Centre for Neuroimaging, London; http://www.fil.ion.ucl.ac.uk/spm/). Functional scans were realigned within each session and then across sessions. These were then normalized to the standard Montreal Neurological Institute (MNI) echo-planar image template. All statistical analyses were performed and are reported using the matrix and voxel sizes associated with the MNI template space (MNI matrix size: $79 \times 95 \times 68$ voxels; MNI voxel size: 2 mm). Before statistical analysis, normalized data were smoothed using a 6-mm isotropic FWHM Gaussian kernel.

We employed conventional mass-univariate approaches for estimating the magnitude of the blood oxygenation level–dependent (BOLD) hemodynamic response to each stimulus event (Holmes and Friston 1998). At the individual–participant level, we computed "first-level" mass-univariate time series models for each participant. The aim of the first-level model is to predict the fMRI time series using regressors constructed from box-car functions representing stimulus presentation times, convolved with a canonical BOLD response function, which captures the temporal profile of the BOLD response to a behavioral event. For each regressor and its associated event type, a β-value was computed to estimate the magnitude of the BOLD response evoked by the events. β-Values for different event regressors were then contrasted statistically to test whether the BOLD responses evoked by different events significantly differed in magnitude. The first-level analyses also employed an AR(1) autocorrelation parameter and a high-pass filter of 128 s. The term β-value or β weight comes from the fact that it is a coefficient from a linear regression model.

Regressors in the first level model corresponded to the relevant events in the task and were convolved with box-cars that corresponded

to their durations. Specifically, we included regressors for the instruction screen (6 s), presentation of an option (4 s), participant response (1 s), and feedback (3 s). The participant response was modeled separately for choices to take the current option versus choices to decline the current option, so we could subsequently contrast these two decisions. In addition, both the feedback and the participant response were parametrically modulated. The feedback was modulated by the reward amount (either $0, $1, $2, or $5). The response was modulated by the difference in value of the two choice options (i.e., Q(s, decline option) – Q(s, take option)). Thus, it seems the participants BOLD responses (and perhaps the participant's mental processes) were modulated by the difference in value, at the time that they made their choice, consistent with previous reports (Kolling et al. 2012). Because we included a parametric modulator for the value of the choices at the time of choice, differences in activation to take versus decline choices should be due to factors other than the change in value between these choices, as the parametric modulator controls for this change.

Once β values had been computed for each of our events of interest, as well as contrasts over these β values (see Results), we carried out a "second-level" analysis in which the β weights were entered into 1-sample t-tests, treating participants as a random effect. Between subjects covariate analysis with the Barratt impulsivity scale (BIS) was also carried out at this level. There were, however, no significant effects of the BIS, so it is not further reported. The second level analysis allowed us to test whether β values were statistically consistent across participants. All results reported below were observed at $P < 0.001$ with a minimum of 100 voxels. They were then tested for family-wise error (FWE) correction at the cluster level at P(FWE) < 0.05 using Gaussian random field theory. This is a conventional method (Friston et al. 2006) that uses the estimated smoothness of the data to correct for the massive number of multiple comparisons at all voxels in the whole brain.

In addition to the GLM approach, we also carried out estimation of finite impulse response (FIR) functions. This allowed us to check that the assumption of a canonical hemodynamic response was reasonable. We plot these FIRs for each regressor. However, we do not carry out inference using the FIRs, as this is generally less powerful than using the GLM analysis. The FIR plots shown are averaged within a 10-mm radius sphere centered on the peak activation. Results were similar when the entire cluster was used. We also calculated the FIRs by weighting the contribution of each voxel by its z-value. The results of this analysis were also highly consistent. Therefore, we present results using a simple average of voxels values within the sphere.

Results

Task and Behavior

Participants carried out a sequential sampling task (Fig. 1). In the task, participants were shown options drawn from a predetermined list, one at a time. Lists were either 8 or 12 options long and participants were informed of the list length before they began to sample. After seeing each option, they had to decide whether to take the current option or decline the current option and see the next option. If they declined the current option, it was no longer available and they could not return to that option later. All previously declined options were shown on the bottom of the screen to reduce working memory load. When they took an option, they were rewarded on the basis of its rank in the complete list of options, seen and unseen. If they found the best option they earned $5. If they found the second best they earned $3 and if they found the third best they earned $1. Any other options resulted in no reward. If they went to the end of the list and had declined all options, they were forced to accept the last option, independent of its value.

Value estimates in the task was modeled using a MDP. This model calculates the value of either taking or declining each option in a list as

$$Q_n(s_n, a = \text{take option}) = r_n(s_n, a)$$

$$Q_n(s_n, a = \text{decline option}) = \int_S p_n(j|s_n, a)u_{n+1}(j)dj$$

Thus, estimating the value of taking the current option in the list, $Q_n(S_n, a = \text{take option})$, involves computing the expected reward, $r_n(S_n, a)$, for that option. Estimating the value of declining the current option in the list, $Q_n(S_n, a = \text{decline option})$, involves computing the expected value of continuing to sample. This expected value is a function of the probability that better options remain in the list. The distribution of future samples is given by $p_n(j|S_n, a)$, and the value of each of the possible samples is given by, $u_{n+1}(j)$. Thus, $Q_n(S_n, a = \text{decline option})$ is the expected value of the next state. The value of the next state also depends on the value of the state after that, and so forth. Therefore, this problem is normally solved by going to the last sample and computing its value. This can be done because at the last sample, one has to take the option. One cannot choose to decline the last option. You can then work backward from the last option, to compute the value of taking or declining each of the previous options. Once you arrive at the current option, you have calculated the value of declining the current option. We applied this algorithm to estimate the values of declining or taking the current option. This also allowed us to calculate the trial on which the algorithm would stop sampling and take the current option, and compare this to when the subjects stopped, for each list that the subjects viewed. Finally, we also parameterized the algorithm by introducing a cost-to-sample to improve the fit of the algorithm to actual subject behavior. This effectively reduces the value of future samples, by assuming that subjects will implicitly assess a cost on continuing to sample. The model without a cost-to-sample will be referred to as the ideal observer, whereas the model with a cost-to-sample will be referred to as the parameterized model.

In the interest of generating multiple sequences of trials, and using lists that would be ecologically relevant, we framed the task as a series of purchases and related transactions. For example, in 1 case, participants had to find a camera with the maximum number of mega-pixels at a fixed price (Fig. 2A,B). This might correspond to a scenario where someone has a fixed budget, and they are watching sale prices over a series of weeks to find a good camera at their price point. As they advanced through the list, the algorithm-generated value estimates for each option, where the values calculated by the algorithm were for the reward the participants would receive if they chose 1 of the top 3 options in the list. It can be seen that cameras with relatively low pixel counts were considered low value (Fig. 2B). When an option was encountered that had a high pixel count, at sample 5, it had a high value. Its value was determined relative to the other options that had been seen, as well as the number of remaining options that had not been seen. On sample 5, the value of taking the current option exceeded the value of declining the current option, the threshold was crossed, and the model would take that option. In another example, participants were trying to buy an airline ticket (Fig. 2C,D). Some of the cheaper early options were of relatively low value because it was likely, given the number of remaining samples, that one could find a ticket that was cheaper.

In this particular example, however, the best option is presented relatively early, as sample 3. Although this sample has a relatively high expected reward, it is better, on average, to proceed further through the list as there are 9 samples remaining, and only on option 11 would the algorithm take the current option.

We found that the participants examined more options on lists of length 12 than they did on lists of length 8 ($F_{1,31} = 31.0$, $P < 0.001$). However, participants drew a bit less than optimal (Fig. 2E). When compared to the ideal observer, the participants drew less on the same sequences, across list lengths ($F_{1,31} = 16.96$, $P < 0.001$). There was also an interaction between list length and subject type (i.e., ideal observer vs. participants; $F_{1,31} = 24.55$, $P < 0.001$). Post hoc tests showed that the difference from optimal was significant for list lengths of 12 ($t_{(31)} = 4.98$, $P < 0.001$) but not for list lengths of 8 ($t_{(31)} = 1.4$, $P = 0.169$). Thus, participants adjusted their strategy and sampled more for longer lists, but they did not adjust as much as they should have. This led to most participants underperforming the model (Fig. 2F), although a few participants slightly outperformed the model across the 42 trials. This is possible, as the model is only guaranteed to be optimal on average.

As many of the subjects did not perform optimally, we parameterized the ideal observer model to capture the behavior of individual subjects. Subjects generally did not examine a sufficient number of samples, which implies that they undervalued sampling again. To estimate the amount that each subject undervalued declining the current option and sampling again, we estimated their cost-to-sample (see Materials and methods). We found that the average cost-to-sample across subjects was $0.093 (SEM = 0.022). The cost-to-sample significantly improved the model's prediction, relative to the ideal observer, in 16 of 32 individual subjects ($\chi^2(1) > 6.635$, $P < 0.01$). Further, this parameter captured the sampling behavior of the subjects, as it was significantly correlated (Spearman rho ($n = 32$) = -0.67, $P = 0.001$) with the number of samples the subjects drew, on average (Fig. 2G). In the model, increasing the cost-to-sample decreases the value of future samples, and leads to earlier decision (Fig. 2H). The value estimates used below for the fMRI analysis were derived from the parameterized model.

We also examined 2 additional behavioral models to see if they better fit the subject choice data. In the first model the variance of the inferred Gaussian distribution was fixed to the mean. This is the distribution that the subjects assume the data is being drawn from. This was done for the prior and for all subsequent variance estimates, after updating with samples. It is possible that subjects were making this simplifying assumption, and therefore, such a model might better account for the choice data. However, when we fixed the variance of the distributions to the mean without a cost to sample parameter, and compared the models, the model with the variance estimated from the data outperformed (i.e., provided a better fit to the subject choice data) the model with the variance set to the mean ($t_{(31)} = 4.94$, $P < 0.001$). When we introduced a cost to sample parameter in both models and compared the fit of the models to the subject choice data, the model with variance estimated from the data again outperformed the model with the variance set to the mean ($t_{(31)} = 7.95$, $P < 0.001$).

We also fit a model which used actual category values, instead of the payout from the ranks, to predict subject choices. In other words, this model assumed that the subjects were simply trying to maximize or minimize the actual

parameter (e.g., mileage on a car or the interest rate paid on a credit card). When we compared the models without a cost-to-sample parameter, there was no statistically significant difference between the model that used actual category values, instead of the payout from the ranks ($t_{(31)} = 0.54$, $P = 0.596$). We also introduced a cost-to-sample parameter and compared the optimized cost-to-sample model based on category values to the original model which used sample ranks and found that there was again no difference in the models ($t_{(31)} = 1.61$, $P = 0.117$).

Functional Imaging Results

Participants were given monetary compensation for selecting 1 of the top 3 options in the list. To examine whether these monetary outcomes were engaging neural circuits related to reward motivation, we first examined effects of the outcome (Table 1). We found, consistent with previous studies, that lateral orbital prefrontal cortex (Fig. 3A) and the bilateral ventral striatum (Fig. 3B) were modulated by the size of the reward delivered for a high ranking choice option.

We next examined the contrast between choices to take the current option versus choices to decline the current option, which is similar to a threshold crossing. This contrast revealed a robustly activated set of areas (Table 1). Specifically, we found an anterior insula area bilaterally (Fig. 4A). There was also a large activation in the dorsal anterior cingulate (Fig. 4B), and the ventral striatum (Fig. 4C). In addition, parietal–frontal areas were also activated (Fig. 4D,E). Specifically, we found a left lateralized dorsal parietal activation (Fig. 4D) as well as a left lateralized prefrontal activation (Fig. 4E). The opposite contrast, declining versus taking the current option, did not

Table 1

Peak activation coordinates and statistics for reported contrasts

Region	x[a]	y	z	t[b]	μL[c]
Value of outcome[d]					
Inferior frontal gyrus	−46	42	−16	5.62	1384
	50	40	−2	6.54	4464
Ventral striatum	−12	8	6	6.24	1648
	14	10	−2	6.08	4936
Middle frontal gyrus	48	34	20	5.52	1616
Precentral gyrus	−36	−14	68	6.70	2304
Posterior insula	−46	−24	20	5.38	840
Parahippocampal gyrus	22	−20	−20	5.87	1496
Calcarine gyrus	18	−10	16	8.16	53 204
Take > decline[e]					
Ventral striatum/medial thalamus	−2	−14	6	9.20	8104
Anterior insula	−34	22	−12	9.67	18 416
	32	22	−12	8.65	5336
Anterior cingulate	−2	40	36	8.51	14 104
Parietal cortex	−50	−46	56	7.46	7160
Inferior frontal gyrus	−36	14	32	6.79	2304
Posterior cingulate	0	−26	28	9.10	4520
Calcarine gyrus	−32	−82	−14	8.68	8736
	24	−78	−12	7.83	19 512
Value difference on declined choices[f]					
Ventromedial prefrontal cortex	−4	54	−10	5.29	2412
Precuneus/calcarine gyrus	10	−58	22	5.28	5052

[a]Peak coordinates are reported in MNI space.
[b]Whole brain corrected at $P < 0.05$ FWE (initial threshold $P < 0.001$ uncorrected).
[c]Cluster volume estimated using 2 mm^3 voxels.
[d]Clusters significant for the parametric modulator on reward outcome.
[e]Clusters significant for the contrast of taking versus declining the current option.
[f]Clusters significant for parametric modulation of responses, at the time of choice, for the decline choices. The parametric modulator was the value difference between declining and taking the current option.
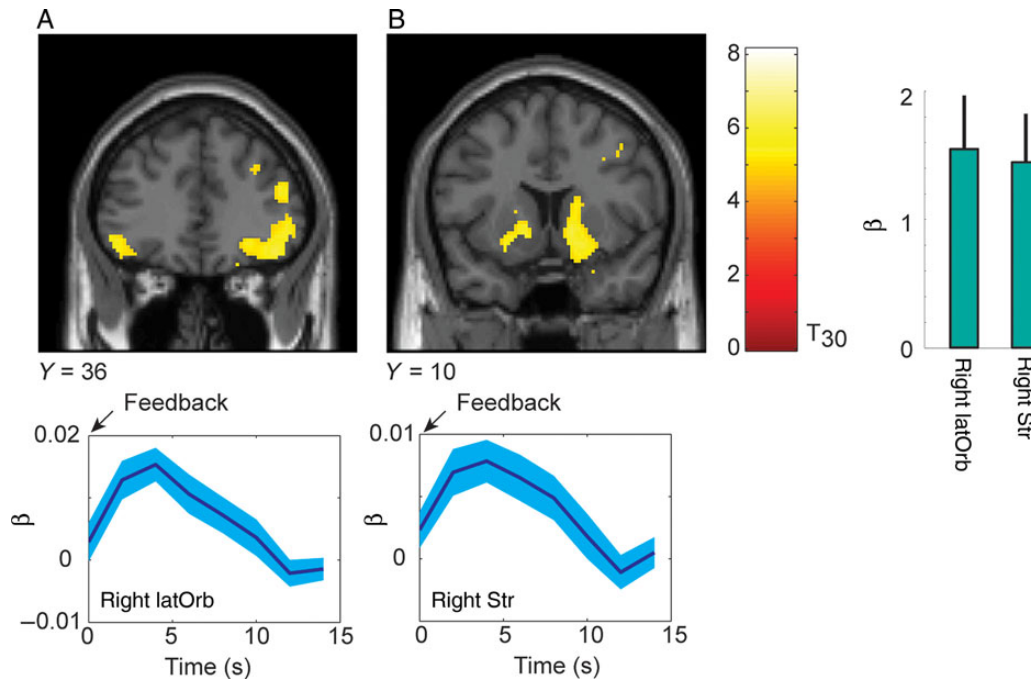
Figure 3. Parametric modulation for reward outcome. Activations plotted at $P < 0.05$ FWE. Time series below each coronal section show mean and standard error from analysis of finite impulse response (FIR), for comparison. The FIR analysis was not used for inference of active clusters. The time course of a parametric modulator, as plotted here, indicates the time-course of significant covariation between the value of the modulator, and the value of the bold signal, over and above the actual event-related response. Histogram shows β weights and 90% CI from the GLM analysis. (A) Activation in lateral-orbital cortex that correlated with the size of the outcome. (B) Activation in ventral striatum.

produce any significant clusters of voxels following whole-brain correction.

In the next analysis, we examined whether activity at the time of choice was parametrically modulated by the value of the choice (Table 1). The parameterized MDP model generated value estimates for every decision (e.g., Fig. 2H). We used these value estimates to generate a parametric modulator of the difference in value of the 2 choice options (i.e., value of declining minus value of taking the current option). We found that this parametric modulator was significant in visual cortex as well as the ventromedial prefrontal cortex (vmPFC; Fig. 5). Thus, these regions both showed larger responses when there was a larger difference in value between the choice to decline and the choice to take the current option, and the response became smaller when the expected value of the take option increased relative to the decline option. There were no significant effects of the parametric modulator for take choices.

We also examined several other variables, to see if they parametrically modulated activation at time of choice. Specifically, we examined reaction time, the number of samples that had been viewed, and the cumulative estimate of the category mean. We found that reaction time did not have any voxels significant at the $P < 0.001$ uncorrected level. However, at $P < 0.005$, there was a cluster in left vmPFC (-14, 54, 2; 89 voxels). It did not survive whole-brain correction, however. For the number of samples viewed, there were 77 suprathreshold voxels at $P < 0.001$ uncorrected. However, none passed the whole-brain corrected threshold. Finally, for the cumulative mean, there were only a few suprathreshold voxels at $P < 0.005$ and none at $P < 0.001$ and none passed whole-brain correction. Thus, none of these other variables accounted for variance significantly at the whole-brain level.

We also used the value estimates generated from the model which used actual category values, instead of rank values to generate the difference in value regressor, and used this as a covariate. Although this model predicted behavior as well as the model based on the payouts calculated from the ranks, it generates value estimates whose scale differs across categories (i.e., values for airline tickets are much higher than values for interest on credit cards). When we used this as a parametric regressor, we did not find any significant voxels at the $P < 0.001$ uncorrected level.

The activation in the present study for the contrast of take versus decline, was highly similar to the activation in Furl and Averbeck (2011) for a related information sampling task. In both tasks, this contrast compares the choice to stop sampling information to the choice to continue sampling information. Therefore, consistent activation across studies would support the hypothesis that this set of areas is important for deciding when the value of the current option exceeds the expected value of future options. We masked the take versus decline contrast in the current study, with the related contrast from the previous study (Fig. 6). Note that we were not asking whether 2 variables activate the same areas within subjects from a single study. Rather, we were asking whether contrasts in different groups of subjects from tasks that have overlapping computational demands activate similar areas. We found that several areas were activated across both studies including the anterior insula (Fig. 6A), the anterior cingulate (Fig. 6B), dorsal parietal cortex (Fig. 6C), and the ventral striatum (Fig. 6D). The lateral prefrontal activation (Fig. 4E) in the current study was not present in the previous study, and therefore, it did not survive the mask.

Discussion

We carried out an fMRI study of participants doing an information sampling task known as the best choice task.
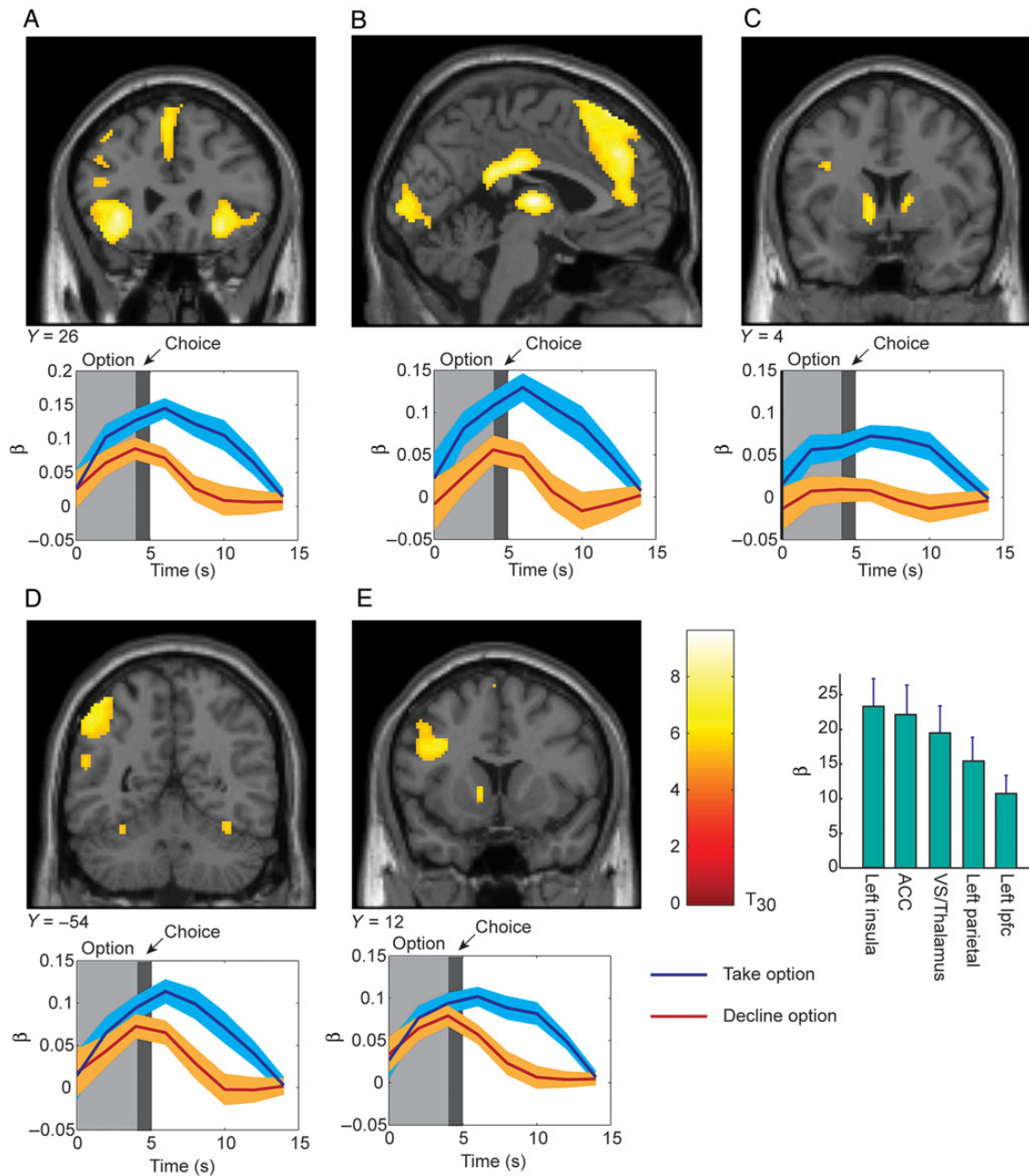
Figure 4. Significant clusters for contrast of Take option versus Decline option. Coronal (and one sagittal) sections with significant voxels are shown on top. The time-course of activations from a FIR analysis is shown below each corresponding coronal section. Time zero is option onset, while choice indicates the time at which the subject entered their decision. Activations plotted at $P < 0.05$ FWE. Histogram shows β weights and 90% confidence interval from GLM analysis for each of the activations shown in (A–E). (A) Activation in left anterior insula. (B) Activation in anterior cingulate shown on a sagittal view. (C) Activation in ventral striatum. (D) Activation in dorsal parietal cortex. (E) Activation in lateral–prefrontal cortex.

Participants did well at the task, but on longer lists they tended to under-sample. While participants were declining options, the value of the declined options, relative to future options, correlated with activation in ventromedial prefrontal cortex (vmPFC). This finding is consistent with previous studies of value representations in vmPFC (Lim et al. 2011; Kolling et al. 2012). When participants committed to a decision relative to continuing to sample, effectively crossing a threshold, frontal–parietal and cingulate, anterior insula, and ventral striatal areas were activated.

The best choice problem we have studied is an optimal stopping problem (DeGroot 1970). Other recent studies have also

examined stopping problems (Gluth et al. 2012; Kolling et al. 2012), although they mostly focused on value representations rather than the decision to stop. Reaction time variants of perceptual inference tasks studied in monkey (Roitman and Shadlen 2002; Seo et al. 2012), are also stopping problems (Drugowitsch et al. 2012). Perceptual inference tasks are often modeled using drift diffusion or sequential sampling models, in which one assumes that evidence is accumulated until a bound is reached (Ratcliff 1978). Such an approach is not readily applicable to the task used in the present study. In the best choice task we have studied, one assumes there is some distribution of values in the world associated with each
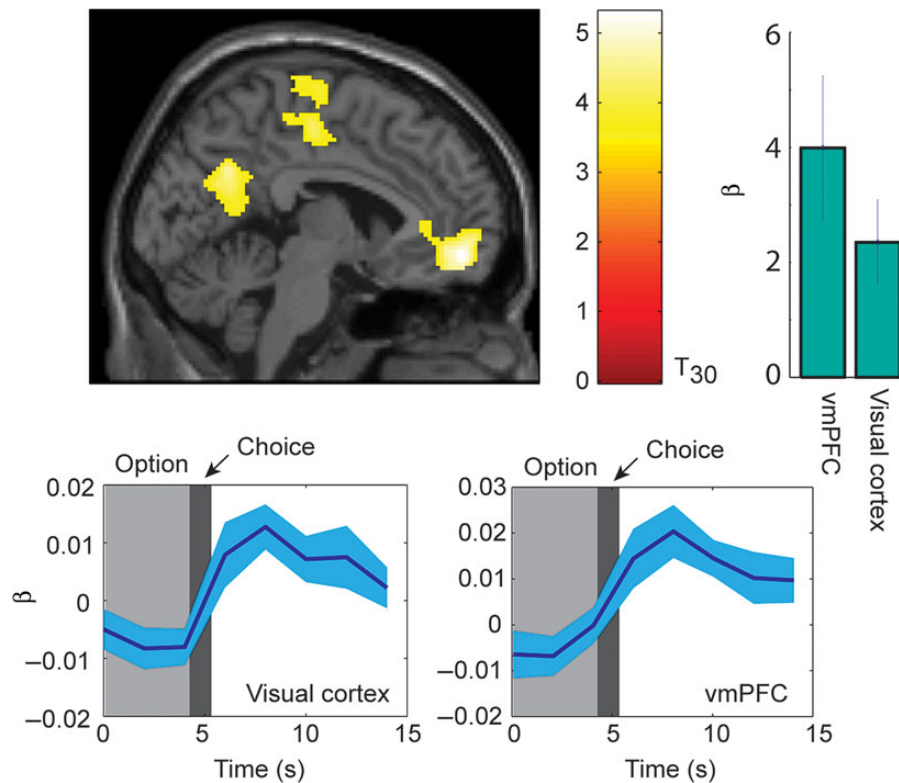
Figure 5. Significant effects of parametric modulator for declined choices. The time-course for a parametric modulator shows significant covariation between BOLD response and value of parametric modulator, over time. Activations plotted at $P < 0.001$ uncorrected. Histogram shows β weights from the GLM analysis that corresponds to activations. Significant activations can be seen in visual cortex and vmPFC. Lower panels show time series from FIR analysis in visual cortex and vmPFC, as indicated.

category. Each subsequent sample improves the estimate of this distribution. This corresponds most closely to inferences about the state of the world in perceptual inference. However, although collecting additional samples can lead to a better estimate of the underlying distribution, this will not necessarily lead to the selection of a better option, because one cannot return to previously presented options, once they are declined. Thus, each option serves not only to update level of belief about the state of the world. Each option is also a unique choice candidate. This scenario corresponds well to a large class of real-world decision problems from finding a mate (the fiancé(e) problem) to hiring a new employee (the secretary problem) as well as tasks like buying airline tickets or locking in on a mortgage rate, when prices fluctuate. In many of these scenarios, once one has declined an option, one cannot return to it.

While there is a substantial body of work on information integration, or the representation of belief, in perceptual decision-making tasks using both animal (Shadlen and Newsome 1996; Roitman and Shadlen 2002) and human subjects (Heekeren et al. 2006, 2008; Ho et al. 2009), there is relatively little known about how the decision to stop gathering information is arrived at. Computational models have hypothesized that the superior colliculus might be relevant (Lo and Wang 2006). The colliculus might be relevant for eye movements, particularly in fast, millisecond scale decisions in saccadic reaction time tasks. However, a general role in committing to a choice option when deliberation extends over seconds, independent of the effector, seems less likely. We have found that the decision to commit to an option robustly engaged frontal–parietal areas as well as cingulate, anterior insula and ventral striatal areas. Thus, this set of areas is a candidate network for the decision to commit to an option and stop gathering information. One possible alternative explanation is that this network is simply registering the relative value of the chosen option, and that this value is greater than the value of all of the declined options. However, the effect of the relative value on the BOLD response was estimated with the parametric modulator. Thus, the direct contrast of taken versus declined options should not have been sensitive to the change in relative value, as variance related to this should have been taken up by the parametric modulator.

As further evidence that this set of areas underlies the decision to commit, it was consistent with the set of areas found in our previous study of the beads task for a similar contrast (Fig. 6). However, the current study extends the finding in the beads task in 2 important ways. First, in the previous study, the participants had to remember the beads drawn, so there was a working memory component to the task. Thus, activation of parietal and prefrontal areas may have been related to this aspect of the task. In the current task, previous samples were displayed on the screen, so subjects did not have to use working memory to retain the distribution of samples. Second, the decision process in the current task cannot be modeled within a drift diffusion framework. However, the beads task can be modeled with this approach (Averbeck et al. 2011). Thus, the finding in the current study and our previous study suggest that these areas underlie the decision to commit to an option, independent of the mechanism that underlies belief updating.
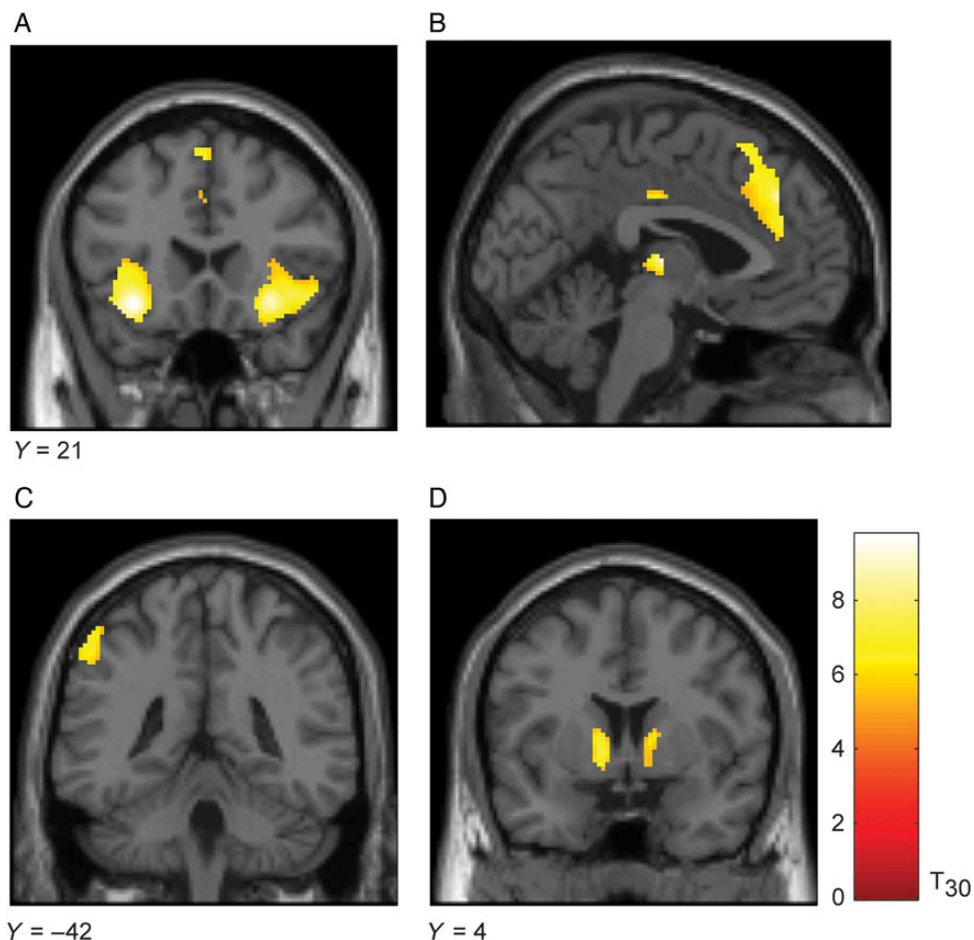
Figure 6. Take versus decline activation, masked by significant clusters (*P* < 0.05 FWE) from Furl and Averbeck (2011). Compare with Figure 4 from this earlier paper in which information sampling was assessed in the beads task. (*A*) Anterior insula activation; (*B*) ACC activation; (*C*) dorsal parietal activation; (*D*) ventral striatal activation.

While we have identified a set of areas that seem to underlie this process, the individual contributions of the areas, if in fact, there are differential contributions, are not yet clear. However, it is possible that the parietal–frontal areas underlie belief representations that are relatively independent of value. In our previous study, we found that parietal–frontal areas correlated, between subjects, with the amount of information they gathered before committing to a decision (Furl and Averbeck 2011). Thus, these areas may mediate probabilistic belief estimates, although reward magnitude is also likely represented to some extent (Platt and Glimcher 1999; Kim et al. 2009; Seo et al. 2009). The striatum and the anterior insula, on the other hand, may relate more to the reward magnitude of the choice option, somewhat independent of underlying belief estimates (Haber and Knutson 2010). In support of this, we found that the ventral striatum was parametrically modulated by the size of the reward outcome.

The role of the anterior cingulate in our task is perhaps less clear, given this area's involvement in a vast array of mental processes. Relevant to this question, however, other recent work has examined tasks in which subjects could either "forage" or take a currently offered gamble (Kolling et al. 2012). This task is similar to ours in some respects, although it differs in important aspects. Specifically, their task contrasts foraging choices in which subjects decide to (re-)

sample from a distribution of possible gambles rather than engaging the currently offered gamble. The difference between engaging the current gamble and searching for a better gamble, which is similar to our contrast of taking versus declining a current option, is not reported, although this contrast was a focus of our study. One of the main contrasts of interest in their study compared the foraging period to the decision period. This would be the closest contrast to our comparison of declining versus taking a current option, although it may not be directly related. However, they report ACC activation for this contrast and we find no activation for our contrast. Thus, our task appears to engage different neural mechanisms. In addition to different patterns of brain activation, there are substantive behavioral differences between these 2 tasks. In the foraging task, the participants know the distribution that they are sampling from because it is presented on the screen. Thus, they can set an internal threshold and sample until they get one of the option pairs that they prefer. The behavioral analysis showed that they took into account the cost of searching, but it was not clear how they incorporated this information, as no formal model was given against which the participant performance could be compared. In addition to these differences the foraging task is also an infinite horizon problem, as there is no limit to how long they can sample.

Conclusion

We have identified a set of brain areas that are correlated with the decision to commit to a choice option. This is similar to a threshold crossing process. However, in the current experiment, one is not drifting toward a bound. Rather, each sample serves both to update one's estimate of the underlying distribution (i.e., the state of the world, from which value estimates for each sample can be derived) as well as being a choice candidate. This scenario typifies many real-world decisions, where prices fluctuate over time, or potential partners have to be explored sequentially.

Authors' Contributions

V.D.C and B.B.A conceived and designed the experiment. V.D.C performed the experiment. B.B.A developed the Markov decision process model. V.D.C and B.B.A analyzed the data. V.D.C and B.B.A wrote paper.

Funding

Notes

Conflict of Interest: None declared.

References

Averbeck BB, Evans S, Chouhan V, Bristow E, Shergill SS. 2011. Probabilistic learning and inference in schizophrenia. Schizophr Res. 127:115–122.

DeGroot MH. 1970. Optimal statistical decisions. Hoboken: Wiley.

Ditterich J. 2006. Stochastic models of decisions about motion direction: behavior and physiology. Neural Netw. 19:981–1012.

Drugowitsch J, Moreno-Bote R, Churchland AK, Shadlen MN, Pouget A. 2012. The cost of accumulating evidence in perceptual decision making. J Neurosci. 32:3612–3628.

Ferguson TS. 1989. Who solved the secretary problem? Statist Sci. 4:282–296.

Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. 2006. Statistical parametric mapping: the analysis of functional brain images. Oxford: Academic Press.

Furl N, Averbeck BB. 2011. Parietal cortex and insula relate to evidence seeking relevant to reward-related decisions. J Neurosci. 31:17572–17582.

Gelman A, Carlin JB, Stern HS, Rubin DB. 2004. Bayesian data analysis. New York: Chapman and Hall/CRC.

Gluth S, Rieskamp J, Buchel C. 2012. Deciding when to decide: time-variant sequential sampling models explain the emergence of value-based decisions in the human brain. J Neurosci. 32:10686–10698.

Gold JI, Shadlen MN. 2001. Neural computations that underlie decisions about sensory stimuli. Trends Cogn Sci. 5:10–16.

Haber SN, Knutson B. 2010. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology. 35:4–26.

Heekeren HR, Marrett S, Ruff DA, Bandettini PA, Ungerleider LG. 2006. Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. Proc Natl Acad Sci USA. 103:10023–10028.

Heekeren HR, Marrett S, Ungerleider LG. 2008. The neural systems that mediate human perceptual decision making. Nat Rev Neurosci. 9:467–479.

Hill TP. 2009. Knowing when to stop: How to gamble if you must-the mathematics of optimal stopping. Am Sci. 97:126–133.

Ho TC, Brown S, Serences JT. 2009. Domain general mechanisms of perceptual decision making in human cortex. J Neurosci. 29:8675–8687.

Holmes AP, Friston KJ. 1998. Generalisability, random effects and population inference. NeuroImage. 7:s754.

Kim S, Hwang J, Seo H, Lee D. 2009. Valuation of uncertain and delayed rewards in primate prefrontal cortex. Neural Netw. 22:294–304.

Koestler A. 1960. The watershed. A biography of Johannes Kepler. Garden City, NY: Anchor Books.

Kolling N, Behrens TE, Mars RB, Rushworth MF. 2012. Neural mechanisms of foraging. Science. 336:95–98.

Lim SL, O'Doherty JP, Rangel A. 2011. The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. J Neurosci. 31:13214–13223.

Lindley DV. 1961. Dynamic-programming and decision-theory. Appl Statist. 10:39–51.

Lo CC, Wang XJ. 2006. Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nat Neurosci. 9:956–963.

Platt ML, Glimcher PW. 1999. Neural correlates of decision variables in parietal cortex. Nature. 400:233–238.

Puterman ML. 1994. Markov decision processes: discrete stochastic dynamic programming. New York: Wiley.

Ratcliff R. 1978. A theory of memory retrieval. Psychol Rev. 85:59–108.

Roitman JD, Shadlen MN. 2002. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J Neurosci. 22:9475–9489.

Seo H, Barraclough DJ, Lee D. 2009. Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J Neurosci. 29:7278–7289.

Seo M, Lee E, Averbeck BB. 2012. Action selection and action value in frontal-striatal circuits. Neuron. 74:947–960.

Shadlen MN, Newsome WT. 1996. Motion perception: seeing and deciding. Proc Natl Acad Sci USA. 93:628–633.