Neuron, Volume *93*

Supplemental Information

Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments

Yuan Chang Leong, Angela Radulescu, Reka Daniel, Vivian DeWoskin, and Yael Niv

Supplemental Figures

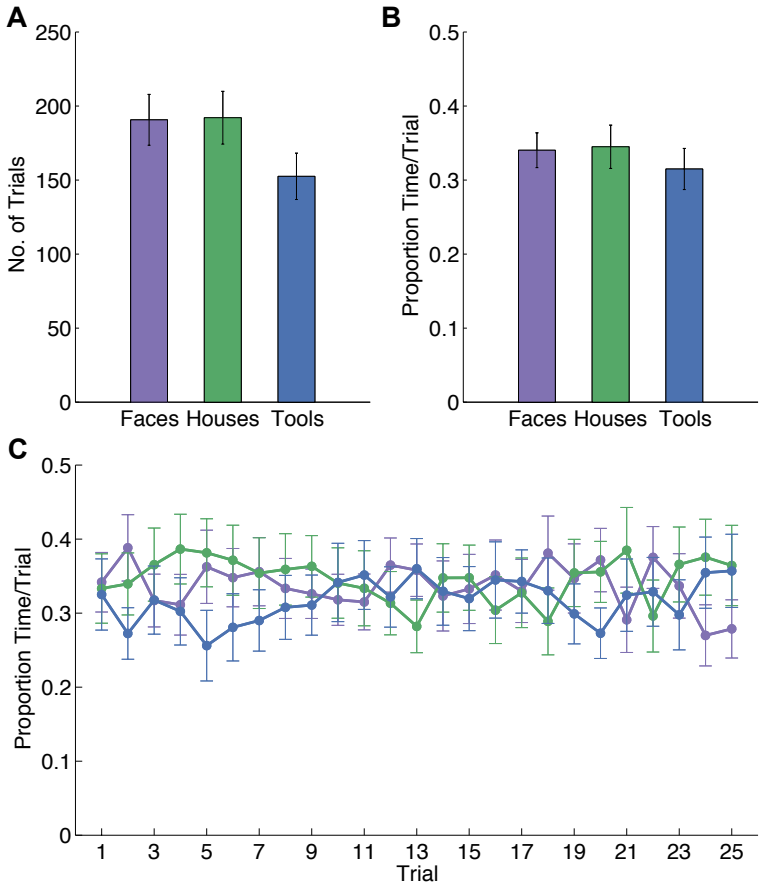

Figure S1, Related to *Modulation of Attention, Experimental Procedures*. The eye-tracking measure was not biased towards faces, landmarks or tools. A. Number of trials on which each dimension was the maximally attended dimension (that is, the dimension that participants looked at for the longest duration for that trial). There was no significant difference between dimensions (one-way repeated measures ANOVA: $F(2,48)=1.24$, $p = 0.3$). **B.** The average proportion of time participants looked at each dimension on each trial did not differ significantly (one-way repeated measures ANOVA: $F(2,48) = 0.24$, $p = 0.79$). **C.** The average proportion of time participants looked at each dimension on each trial was not significantly different between dimensions over the course of a game (two-way repeated measures ANOVA: $F(2,48) = 0.89$, $p = 0.42$). Error bars: SEM.

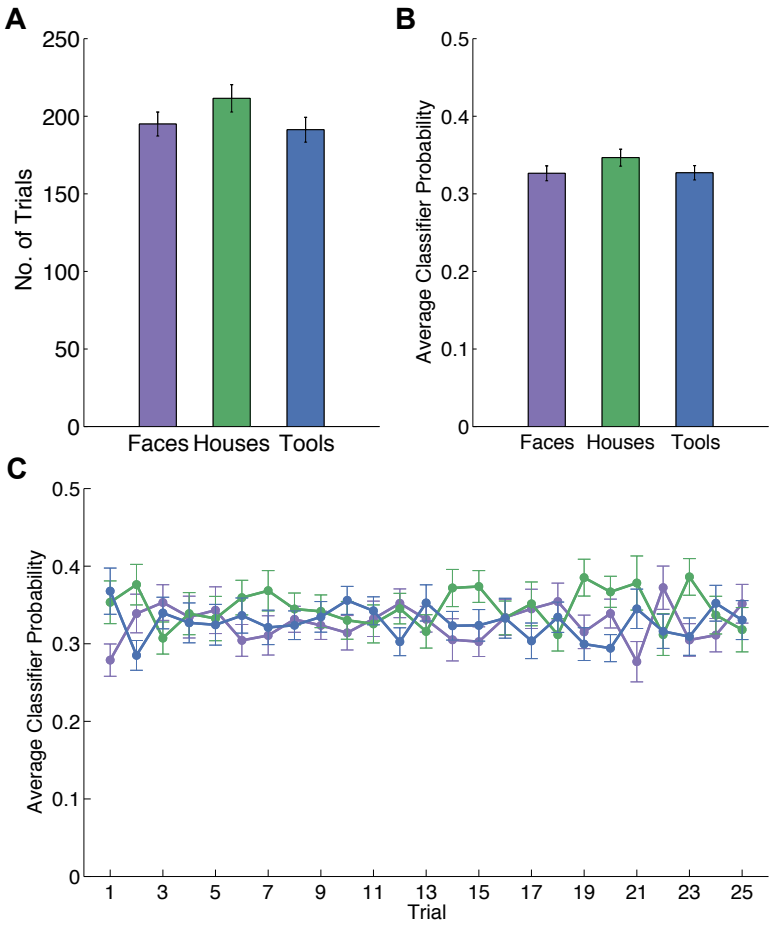

Figure S2, Related to *Modulation of Attention, Experimental Procedures*. The MVPA attention measure was not biased towards faces, landmarks or tools. A. The number of trials on which each dimension was the maximally attended dimension (that is, the dimension with the highest classification probability for that trial) was not significantly different (one-way repeated-measures ANOVA: $F(2,48) = 1.16$, $p = 0.32$). **B.** The average classifier probability of each dimension averaged over all trials and all participants was not significantly different between dimensions (one-way repeated measures ANOVA: $F(2,48) = 0.88$, $p = 0.42$). **C.** The average classifier probability of each dimension on each trial was not significantly different between dimensions over the course of a game (two-way repeated measures ANOVA: $F(2,48) = 073$, $p = 0.49$). Error bars: SEM.

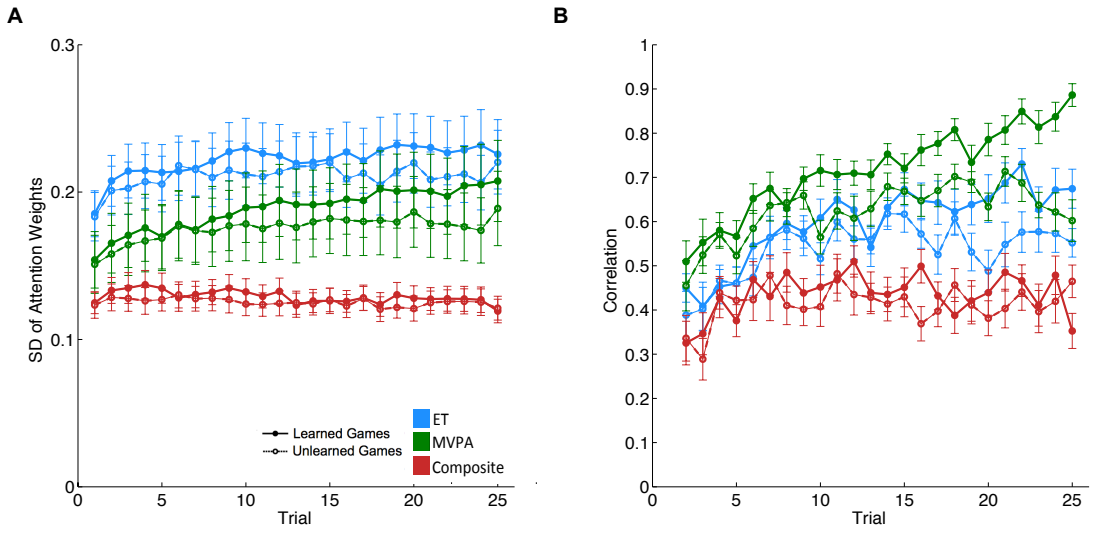

Figure S3, Related to *Modulation of Attention, Experimental Procedures*. The eye-tracking and composite measures of attention become increasingly focused and consistent over the course of a game. A. Standard deviation of the three attention weights as a function of trial in the game, for different attention measures, and for learned and unlearned games. Standard deviation of eye-tracking (linear mixed-effects model: $t(23.1) = 4.8$, $p < 0.001$) and composite attention weights ($t(24) = 2.79$, $p = 0.01$) increased over the course of a game, indicating a sharpening of attention. This increase was greater in learned games than unlearned games (eye-tracking weights: $t(25.4) = 3.9$, $p < 0.001$; composite weights: $t(22.5) = 1.8$, $p = 0.08$). The increase was not observed in the MVPA weights ($t(28.5) = -1.0$, $p = 0.31$). **B.** Pearson correlation between attention weights for consecutive trials, separately for each attention measure, and for learned and unlearned games. We used Pearson correlation to quantify the similarity in the distribution of attention between consecutive trials. The eye-tracking and composite measures of attention changed less from trial to trial as games progressed (eye-tracking weights: $t(23.0) = 6.4$, $p < 0.001$; composite weights: $t(24.0) = 5.6$, $p < 0.001$). This effect was more pronounced for learned games than for unlearned games (eye-tracking weights: $t(46.4) = 4.7$, $p < 0.001$; composite weights: $t(36.5) = 2.96$, $p = 0.005$), and not observed in the MVPA weights ($t(24) = 0.04$, $p = 0.97$). These results suggest that the MVPA measure might be noisier than the eye-tracking measure. However, an alternative possibility is that while the eye-tracking measure reflects goal-oriented, value-driven attention that sharpens as participants become increasingly certain about the most-rewarding feature, the MVPA measure captures random fluctuations in attention that nevertheless affect value computation and value update (c.f. deBettencourt et al., 2015). Our finding that combining the eye-tracking and MVPA measures of attention improves the model's ability to predict participants' choices (Fig. S4) indicates that the MVPA measure did contribute independently to our measure of attention. Learned games: games in which participants chose the most rewarding feature on each of the last five trials. Error bars: SEM.

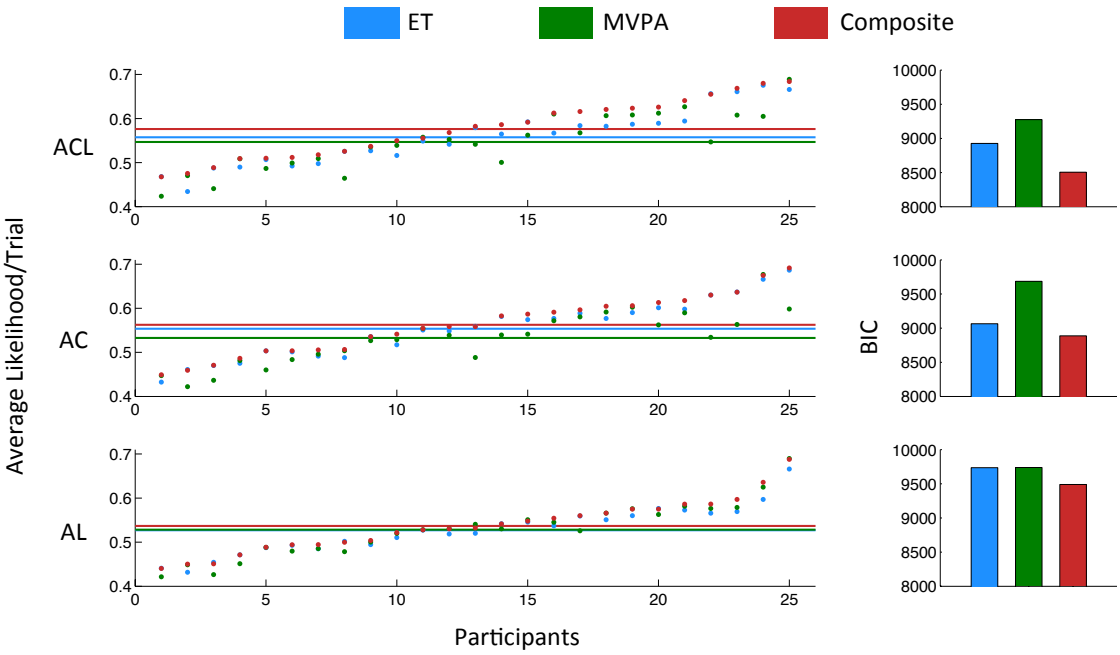

Figure S4, Related to Figure 3. Comparison of models with composite, eye tracking and MVPA measures of attention. Model fits of the ACL model (top), AC model (middle) and AL model (bottom) with eye tracking (blue), MVPA (green) and composite (red) measures of attention, all showed a better fit (higher average likelihood per trial and lower BIC score) for the composite measure of attention. Participants are ordered by average choice likelihood of the model that best explained their data. Comparison based on average likelihood/trial is shown on the left, while comparison based on BIC is shown on the right. For all three models, average likelihood per trial was significantly higher when using the composite measure than when using the eye tracking (ACL: $t_{24} = 5.80, p < 0.001$; AC: $t_{24} = 5.24, p < 0.001$; AL: $t_{24} = 4.48, p < 0.001$) or MVPA measure (ACL: $t_{24} = 4.86, p < 0.001$; AC: $t_{24} = 5.20\ p < 0.001$; AL: : $t_{24} = 4.13\ p < 0.001$).

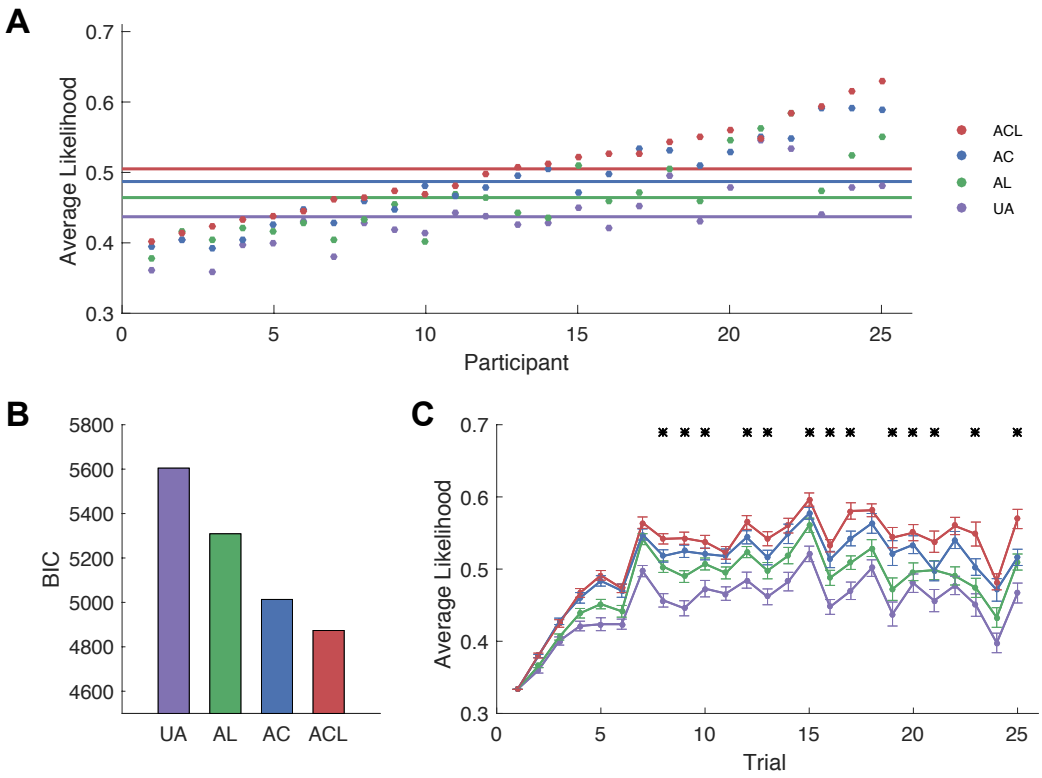

Figure S5, Related to *Figure 3*. The ACL model provides the best fit for unlearned games. Our study was aimed at investigating the neural and behavioral dynamics of learning what to attend to. As such, unlearned games were in many ways more revealing of the underlying learning dynamics, as there was no period of learned asymptotic behavior in these games. Here, we compare the models based on unlearned games only, and show that despite the reduced power, the ACL model still provides the best fit to the data. **A.** Average choice likelihood per trial for each model and each participant (ordered by likelihood of the model that best explained their data), calculated for unlearned games only (that is, games in which the participant did *not* consistently select the correct stimulus in the last five trials of the game). The ACL model explained the data significantly better than other models (ACL vs AC: $t(24) = 5.4$, $p < 0.001$; ACL vs AC: $t(24) = 5.9$, $p < 0.001$; ACL vs UA: $t(24) = 8.4$, $p < 0.001$). Solid lines: mean for each model across all participants. **B.** BIC scores for the four models aggregated over all participants also support the ACL model. **C.** In unlearned games, the average choice likelihood of the ACL model was significantly higher than the next best model from as early as the 8th trial. Error bars: within-subject SEM.

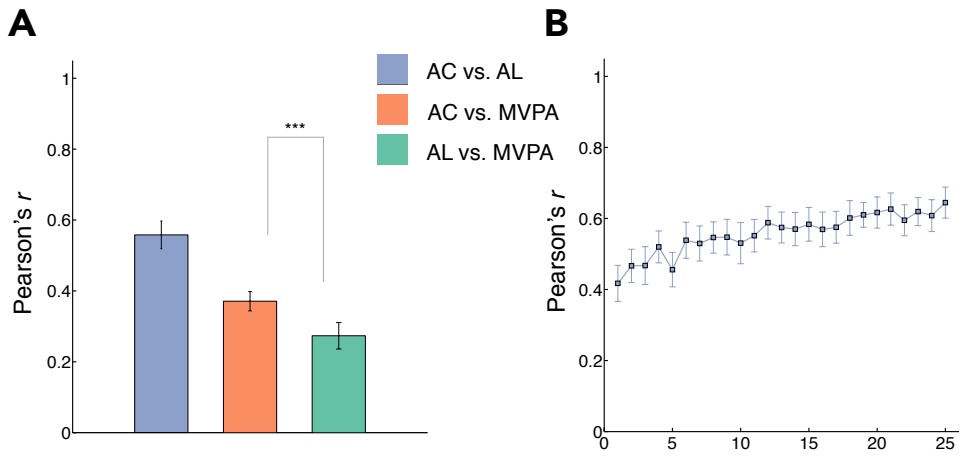

Figure S6, Related to *Measures of Attention*, *Experimental Procedures*. See also Supplemental Methods below. Comparison of attention weights derived by averaging over different intervals. Average Pearson correlation coefficient (r) obtained by computing pairwise correlations between attention vectors on every trial and averaging within each subject and then across subjects. Attention at choice (AC) and attention at learning (AL) were moderately correlated ($r = 0.56$). The MVPA measure of attention was significantly more correlated with AC rather than AL, suggesting that the MVPA measure might reflect attention at choice more than attention at learning (average r between MVPA and AC: 0.37, average r between MVPA and AL: 0.27, $t(24) = 4.63$, $p < .001$). **B.** Average trial-by-trial Pearson's r between AC and AL attention vectors increased throughout the game $F(24,24) = 4.95$, $p < .001$. Results were averaged across games within participant, and then across participants. Error bars: SEM. *** p < .001.

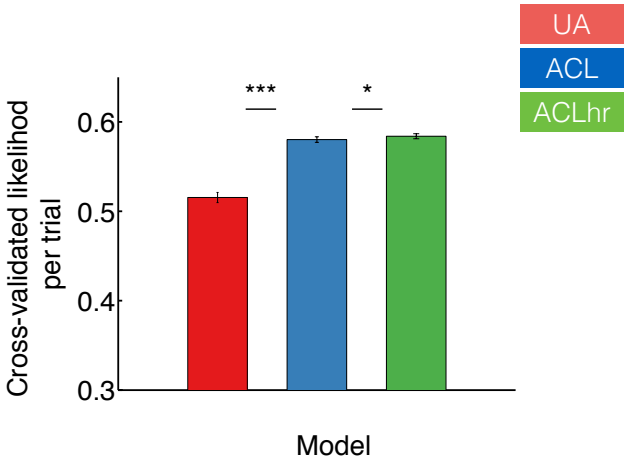

Figure S7, Related to *Figure 3*. A reinforcement learning model that uses separate measures of attention at choice and at learning (ACLhr) predicts the choice data significantly better than the ACL model that uses the same attention measure for both. See also Supplemental Methods below. Shown are results of a paired-sample t-test for the cross-validated likelihood per trial obtained by performing leave-one-game-out cross-validation. Error bars: SEM. *** p < .001; ** p < .01; * p < .05.

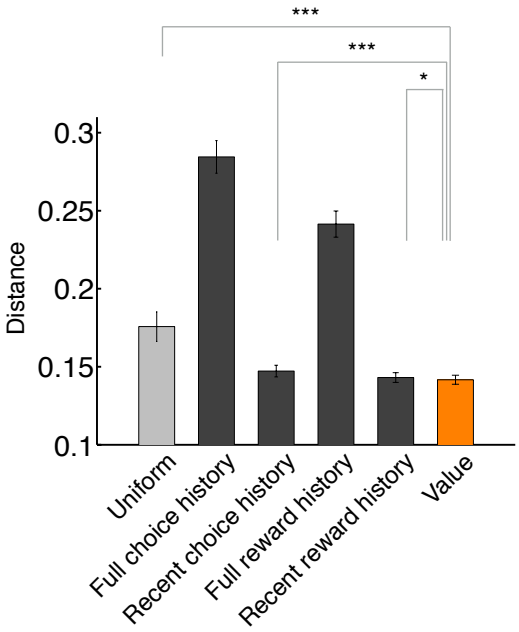

Figure S8, Related to *Figure 5*. A model that allocates attention based on learned values matches the composite attention measure better than models where attention is allocated based on past choices or rewards. Comparison of models of attention according to the root mean squared deviation (RMSD) of the model's predictions from the empirical data (lower values indicate a better model). For the uniform model (light gray), we computed the average per-trial RMSD between the observed attention vector on each trial and [1/3 1/3 1/3]. For the remaining models, we computed the RMSD by repeatedly fitting the models to all games except one and testing on the holdout game. Plotted is the subject-wise average per-trial RMSD from the composite measure of attention, calculated on the holdout games. The winning model (Value) is shown in orange. Error bars: SEM. *** $p < .001$; ** $p < .01$; * $p < .05$.

Supplemental Tables

Table S1, Related to *Choice models, Experimental Procedures*. Best-fit parameters for each model with accompanying constraints. Models parameters fit to all data from each participant separately. Parameters were optimized to minimize the negative log posterior probability of the participant's choice data given the model. β: softmax gain (inverse temperature); η: learning rate; ω_{ET}: smoothing weight for the eye-tracking attention measure; ω_{MVPA}: smoothing weight for the MVPA attention measure. Because the β parameter takes on unbounded values, to stabilize model optimization and to prevent numerical overflows, a Gamma(2,3) prior distribution over this parameter was used.

Model	Parameters	Constraints	Priors	Fit value ± SEM
ACL	β	$0 \leq \beta \leq \infty$	Gamma(2, 3)	13.5 ± 1.29
	η	$0 \leq \eta \leq 1$		0.39 ± 0.03
	ω_{ET}	$0 \leq \omega_{ET} \leq 1$		0.40 ± 0.05
	ω_{MVPA}	$0 \leq \omega_{MVPA} \leq 1$		0.29 ± 0.03
AC	β	$0 \leq \beta \leq \infty$	Gamma(2, 3)	16.0 ± 1.54
	η	$0 \leq \eta \leq 1$		0.36 ± 0.04
	ω_{ET}	$0 \leq \omega_{ET} \leq 1$		0.53 ± 0.06
	ω_{MVPA}	$0 \leq \omega_{MVPA} \leq 1$		0.26 ± 0.03
AL	β	$0 \leq \beta \leq \infty$	Gamma(2, 3)	11.9 ± 1.2
	η	$0 \leq \eta \leq 1$		0.50 ± 0.04
	ω_{ET}	$0 \leq \omega_{ET} \leq 1$		0.43 ± 0.06
	ω_{MVPA}	$0 \leq \omega_{MVPA} \leq 1$		0.40 ± 0.05
UA	β	$0 \leq \beta \leq \infty$	Gamma(2, 3)	18.3 ± 1.61
	η	$0 \leq \eta \leq 1$		0.33 ± 0.03

Table S2, Related to *Figure 4*. Clusters significantly correlated with model-based estimates of expected value and prediction error. All clusters survived cluster correction at the $p < 0.05$ level with cluster-forming threshold of $p < 0.001$. Coordinates are in MNI space and correspond to the center of mass of the cluster. In general, estimates from the ACL model were most closely correlated with neural data.

Region	*x (mm)*	*y(mm)*	*z(mm)*	Extent (voxels)
		Value Regressor		
$Value_{ACL}$				
vmPFC	-2	59	3	318
$Value_{UA}$				
R Occipital Pole	13	-100	1	333
L Occipital Pole	-21	-100	-4	236
		Prediction Error Regressor		
PE_{ACL}				
R Striatum	10	7	2	100
L Striatum	-6	5	3	80
L Superior Temporal Sulcus	-51	-62	14	343
R Intraparietal Sulcus	31	-78	33	209
L DLPFC	-49	19	18	190
R Parahippocampal Cortex	29	-37	-20	130
L Parahippocampal Cortex	-25	-43	-17	107
L Extrastriate Cortex	-22	-68	-12	112
L Precuneus	-5	-57	45	110

Table S3, Related to *Figure 6*. Brain areas correlated with attention switches. All clusters survived cluster-size correction ($p < 0.05$) with cluster-forming threshold of $p < 0.001$. Coordinates are in MNI space and correspond to the center of mass of the cluster.

Region	***x (mm)***	***y(mm)***	***z(mm)***	**Extent (voxels)**
	Switch trials – stay trials			
R dlPFC	44	35	30	277
L dlPFC	-46	29	32	235
Precuneus R IPS L IPS	-6	-61	47	2063
L preSMA	-6	13	49	151
R FEF	33	2	58	124
L fusiform cortex	-44	-52	-24	260
Cerebellum Lingual gyrus	3	-79	-12	1038
Cerebellum	42	-74	-25	290

Table S4, Related to *Figure 8*. PPI analysis with vmPFC activity as seed regressor and stay trials and switch trials as two task regressors. Shown are areas that showed a significant (negative) correlation with the stay-trials PPI regressor. All clusters survive FWE whole-brain cluster size correction ($p < 0.05$) with cluster-forming threshold of $p < 0.001$. No significant clusters were found for the switch-trials PPI regressors.

Region	***x (mm)***	***m)***	***z(mm)***	**Extent (voxels)**
R dlPFC	46	30	32	234
L dlPFC	-47	24	29	238
preSMA	1	42	36	108
R vlPFC	29	57	2	290
L vlPFC	-35	50	9	273
R striatum	5	9	6	130
L striatum	-14	22	3	103

Supplemental Experimental Procedures

Support vector machine classifier

Classification of fMRI data was performed using the SVM routine LinearNuSVMC (with Nu = 0.5) implemented in the PyMVPA package (Hanke et al., 2009). For multiclass problems, the algorithm first performs pairwise classification for each class (e.g., Face vs. Not Face, Landmark vs. Not Landmark and Tool vs. Not Tool). Pairwise classification probabilities are then calculated for each comparison using Platt scaling, which fits a logistic regression model to classifier evidence. Classifier evidence here refers to the signed distance between the multivariate measurement on a specific trial and the decision boundary for each class. The probability that a specific datapoint comes from each of the classes is then estimated by solving a linear system of equations with the pairwise-classifier probabilities, under constraints that the probabilities for all classes are positive and sum to 1 (Wu et al., 2004). The result of this procedure was therefore a vector of three probabilities (summing to 1) for each trial, which we used as the MVPA component of participants' attention to the respective dimensions on that trial.

Model comparison of choice models based on Bayesian information criterion

We also compared the choice models based on the Bayesian Information Criterion (BIC, Schwarz, 1978). We first optimized model parameters by finding participant-specific parameters that minimized the negative log likelihood of the participant's data given the model, using data from all games. These parameters were then used to compute the BIC approximation of model evidence, E_m:

$$E_M \approx log(p(D|M,\hat{\theta}_M)) - \frac{||\hat{\theta}||}{2} log\ N$$

where $p(D|M, \hat{\theta}_M)$ is the likelihood of the participant's choice data D given model M and maximum likelihood parameters $\hat{\theta}_M$, $||\hat{\theta}||$ is the number of free parameters in the model and N is the number of data points (trials). BIC values were then summed across participants to compare between models.

Modulation of attention by value and reward

As a measure of the trial-by-trial attention bias, we computed the standard deviation of attention weights on each trial. Linear mixed models were used to test for the main effect of trial in game on attention bias, as well as for the interaction between trial in game and whether the participant successfully learned that game. Models were estimated using the lmerTest R package. We tested for significance using t-tests, with Satterthwaite approximations to degrees of freedom. To assess how much the attention bias changed with each trial, we computed the Pearson correlation between the attention weights on consecutive trials. Linear mixed effects models were again used to test for the main effect of trial in game on the correlation of consecutive attention weights, as well as for the interaction between trial in game and whether the participant successfully learned that game (Fig. S1, S2, S3).

To investigate the relationship between attention bias and value, we performed a tercile split to bin trials according to strong, moderate and weak attention biases. We calculated, for each bin, the fraction of trials on which the most attended dimension was also the dimension with the highest feature value. We then tested if this fraction was higher on trials with stronger attention biases. Statistical significance for each pairwise comparison (i.e. strong vs. moderate, strong vs. weak, moderate vs. weak) was assessed using a bootstrap analysis. Specifically, attention weights for each game of each participant were replaced with those of a randomly selected game from the same participant. The ACL model was then run using these attention weights to generate estimates of feature values from participants' actual choices and outcomes, creating a "fictitious" dataset that controlled for the dependence between attention weights and feature values inherent in the ACL model. This process was repeated 1000 times, and a null distribution of t-statistics was generated for each pairwise comparison by performing the corresponding paired t-test on each iteration of the fictitious dataset. p-values were then determined by comparing the t-statistic obtained from the unshuffled data to the corresponding null distribution for that comparison.

We performed another tercile split to bin trials according to the standard deviation of the highest feature values (SDV) in each dimension. We then averaged the attention bias for each bin, and tested if the attention bias on high SDV trials was stronger than that on middle SDV trials and on low SDV trials, and if the attention bias on middle SDV trials was stronger than that of low SDV trials. Statistical significance was assessed using the same bootstrap method. For each SDV bin, we also calculated the fraction of trials on which there was a switch in attention. We defined a trial with an attention switch as one where the maximally attended dimension (i.e. dimension with the highest attention weight) was different from that in the previous trial. We then tested if the fraction of switches was higher in low SDV trials than in the middle and high SDV trials; and if the fraction of switches was higher in middle SDV trials than in high SDV trials. Statistical significance was again assessed using the bootstrap

method. All results were qualitatively similar if trials were binned based on the standard deviation of all feature values instead of SDV.

Finally, we ran a logistic regression to predict attention switches from reward history in the preceding five trials. We then tested if the regression coefficient on each trial was significantly different from zero. We excluded the first five trials of each game in all analyses of attention switches, as behavior might have been more random early on in the game.

Dissociating attention at choice and attention at learning

Previous empirical and theoretical work makes a distinction between attention for choice and attention for learning (Dayan et al., 2000). Unlike with the MVPA measure, the temporal resolution of eye-tracking allowed us to separately measure attention at the time of choice (using data from 200ms after stimulus onset and up to the time of choice), and attention at the time of learning (using measurements in the 500ms of outcome presentation), and to analyze these separately.

We first tested whether attention weights were different during these two time periods. For this, we computed the Pearson correlation coefficient between the two attention vectors on each trial (Fig. S6A), and found that on average, attention at choice and attention at learning were moderately correlated, with the correlation increasing over the course of a game, suggesting that as participants figured out the relevant dimension, they attended to the same dimension in both phases of the trial (Fig. S6B).

We next asked whether attention at choice and attention at learning had different effects on task behavior. For this, we fit a modified version of the ACL model (which we call the "high resolution ACL model", or ACLhr) that used separate attention weights at choice and learning. We found that the ACLhr model predicted choices slightly but significantly better than an ACL model, which used the same eye-tracking attention weights (combined across choice and learning) for both phases (Fig. S7). These results suggest that attentional processes at choice and at learning may reflect dissociable contributions to decision-making.

We note, however, that our design was not optimized to disentangle attention at choice from attention at learning. In particular, while participants had 1.5 seconds to make their choice, the outcome was presented for only 500ms. As such, there were fewer measurements during the time of outcome presentation than during the time of choice, resulting in a noisier estimate of attention at learning. Furthermore, the outcome was presented above and below the chosen stimulus, which meant that saccading to the outcome itself could contaminate our measure of attention at learning, as well as further reduce the time in which participants could look at the chosen stimulus after the outcome is revealed.

Supplementary References

Dayan, P., Kakade, S., and Montague, P.R. (2000). Learning and selective attention. Nat. Neurosci. *3*, 1218–1223.

deBettencourt, M.T., Cohen, J.D., Lee, R.F., Norman, K.A., and Turk-Browne, N.B. (2015). Closed-loop training of attention with real-time brain imaging. Nat. Neurosci. *18*, 470–475.

Hanke, M., Halchenko, Y.O., Sederberg, P.B., Hanson, S.J., Haxby, J.V., and Pollmann, S. (2009). PyMVPA: a Python Toolbox for Multivariate Pattern Analysis of fMRI Data. Neuroinformatics *7*, 37–53.

Schwarz, G. (1978). Estimating the Dimension of a Model. Ann. Stat. *6*, 461–464.

Wu, T.-F., Lin, C.-J., and Weng, R.C. (2004). Probability Estimates for Multi-class Classification by Pairwise Coupling. J. Mach. Learn. Res. *5*, 975–1005.