Effects of category-specific costs on neural systems for perceptual decision-making

Stephen M. Fleming¹, Louise Whiteley², Oliver J. Hulme¹,³, Maneesh Sahani², Raymond J. Dolan¹

¹Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG

²Gatsby Computational Neuroscience Unit, University College London, 17 Queen Square, London WC1N 3AR

³Laboratory of Neurobiology, UCL Institute of Ophthalmology, Bath Street, EC1V 9EL

Correspondence: Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG. s.fleming@fil.ion.ucl.ac.uk

Running head: Prospective costs in perceptual decisions

Keywords: Decision; psychophysics; fMRI; perception; value


Pages - 36; Figures - 6; Tables – 1; Supplementary Material – 5 Figures, 2 Tables.
Perceptual judgments are often biased by prospective losses, leading to changes in decision criteria. Little is known about how and where sensory evidence and cost information interact in the brain to influence perceptual categorisation. Here we show that prospective losses systematically bias the perception of noisy face-house images. Asymmetries in category-specific cost were associated with enhanced BOLD signal in a frontoparietal network, in keeping with the hypothesis that loss functions enact a particular task set that is communicated to visual regions. We observed selective activation of parahippocampal gyrus for changes in category-specific cost, in a region adjacent to that sensitive to stimulus information. Across subjects, greater shifts in decision criteria were associated with greater activation of the anterior cingulate cortex (ACC). Our results support a hypothesis that costs bias an intermediate representation between perception and action, expressed via general effects on frontal cortex, and selective effects on extrastriate cortex. These findings indicate that asymmetric costs may affect a neural implementation of perceptual decision making in a similar manner to changes in category expectation, constituting a step towards accounting for how prospective losses are flexibly integrated with sensory evidence in the brain.
INTRODUCTION

Visual perception has long been considered a process of inference about the most likely explanation of the stimulus – of inferring the state of the world most likely to have caused the pattern of photons hitting the retina (Helmholtz 1856). However, in an ecological context, perceptual categorisation needs to take into account not just probabilities, but also gains and losses (Bohil & Maddox 2001; Davison & Tustin 1978; Kersten et al. 2004). Think of a radiologist trying to diagnose whether a tumour is present or not in an X-ray. The sensory data may only weakly signal the probability of a tumour, but the potential costs of making a false alarm (further investigation of the occasional healthy person) are far less than the costs of missing a real tumour. In these circumstances, the perceptual judgment may be biased by the prospective loss, creating more false alarms than misses. These shifts in decision criteria are clearly important for survival: for example, in the North American forest, brown bears are more dangerous than black bears. If perception is impoverished, it is better to assume a particular bear-like object is a brown bear. Such scenarios are widespread in perception, and raise the question of how sensory evidence and prospective losses interact in the brain to bias perceptual categorisation in humans (Gold & Shadlen 2002; Heekeren et al. 2008).

Evidence from psychophysics suggests that prospective costs have strong effects on human perceptual decision criteria (Green & Swets 1966; Whiteley & Sahani 2008; Landy et al. 2007). Changes in value linked to particular regions of space are thought to alter intermediate representations between sensory coding and motor planning (Liston & Stone 2008), and to modulate spatially selective regions of early visual (Serences 2008) and somatosensory (Pleger et al. 2008) cortex, potentially via recruitment of fast attention-like mechanisms (Maunsell 2004; Serences 2008). However, it is unclear whether costs associated with particular categorical outcomes, such as tumour present/not present, or, as in this study, the presence of a face or house, are similarly mediated via category-sensitive visual areas (Fleming 2009). An alternative suggestion is that the decision is computed in frontoparietal regions thought to compare category evidence against a particular criterion (Heekeren et al.
at which point potential losses and gains are taken into account and used to bias response selection. This suggestion is supported by recent single-unit recording evidence showing that inducing shifts in decision criteria through changing a learnt category boundary (the speed of moving dots) modulates neural firing in the frontal eye fields (Ferrera et al. 2009). A third possibility is that changes in the payoff matrix create a particular “task-set” in fronto-parietal regions, which then acts to bias category-specific representations in visual cortex via backward connections (cf. Summerfield et al. 2006a). This suggestion is in accord with the tight linkage between activity in category-specific ventral visual regions and subjective reports of perceiving faces or houses (McKeef & Tong 2007; Summerfield et al. 2006b), even when the stimulus remains constant (Tong et al. 1998).

To examine how prospective losses bias perceptual categorisation, we manipulated the costs associated with visual categories (faces and houses) while obtaining brain data using functional magnetic resonance imaging (fMRI). We predicted that if biases are expressed through changes in classically defined object representations, we should observe asymmetries in the activity of face- and house-selective regions located in fusiform and parahippocampal gyri in inverse proportion to the loss associated with a particular category. Alternatively, if losses solely bias evidence accumulation, effects of category-specific cost may be constrained to fronto-parietal regions known to compare evidence for perceptual decisions. Finally, if the “task set” hypothesis is correct, we would predict that the categorical decision emerges out of an interaction between fronto-parietal and ventral visual activity.
MATERIALS AND METHODS

Subjects

Nineteen right-handed subjects participated in the psychophysics session (7 male; 19 – 44 years of age; mean age, 25.0 years). All had normal or corrected-to-normal vision, and no history of psychological or neurological illness. Of these participants, sixteen were scanned. One participant was excluded at this stage due to a change of response strategy in the scanner that led them to disregard the face/house image, leaving fifteen subjects (5 male; 19 – 27 years of age; mean age, 23.9 years) in the analysis. Samples of this size have been shown to be highly sensitive to true underlying effects in a classical inference framework, assuming, as in the present study, a relatively large number of scans per subject (Friston et al. 2002). The study was approved by the Institute of Neurology (University College London) Research Ethics Committee.

Stimuli

We used 10 neutral faces (5 male, 5 female) from the KDEF face set (Lundqvist et al. 1998) and 10 houses (photographed by the first author). The stimuli were all cropped to be of equal size and converted to grayscale. To create a stimulus continuum, we adapted a technique used by Heekeren et al. (2004). Fourier transforms (FT) of each image were computed, producing 20 magnitude and 20 phase matrices. The average magnitude of all house and face stimuli was then stored. On each trial, a linear combination of one randomly selected house and face phase matrix was computed, plus a constant proportion (0.35) of a stored white noise matrix (see Supplementary Fig. 1). The resulting phase matrix was then recombined with the average magnitude matrix of the whole stimulus set using an inverse FT. Finally each image was normalised to have average luminance equal to that of the screen background and constant RMS contrast.

Face/house images were presented for 100ms on a grey background using Cogent 2000 (www.vislab.ucl.ac.uk/cogent.php) running in MATLAB. In the psychophysics
experiment, stimuli were presented using a 20.1 inch Dell 2001FP monitor running at a refresh rate of 60 Hz, situated in a dimly lit room. All images subtended 4 degrees of visual angle at a viewing distance of 60 cm. During the fMRI experiment, stimuli were presented using an NEC LT157 LCD projector, viewed by subjects via an adjustable mirror. At the beginning of each scanning session, a custom-written Cogent routine adjusted stimulus size and position to match that used in the psychophysics.

**Psychophysics**

The experiment was divided into two separate sessions. The first session involved acquiring psychophysics data outside the scanner; the second session repeated the same task during fMRI data acquisition. Participants were not informed of the image continuum, but were instead asked to categorise each briefly presented stimulus as either a noisy face or house. Participants found this task natural and were unaware of any blend between the two image categories when debriefed. Before introducing a monetary component to the task, each participant completed 540 trials of simple face/house discrimination using the same stimulus timings as in the main experiment. Face and house responses were made using left and right-hand key presses respectively. There were 15 stimulus levels, spaced in equal steps from 100% house to 100% face phase, enabling us to plot out each individual’s psychometric function. The point of subjective equality (PSE) for each subject was then used to define face and house categories for the category-specific cost task.

The category-specific cost task involved further face/house discrimination under asymmetric losses for incorrect responses. There were three levels of the cost factor: face value (FV; -50p for an incorrect “house” response, -10p for an incorrect “face” response), neutral value (NV; -30p for an incorrect “house” response, -30p for an incorrect “face” response) and house value (HV; -10p for an incorrect “house” response, -50p for an incorrect “face” response). Before each block of trials, subjects were given an endowment of £10, and informed that they would keep any money they did not lose on the task. We used losses rather
than gains as for the small amounts of money used here, losses were hypothesised to
generate a greater behavioural impact on decision criteria than gains (Kahneman & Tversky
1979). Cumulative feedback screens displaying the current total were provided only every 15
trials, to avoid incremental learning of decision strategy via trial-by-trial adjustments
(Whiteley & Sahani 2008).

Image phase (15 levels) was randomised and orthogonal to the cost factor, which was
signalled to participants prior to the face/house stimulus on every trial (Fig. 1a). The cost
level changed every two trials. Subjects completed nine experimental blocks of 140 trials
each, spanning a single session lasting around 3 hours including breaks. Note that when the
penalty for answering “house” incorrectly is greater, a reasonable strategy is to answer “face”
more often when uncertain of the answer.

fMRI experiment

The fMRI experiment took place within a week of the psychophysics, and employed the same
task with minor alterations. Subjects completed four runs of 105 trials. The initial endowment
for each block was £12, and feedback was given every 10 trials. Each trial began with a cost
cue presented for 1820ms, followed by a variable interval of 100ms – 3000ms during which a
fixation cross was presented. The face/house image was then presented for 100ms, and
subjects were able to respond immediately following the onset of the face-house image.
Following the offset of the face-house image, a fixation cross was presented for a variable
interval of 1625ms – 3625ms prior to the start of the next trial. The buttons indicating face
and house responses were switched halfway through the session, so that each subject made
face and house decisions with both left and right button presses. To avoid switch costs, a
short training run was given with the new response mapping without any imaging data being
collected.

Stimuli were presented in a permuted randomised fashion, so that the full phase range
was covered every ~ 7 trials. Similarly, the three cost levels were cycled every 6 trials
(changing every two trials, as in the out-of-scanner psychophysics), while keeping stimulus phase and cost orthogonal. This cycling over ~ 30 s was chosen to match the filter properties of the canonical haemodynamic response function (HRF), maximising power for estimating the cost- and stimulus-related parameters in our event-related analysis.

fMRI acquisition

Images were acquired using a 3T Allegra scanner (Siemens, Erlangen, Germany). BOLD-sensitive functional images were acquired using a gradient-echo EPI sequence (48 transverse slices; TR, 3.12 s; TE, 65 ms; 3 x 3 mm in-plane resolution; 2 mm slice thickness; 1 mm gap between adjacent slices; z-shim, + 0.6 mT/m; positive phase encoding direction; slice tilt, - 45 degrees) optimised for detecting changes in the parahippocampal region and fusiform gyrus (Weiskopf et al. 2006). Four runs of 213 volumes were collected for each subject, followed by a T1-weighted anatomical scan and local field maps.

Behavioural data analysis

Subjects’ psychophysical responses outside the scanner were modelled using a cumulative normal psychometric function incorporating a random lapse term (Wichmann & Hill 2001), assuming binomial response counts (see Whiteley & Sahani 2008, for full details of the mathematical model). The curve for each cost condition (indexed by $j$) had three free parameters: the mean ($\mu_j$) reflecting the PSE, the slope ($\rho_j$) reflecting a subject’s uncertainty over the whole stimulus range, and the lapse rate ($\varepsilon_j$) reflecting motor errors and lapses of attention. In the equation below, $CP_j$ gives the probability of answering “face” for each given stimulus phase combination, $x_i$, in the $j$th cost condition:

$$CP_j = (1 - \varepsilon_j) \cdot \frac{1 + \text{erf}(\sqrt{\pi} \cdot \rho_j \cdot (x_i - \mu_j))}{2} + \frac{1}{2} \varepsilon_j$$

where $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$
We used gradient descent algorithms to find the parameter values that produced optimal curve fits to the observed data. We additionally implemented a Bayesian model comparison to determine whether sharing each of the parameters $\mu, \rho$ and $\epsilon$ between cost conditions gave better fits to the data than allowing each to be optimised separately. We fitted curves to the in-scanner data in the same manner as for the psychophysics data for visualisation purposes.

To define category uncertainty for a given stimulus phase, we rectified each individual’s psychometric function under the neutral cost condition around the PSE and normalised the result such that the range varied from 0 to 1.

$$U_i = \frac{|0.5 - CP_i| - 0.5}{0.5}$$

Following Grinband et al. (2006), this procedure defines the PSE as having maximal categorical uncertainty and 100% house/face phase as having minimal categorical uncertainty. Note that this use of the term “uncertainty” refers to the difficulty of categorisation for a particular face-house phase composition, as opposed to the overall uncertainty about the task expressed in the slope of the psychometric function across the full phase axis. The latter might be expected to change with, for example, practice, stimulus duration, or lighting conditions.

The psychometric function fits to the in-scanner data were not a robust basis for inference, given the lower number of trials per data point compared to the psychophysics session. Consequently, we conducted further behavioural analysis using the framework of signal detection theory (SDT) (Green & Swets 1966). Stimuli were classified as being faces or houses, depending on which side of the PSE they fell, yielding a classic $2 \times 2$ stimulus-response table for each cost condition. This approach implicitly approximates the stimulus continuum as being drawn from two overlapping Gaussian distributions, one for each category. This allowed us to compute subject-specific measures of sensitivity ($d'$) and criterion ($c$) separately for each cost condition. Despite being a cruder measure of behavioural
performance than the psychometric function fitting described above, this method provides a useful index of whether value primarily affects sensory discrimination or decision/response criteria (Macmillan & Creelman 2005), and circumvents the problem of fewer trials in the scanner leading to unreliable psychometric function fits.

fMRI data preprocessing and analysis

Functional data were analysed using SPM5 (Statistical Parametric Mapping; www.fil.ion.ucl.ac.uk/spm). The first five volumes of each run were discarded to allow for T1 equilibration. Using the FieldMap toolbox (Andersson et al. 2001), field maps were estimated from the phase difference between the images acquired at the short and long TE. The EPI images were then realigned and unwarped using the created field map, and slice-timing correction applied to align each voxel’s timeseries to the acquisition time of the middle slice. Each subject’s T1 image was segmented into grey matter, white matter and CSF, and the segmentation parameters were used to warp the T1 image to the SPM Montreal Neurological Institute template. The resulting normalization parameters were then applied to the functional data. Finally, the normalized images were spatially smoothed using an isotropic 8mm full-width half-maximum Gaussian kernel.

fMRI timeseries were regressed onto a composite general linear model (GLM) containing delta (stick) functions representing the onsets of the cost cue, stimulus, response and cumulative feedback (see Supplementary Table 1). These delta functions were convolved with the canonical HRF, and low-frequency drifts were excluded with a high-pass filter (128 s cutoff). Short-term temporal autocorrelations were modeled using an AR(1) process. The stimulus delta functions were separated into three regressors dependent on the cost condition on each trial (face value - FV, neutral value - NV and house value - HV). Each stimulus onset was parametrically modulated by two subject-specific functions. The first was the choice probability ($CP$) curve fitted to the out-of-scanner psychophysics data in the neutral value condition. The second was the categorical uncertainty function ($U$), again derived from the
out-of-scanner psychophysics data, and orthogonalised with respect to choice probability (see above for mathematical definitions). The cumulative feedback stick function was also modulated with the amount of money lost on the previous 10 trials. To investigate interactions of value and response hand, the response delta function was separated by cost, decision and response hand, giving a 3 (cost; FV vs. NV vs. HV) × 2 (decision; face vs. house) × 2 (response; left vs right) factorial combination. Motion correction regressors estimated from the realignment procedure were entered as covariates of no interest.

**Statistical inference**

Statistical significance was assessed using linear compounds of the model parameters (regression coefficients of the trial-specific stimulus functions above), for each subject. These contrast images were then entered into a second-level random effects analysis using a one-sample t-test against zero to assess group-level significance. Cluster-based statistics (Friston et al. 1994) were used to define significant activations based both on their intensity and spatial extent. Clusters were defined using a height threshold of $P < 0.001$ and corrected for multiple comparisons across the whole brain using family-wise error correction (FWE) and a threshold of $P < 0.05$. Images are displayed at the cluster-defining threshold of $P < 0.001$ using MRIcron (http://www.sph.sc.edu/comd/rorden/mricron/). Small-volume correction (SVC) was applied to category-specific responses by using anatomical masks for fusiform and parahippocampal gyri as specified in the PickAtlas toolbox (Maldjian et al. 2003).

Percent signal change was extracted from clusters of interest for further analysis by averaging over subjects and sessions using MarsBar (Brett et al. 2002). Estimated time courses within clusters are plotted at seven TRs following stimulus onset using a finite impulse response (FIR) model. We note that timecourses are plotted for illustration purposes only, inference having first been carried out using appropriate adjustments for multiple comparisons within SPM.
RESULTS

We asked subjects to categorise randomly presented noisy images as either faces or houses (Fig. 1a and Supplementary Fig. 1). Before each block of trials subjects received a financial endowment, and were informed that they could keep whatever amount they did not lose in penalties for incorrect answers. On each trial variable monetary losses were imposed for incorrect face and house decisions: either the face response had lower cost (face value condition; FV), the house response had the lower cost (house value condition; HV), or they were balanced (neutral value condition; NV). In our fMRI experiment subjects performed the perceptual categorisation task inside the scanner, enabling us to obtain measures of regional brain activity associated with task performance over time. Face and house responses were made with both left and right button presses, allowing us to further investigate whether changes in value interacted with response-specific brain activity, by decoupling the latter from the decision category.

Behavioural results

Subjects’ average point of subjective equality (PSE) in the face-house discrimination pre-test was 53.9 ± 9.15 % face phase. Categorisation probability data from a representative subject’s psychophysical results are shown in Figure 1b. To explore the effects of asymmetric cost on choice probability, we fit psychometric functions to the data with either shared or separate mean ($\mu_j$), slope ($\rho_j$), and lapse rate ($\varepsilon_j$) parameters for the three cost conditions (indexed by $j$). We then carried out Bayesian model comparison, thereby revealing which of the eight possible parameter structures (single vs. shared mean × single vs. shared slope × single vs. shared lapse rate) best accounted for the effects of manipulating asymmetric cost on choice probabilities.

All subjects consistently shifted their responses towards the category carrying lower cost, as expected (Fig. 2b and Supplementary Fig. 2). Paired-sample t-tests confirmed that average shifts were significantly different from NV for both FV ($t_{(18)} = 5.95, P < 0.0001$) and
HV trials \(t_{(18)} = 4.98, P < 0.001\). There were also small, but significant, differences in psychometric function slope between value conditions (white markers in Fig. 2b; one-way ANOVA, significant effect of value: \(F_{(2,36)} = 4.61, P < 0.05\)). Consistent with these results, Figure 2a shows that the model with both separate means and slopes was the best model of the data, despite the Occam’s razor-like penalty for greater model complexity inherent in Bayesian model comparison. However, the magnitude of the difference between the summed log model evidences for shared and separate slopes is rather small, making firm conclusions about differences in slope between conditions difficult.

As fewer trials precluded fitting reliable psychometric curves to the choice data in the scanner, we carried out a signal detection analysis (Green & Swets 1966) to characterise in-scanner behaviour, collapsing stimuli into either face or house categories based on each individual subject’s PSE. This analysis confirmed that asymmetric cost led to deviations of the decision criterion in the predicted direction, relative to the neutral value condition (\(c\); FV, \(t_{(14)} = 5.82, P < 0.0001\); HV, \(t_{(14)} = 5.78, P < 0.0001\)) but did not change category discriminability (\(d'\); \(F_{(2,28)} = 0.41, P > 0.5\); Fig. 2c).

Importantly, mean reaction times (RT) did not differ across value conditions (psychophysics, \(F_{(2,36)} = 0.70, P > 0.4\); in-scanner, \(F_{(2,28)} = 1.67, P > 0.2\), suggesting that any bias-related differences we find in brain activity are not driven by systematic differences in task difficulty (Fig. 2d). RT was however significantly correlated with categorical uncertainty (see Methods and Supplementary Fig. 3; psychophysics, mean \(r = 0.79 \pm 0.092, n = 19\); in-scanner, mean \(r = 0.56 \pm 0.21, n = 15\)).

fMRI results

Cost-related regions

We first identified regions involved in processing the extra demand of integrating asymmetric cost by computing the COST > NEUTRAL contrast (FV + HV > 2NV). A frontoparietal
network (Fig. 3 and Table 1) was consistently active for both types of asymmetric cost condition compared to neutral ($P < 0.05$, whole-brain corrected), suggesting its involvement in the biasing of perceptual decisions as a function of category-specific cost. In addition to frontoparietal areas, we also found increased activity in a cluster spanning the subthalamic nucleus (STN) region, thalamus and caudate nucleus ($P < 0.05$, whole-brain corrected), consistent with the proposal that subcortical loops are involved in setting response criteria (Lauwereyns et al. 2002; Lo & Wang 2006). Recent modelling work has shown the type of onset function (stick vs. epoch) can affect activation patterns in decision-making tasks, even when decision times are short (Grinband et al., 2008). To check that our cost-related activations were robust to the type of onset function used, we constructed a second design matrix in which the decision period was modelled using a variable epoch. Similar activations were observed in both models, with the overlap shown in Supplementary Figure 5.

**Stimulus-selective regions**

We next identified functional regions of interest (ROIs) sensitive to stimulus category information. In our group analysis, a cluster in right fusiform gyrus (FG; $P < 0.05$, small volume corrected (SVC); Fig. 4a) correlated with the probability that the image was categorised as a face, averaging over cost condition. Conversely, activity correlating with the probability of an image being categorised as a house was expressed in bilateral parahippocampal gyrus (PHG; right, $P < 0.05$, SVC; left, $P < 0.05$, SVC; Fig. 4b). No other clusters correlating with the categorical choice probability survived whole-brain correction.

To test our first hypothesis, that category-specific costs directly affect responses in the ventral visual stream, we computed the signal change for each cost condition in each stimulus-selective ROI identified above. One-way ANOVAs (FV vs NV vs. HV) revealed effects of cost on right PHG ($F_{(2,28)} = 4.80$, $P = 0.002$), but not left PHG or right FG ($F_{(2,28)} < 2.5$, $P > 0.1$). Further investigation of the pattern of differences in right PHG revealed increases in the HV condition compared to NV ($t_{(14)} = 3.09$, $P = 0.008$), a trend for increases
in the FV condition ($t_{(14)} = 1.77$, $P = 0.10$), but no evidence for differences between the
category-specific FV and HV conditions ($t_{(14)} = 1.26$, $P = 0.23$). A similar trend for non-
specific increases under asymmetric cost conditions can be seen in all three ventral visual
areas (Fig. 4c, d), and an omnibus ANOVA in which region was included as a separate factor
indicated an overall significant effect of cost ($F_{(2,28)} = 5.95$, $P = 0.007$). Together, these
analyses indicate that asymmetric cost has a general driving effect on ventral stimulus-
selective regions, but in a manner that does not appear to discriminate between stimulus
categories.

**Category-specific effects of cost**

The account presented thus far indicates that category-selective stimulus information is to
some degree represented independently of category-specific biases induced by changes in the
payoff matrix. In other words, payoff asymmetries lead to only general, not category-specific,
increases in the signal in voxels sensitive to stimulus (face or house) information. However,
the mechanism by which asymmetric value information brings about a change in perceptual
decision (such as a bias towards responding “face”) remains unclear. To address this
question, we computed the category-specific value contrasts (FV > HV; HV > FV). A cluster
in left parahippocampal gyrus (-33, -36, -15; $P < 0.05$, SVC) responded specifically to
increases in house value (decrease in cost for responding “house”), lateral and anterior to the
stimulus-selective cluster we characterised previously (Fig. 5a). No significant clusters were
evident in the opposite (FV > HV) contrast even at a liberal ($P < 0.01$) defining threshold.
Together, these findings suggest that category-specific costs exert effects on the ventral
visual pathway (at least for the bias towards responding “house”).

We noted that while the direction of decision criterion shifts under asymmetric cost was
consistent across subjects (Fig. 2c), individual differences in the size of this shift were
evident in the behavioural data (Fig. 5b). To explore whether subjects who exhibit greater
decision criterion shifts also show greater activity within regions that are the putative sources
or targets of these shifts, we regressed the category-specific value contrasts (FV > HV; HV >
FV) against between-subjects covariates encoding the amount of behavioural bias in the
relevant asymmetric value condition (FV criterion shift; HV criterion shift). Subjects who
displayed greater criterion shifts in the HV condition tended to activate the anterior cingulate
cortex (ACC; 6, 36, 30) more than subjects who shifted to a lesser degree (Fig. 5c, d; \( P < 0.05 \), whole-brain corrected). Again, as for the simple main effect of FV > HV, no significant
correlations were found with individual differences in the FV criterion (\( P > 0.001 \),
uncorrected).

**Categorical uncertainty**

To look for brain regions responsive to categorical uncertainty, we regressed a parameter
which essentially measures how close to chance the subject is in deciding whether the
stimulus is a face or a house (see Materials and Methods) onto the fMRI signal at the time of
choice. Dorsal medial frontal (paracingulate) cortex (dMFC; 6, 12, 51) and right anterior
insula (42, 24, -3) showed positive correlations with categorical uncertainty (both \( P < 0.05 \),
whole-brain corrected; Figure 6a). We were interested in establishing whether this
uncertainty-related BOLD signal was independent or overlapping with the frontoparietal
regions found to be active under conditions of asymmetric cost. By exclusively masking the
COST > NEUTRAL contrast for regions correlating with categorical uncertainty at a liberal
\( P < 0.05 \), uncorrected) threshold, we found that left vlPFC, left caudate/thalamus/STN and
bilateral FEF were specifically active under conditions of asymmetric cost, independent of
changes in categorical uncertainty (Supplementary Table 2). Conversely, dMFC activity was
independent of changes in category value (Fig. 6b), indicating a partial dissociation in the
brain between regions encoding changes in categorical uncertainty and prospective costs
during perceptual categorisation. The negative effect of the categorical uncertainty regressor
(testing for greater activity in “easy”, certain decisions) was seen in ventromedial prefrontal
cortex (vmPFC; 9, 33, -6; \( P < 0.05 \), whole-brain corrected; Fig. 6a). Interestingly, this effect
of perceptual certainty activated a region of vmPFC adjacent to that correlating with the
amount of money saved at each intermittent feedback screen (Supplementary Fig. 4), suggesting a neural relation between perceptual certainty and reward (cf. Maddox & Bohill, 2003).

**Interaction of cost with motor planning**

The previous analyses identified brain regions that responded preferentially to a particular direction of bias (towards responding “house”). Our design further allowed us to ask whether any bias effects are expressed at the level of the motor system, given that response hand (left or right) was orthogonal to decision (face or house). Interactions of cost asymmetry with response hand were computed by coding each trial as to whether the left or right hand was assigned to a high or low cost response (face or house), and examining the interaction of cost condition with response hand (see Materials and Methods). No effects were found ($P > 0.001$, uncorrected), suggesting that the biasing effects of asymmetric value occur upstream of effector-specific response planning.
Perceptual judgements are often affected by the potential costs associated with different categorisations; for example, during a radiologist’s assessment of an X-ray image or deciding whether a dangerous or benign animal is hidden in the trees. Here we investigated the brain mechanisms that integrate prospective costs and sensory evidence during categorisation. Our behavioural manipulation systematically biased the perception of a noisy image using asymmetric costs, leading to shifts in decision criteria. These shifts functioned to reduce monetary losses, by biasing decisions toward the category with lower cost when the participant was unsure of the answer. Using fMRI, we asked whether category-specific shifts were reflected by changes in frontoparietal areas known to accumulate evidence leading to perceptual categorisation (e.g. Heekeren et al. 2004), in ventral visual regions known to encode category-specific information about faces and houses, and/or via recruitment of regions thought to communicate information about the currently relevant task set (Serences et al. 2004; Summerfield et al. 2006a). Our data best fit the latter, “task set” hypothesis. The requirement to integrate asymmetric cost information into perceptual decisions robustly activated a frontoparietal network, despite conditions being closely matched for expected value and reaction time. In addition, a cluster in the thalamus/caudate was active under asymmetric cost, consistent with subcortical loops being important for the setting of decision criteria (Lo & Wang 2006; Simen et al. 2006). A specific effect on ventral visual areas (parahippocampal gyrus) was found under decreasing cost for houses, anatomically adjacent to the integration of stimulus information. Finally, subjects who showed greater shifts in decision criteria towards houses demonstrated greater activation of the anterior cingulate cortex (ACC), a region thought to be pivotal in the adjustment of decision strategy (Behrens et al., 2009; Botvinick et al., 2001).

**Sources of category-specific bias**
The dorsal frontoparietal network active under asymmetric cost is similar to that commonly activated in studies of transient allocation of attention (Corbetta et al. 2008; Corbetta & Shulman 2002; Yantis et al. 2002), and has been recently implicated in the modulation of early visual cortical activity by rewards tied to particular locations in visual space (Serences 2008). It is plausible that changes in category-specific costs co-opt a similar network. Low-level changes in arousal or task difficulty are unlikely to be explanations for this widespread increase in activity, as RTs and potential gains/losses were matched across conditions.

Instead, our findings indicate that frontoparietal (ventrolateral prefrontal cortex, insula, intraparietal sulcus, frontal eye fields) and subcortical (anterior thalamus/STN) regions are recruited when information about payoffs needs to be incorporated into a perceptual decision. Bilateral activation was found at the junction of the precentral and superior frontal sulci, consistent with the location of the frontal eye fields (FEF) (Lobel et al. 2001), which are known to have a causal role in the modulation of sensory cortex (Ekstrom et al. 2008) and encode shifts in decision criteria (Ferrera et al. 2009). In addition, activation in ventrolateral prefrontal regions including insular cortex is consistent with involvement both in the accumulation of sensory evidence (Ho et al. 2009; Romo et al. 2004) and the incorporation of gains and losses in decision-making (Leon & Shadlen 1999; Watanabe & Sakagami 2007).

Indeed, the network outlined above may be recruited more generally when shifts in decision criteria are induced by manipulations other than asymmetric payoffs. This view is supported by previous findings of modulation of subcomponents of this network when decision criteria are shifted through changes in category boundary – specifically, BOLD signal in anterior thalamus/caudate and insula/vIPFC (Grinband et al., 2006; Li et al., 2009) and single-unit activity in FEF (Ferrera et al., 2009).

The present analyses cannot pin down the source of the bias towards houses and faces, as main effects of category-specific bias were not observed in the network discussed above. However, it is possible that local neural subpopulations within these areas encode biases towards face and house categories. This suggestion is supported by a recent study by Rorie et
al. (2010) in monkeys, demonstrating that asymmetric payoffs in a perceptual decision task bias the initial firing rate of individual neurons in the intraparietal sulcus coding for a saccadic response to one of two particular targets. Similarly, using fMRI, distributed patterns in the inferior frontal gyrus/insula have been found to discriminate the direction of criterion shifts in a visual categorisation task (Li et al. 2009). Our finding that the ACC tracks individual differences in the extent of a signed criterion shift is also consistent with category-specific payoff information being represented in frontal cortex.

**Differential effects of house and face cost**

When the signal change in stimulus-selective ROIs was calculated, general but not selective effects of asymmetric cost were observed. In contrast, a parahippocampal region anterior to the stimulus-selective areas was significantly more active when houses rather than faces were more valuable. Due to the potential impact of our smoothing and thresholding procedures, we are cautious in attributing separate locations to the stimulus- and cost-sensitive regions of the PHG. However, we note evidence suggesting local separation of task- and stimulus-driven regions using neural stimulation coupled with high resolution imaging (Ekstrom et al. 2008); similarly, differences between stimulus- and task-driven localisations have been reported in the fusiform gyrus during perceptual decision making (Philiastides & Sajda, 2007).

Category-specific biases were seen in ventral visual regions for increases in house, but not face, value. Furthermore, across subjects, a correlation with decision criteria was seen in the ACC for bias towards houses, but not faces. While we are cautious about over-interpreting null results, we note that previous studies examining attentional and decisional biases towards faces and houses have also found asymmetric effects of the two categories. Specifically, Summerfield et al. (2006) found that mistaken categorisations of houses as faces were accompanied by increases in fusiform gyrus (FG) activity, but that the opposite mistake did not increase PHG responses. In contrast, Serences et al. (2004) found that shifts in object-based attention towards houses recruited parietal and frontal regions to a greater degree than
shifts towards faces. Both these findings and the asymmetry in the results of the present study can be reconciled by assuming that subjects have a dominant prior to respond “face”. This hypothesis is supported by informal debriefing – some subjects in our study commented that they performed the task by responding “house” whenever evidence for a face was scant (see also Summerfield et al. 2006b). In the case of our data, increases in value for houses would lead to top-down shifts in PHG activity to overcome this implicit prior towards responding face, but the converse may not be necessary. Whether visual phenomenology also changes under such shifts in decision criteria is an open question, one that could potentially be addressed by eliciting detailed reports from subjects under biased and unbiased conditions (Jack & Roepstorff 2002).

Responses to categorical uncertainty

Consistent with previous reports, we found that activity in dorsal medial frontal (paracingulate) cortex (dMFC) and anterior insula correlates with categorical uncertainty (Grinband et al. 2006; Preuschoff et al. 2008; Philiastides & Sajda, 2007). There has been recent debate about the functional role of the medial frontal/paracingulate cortex in perceptual decisions (Heekeren et al. 2008). Here we report preliminary evidence for segregation of networks responding to changes in categorical uncertainty and category value. Dorsal paracingulate activity correlated with increases in categorical uncertainty, independent of changes in value; conversely, the frontal eye fields and caudate/thalamus/STN were specifically active during decisions requiring integration of category value information. As uncertainty was correlated with reaction time (RT), we are unable to dissociate the contributions of decision time to these activations (cf. Grinband et al., 2006; 2008). Interestingly however, the dMFC region lies just dorsal to the ACC, which responded to the degree of decision criterion shift across individuals. Given that such shifts are only required when subjects are uncertain about the sensory data (Maddox & Bohil 2003), the close anatomical relationship between these regions may be optimal for integration of categorical uncertainty during shifts in decision criteria.
In contrast, the opposite contrast (examining brain activity that increases for “easy”, certain choices), revealed a cluster in ventromedial prefrontal cortex that was close to that responding to the explicit delivery of monetary reward information. This finding supports recent evidence that the ventromedial prefrontal cortex may signal a perceptual “match” between observed and predicted stimulus information (Summerfield & Koechlin 2008), and suggests that perceptual accuracy itself may act as a reinforcer (Bohil & Maddox 2001). However, in our experiment this signal could also be related to the ongoing assessment of the expected value of the current decision (Boorman et al. 2009; O'Doherty 2004), as both perceptual accuracy and potential rewards are highly correlated on any given trial.

**Costs and priors**

A recent theoretical suggestion holds that cost functions and priors perform a common role in perceptual inference (Friston et al. 2009), relying on backward communication of expectations. This idea has broad historical precedent in psychology, beginning with the "New Look" school of perception in the 1950s (Bruner 1957; Bruner & Goodman 1947). These studies emphasised the role of needs and desires in altering visual perception, and have been echoed in recent studies showing that, for example, being motivated to receive a particular outcome leads to perceptual biases (Balcetis & Dunning 2006). Two elegant studies of changes in categorical priors have demonstrated increased backwards connectivity from frontal cortex to stimulus-selective visual areas as a function of changes in expectation (Summerfield et al. 2006; Summerfield & Koechlin 2008). A similar mechanism may account for our findings: within frontal cortex, the ACC tracked the degree to which cost information was used to bias decisions towards the house category across individuals, whereas a bias towards the house category selectively activated parahippocampal gyrus. Further work is needed to examine the neural overlap between cost- and prior-induced biases in visual decision making. We also note that the manipulation of category value in our study was achieved through the prospect of asymmetric losses, in keeping with the psychometric tradition of using loss functions to manipulate behaviour (Green & Swets 1966; Landy et al.
2007; Whiteley & Sahani 2008). However, given evidence that the brain may process the prospect of losses and gains differently (Liu et al. 2007; Yacubian et al. 2006), it would be useful to establish whether losses and gains differentially impact upon mechanisms for perceptual decision making.

Summary

To conclude, our findings extend previous reports that costs attributed to perceptual decision outcomes have consistent effects on stimulus categorisation, with subjects acting to minimise prospective losses. We show that this effect of cost on perceptual decisions is robustly associated with BOLD signal increases in a frontoparietal network, in keeping with the hypothesis that loss functions enact a particular task set. When the cost for responding “house” decreased, we observed selective activation within the parahippocampal gyrus. Across subjects, greater shifts in decision criteria were associated with greater activation of the medial frontal cortex (ACC). These findings are consistent with the hypothesis that asymmetric costs alter an intermediate representation between perception and action, expressed via general effects on frontal cortex, and selective effects on extrastriate cortex. Returning to our radiologist’s difficult decision, the implicit costs associated with different diagnoses may set up top-down priors that subtly bias sensorimotor dynamics, leading to more false alarms and less misses. Our results constitute a step towards accounting for how prospective losses are flexibly integrated with sensory evidence during effective perceptual decision making.
Acknowledgements We thank Tali Sharot and Karl Friston for comments on an earlier version of this manuscript, and Nicholas Wright for helpful discussions. This work was supported by a Wellcome Trust Programme Grant (to R.J.D.), MRC funding within the UCL 4 Year PhD Programme in Neuroscience (S.M.F.), a Wellcome Trust PhD Studentship (to L.W.) and a Gatsby Charitable Foundation Grant (to M.S.). Address correspondence to Stephen M. Fleming, Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London WC1N 3BG, UK. Email: s.fleming@fil.ion.ucl.ac.uk.
REFERENCES


Balcetis E, Dunning D. See what you want to see: motivational influences on visual

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. Learning the value of

Bohil C, Maddox W. Category discriminability, base-rate, and payoff effects in perceptual

Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS. How green is the grass on the
other side? Frontopolar cortex and the evidence in favor of alternative courses of

Botvinick MM, Braver TS, Barch DM, Carter CS, Cohen JD. Conflict monitoring and

Presented at the 8th International Conference on Functional Mapping of the Human
Brain. 2002.


Bruner J, Goodman C. Value and need as organising factors in perception. J Abnorm Soc

Corbetta M, Patel G, Shulman GL. The reorienting system of the human brain: from

Davison M, Tustin R. The relation between the generalized matching law and signal-

Ekstrom LB, Roelfsema PR, Arsenault JT, Bonmassar G, Vanduffel W. Bottom-up

Ferrera VP, Yanike M, Cassanello C. Frontal eye field neurons signal changes in decision

Fleming SM. Shaping what we see: Pinning down the influence of value on perceptual

Friston KJ, Glaser DE, Henson RNA, Kiebel S, Philips C, Ashburner J. Classical and

Friston KJ, Daunizeau J, Kiebel SJ. Reinforcement learning or active inference? PloS One


Helmholtz H. *Treatise on physiological optics*. Thoemmes Continuum, 1856.


28


Figure 1. Perceptual decision task and example psychophysics data

(a) Experimental procedure. Subjects viewed a cost signal screen informing them of the potential losses for an incorrect face or house categorisation at the start of each trial. They were then asked to categorise an image randomly drawn from the face-house phase continuum as a “face” or “house”. Timings shown are for the fMRI experiment. (b) Illustrative psychophysics data from one subject (LB). Crosses show choice probabilities for each stimulus phase and cost combination; lines show psychometric curves fit to the data.
Figure 2. Behavioural results

(a) Bayesian model comparison was used to show that the best model for the psychophysics data was one with separate mean and slope parameters for each cost condition. The chart shows Laplace approximation to the total log marginal likelihood across subjects and across shared and separate error parameters (it seems possible that attentional lapses would vary with value condition, which does not bear on the hypothesis of interest) - a smaller negative value indicates a better model. Note that each unit difference in log likelihood corresponds to an e-fold ratio of model probabilities. (b, c) Average parameters of the psychometric function fits to the psychophysics data, N = 19 (b), and a corresponding signal detection analysis of the in-scanner data, N = 15 (c). Bars represent the point of subjective equality (PSE)/criterion in FV, NV and HV conditions. White markers indicate the average slope/d' parameter in each value condition for comparison. (d) Mean reaction times (RT) averaged across changes in stimulus information for each cost condition. In all panels, error bars denote SEM; two asterisks (**), P < 0.005; one asterisk (*), P < 0.05 in comparison with NV.
(a) Axial (z = 54) and sagittal (x = -36 and -12) sections showing brain activations reflecting the main effect of asymmetric cost [(FV + HV) > 2NV], averaged over category. Shown are significant clusters in left ventrolateral prefrontal cortex (vlPFC), intraparietal sulcus (IPS), bilateral frontal eye fields (FEF) and subcortical regions (STN – subthalamic nucleus region; Th – thalamus; see also Table 1). Labelled activations are significant at $P < 0.05$, cluster FWE whole-brain corrected. (b) Haemodynamic response time courses aligned to stimulus onset for the three different cost conditions, plotted for the significant cluster in vlPFC.
Figure 4. Stimulus-selective brain activity

(a) Coronal (y = -48) section showing parametric effects of the probability an image was classified as a face in right fusiform gyrus (FG; 39, -48, -24; Z-score = 4.07; $P < 0.05$, small-volume corrected). (b) Coronal (y = -42) section showing parametric effects of the probability an image was classified as a house in bilateral parahippocampal gyrus (PHG; left: -21, -42, -15; Z-score = 3.68; right: 33, -42, -9; Z-score = 5.26; both $P < 0.05$, small-volume corrected). (c, d) Percent signal change as a function of value condition in stimulus-selective ROIs defined from clusters shown in (a) and (b).
Figure 5. Category-specific effects of asymmetric cost

(a) Axial (z = -15) showing the region in left parahippocampal gyrus (red) active during decreased cost (increased value) for houses (HV > FV; -33, -36, -15; Z-score = 3.84; \( P < 0.05 \), small-volume corrected). Shown in blue are clusters selective for house stimulus information (Fig. 4b) for comparison. (b) Intersubject variation in decision criteria, with subjects ordered by their decision criterion in the NV condition. The arrow shows the difference (extent of behavioural shift under HV) used as a covariate for the relevant contrast testing for HV-specific effects of asymmetric value shown in (a). (c) Sagittal (y = 6) and axial (z = 30) sections showing a region in the anterior cingulate (ACC) that shows greater activity in subjects who show greater behavioural shifts in the HV condition (6, 36, 30; Z-score = 4.29; \( P < 0.05 \), whole-brain corrected). (d) Averaged HV > FV beta within the ACC cluster shown in (c) plotted against the criterion shift in the HV condition, across subjects. Inference was carried out using appropriate corrections for multiple comparisons in the SPM framework; this plot is simply provided for illustration purposes.
Figure 6. Effects of categorical uncertainty

(a) Saggital section (y = -2) showing positive (red) and negative (blue) correlations with a regressor indexing categorical uncertainty (see Methods). Dorsal medial frontal cortex (dMFC; 6, 12, 51; Z-score = 3.70; $P < 0.05$, whole-brain corrected) and insula (not shown; 42, 24, -3; Z-score = 4.11; $P < 0.05$, whole-brain corrected) showed positive correlations with uncertainty. A cluster in ventromedial prefrontal cortex (vmPFC) showed increased activity for easier perceptual decisions (6, 36, 30; Z-score = 4.20; $P < 0.05$, whole-brain corrected).

(b) Haemodynamic response timecourses for the three different cost conditions, plotted for the significant cluster in dMFC. While showing strong correlations with the categorical uncertainty regressor, this region was insensitive to changes in category value (cf. Fig. 3b).
Table 1 Summary of significant activations for the COST > NEUTRAL contrast reported in the main text.

<table>
<thead>
<tr>
<th>Label</th>
<th>Voxels at $P &lt; 0.001$</th>
<th>Z-score</th>
<th>$P$ value (cluster FWE corrected)</th>
<th>Peak voxel MNI coordinates</th>
<th>Laterality</th>
</tr>
</thead>
<tbody>
<tr>
<td>IFG (p. opercularis)</td>
<td>106</td>
<td>4.92</td>
<td>$&lt; 0.001$</td>
<td>-36, 3, 24</td>
<td>L</td>
</tr>
<tr>
<td>FEF</td>
<td>36</td>
<td>4.56</td>
<td>0.004</td>
<td>-27, -3, 54</td>
<td>L</td>
</tr>
<tr>
<td>Caudate/thalamus/STN</td>
<td>93</td>
<td>4.28</td>
<td>$&lt; 0.001$</td>
<td>-15, -18, 0</td>
<td>L</td>
</tr>
<tr>
<td>FEF</td>
<td>42</td>
<td>4.10</td>
<td>0.002</td>
<td>26, -9, 54</td>
<td>R</td>
</tr>
<tr>
<td>IPS</td>
<td>49</td>
<td>4.26</td>
<td>0.001</td>
<td>-36, -39, 45</td>
<td>L</td>
</tr>
<tr>
<td>IFG (p. triangularis)</td>
<td>31</td>
<td>3.94</td>
<td>0.009</td>
<td>-45, 21, 0</td>
<td>L</td>
</tr>
<tr>
<td>Insula/Putamen</td>
<td>33</td>
<td>3.90</td>
<td>0.007</td>
<td>36, 15, -6</td>
<td>R</td>
</tr>
<tr>
<td>pMTG</td>
<td>43</td>
<td>4.04</td>
<td>0.001</td>
<td>-36, -72, 21</td>
<td>L</td>
</tr>
</tbody>
</table>

Abbreviations: IFG – inferior frontal gyrus; IPS – intraparietal sulcus; FEF– frontal eye fields; pMTG – posterior middle temporal gyrus; STN – subthalamic nucleus region.
Fig. 1

(a) Time intervals and stimuli:
- 1820ms
- 100-3000ms
- 100ms
- 1625-3625ms

(b) Proportion of “face” responses over Face phase for different conditions:
- FV
- NV
- HV
Fig. 2

(a) Psychophysics

Marginal log-likelihood

-3000 -3500 -4000 -4500

-0.6 -0.4 -0.2 0 0.2 0.4 0.6

d

Criterion

0 0.2 0.4 0.6

-0.6 -0.4 -0.2 0 0.2 0.4 0.6

-0.6 -0.4 -0.2 0 0.2 0.4 0.6

(b) Psychophysics

Average PSE (\(\mu\))

0.3 0.4 0.5 0.6 0.7

3 4 5

Average slope (\(\rho\))

2 3 4 5

(c) fMRI

Average slope (\(\rho\))

2.5 2.0

1.5 2.0

(d) RT(ms)

700 750 800 850

n.s.
Fig. 5

(a) [Image of a brain scan with a cluster highlighted in red and blue.]

(b) Graph showing decision criterion across subjects for FV, NV, and HV conditions. The x-axis represents subject number, and the y-axis represents criterion. Different markers and lines indicate different conditions.

(c) [Image of another brain scan with a different set of clusters highlighted in red and blue.]

(d) Scatter plot showing parameter estimate (HV - FV) against criterion shift in HV condition for each subject.
Fig. 6

(a) An MRI image showing regions in the brain, labeled as dMFC and vmPFC.

(b) A graph showing the % signal change over time (s) for different conditions labeled as FV, NV, and HV.