

# Neurocomputational mechanisms underlying motivated seeing

Yuan Chang Leong<sup>1\*</sup>, Brent L. Hughes<sup>2</sup>, Yiyu Wang<sup>3</sup> and Jamil Zaki<sup>1</sup>

**People tend to believe that their perceptions are veridical representations of the world, but also commonly report perceiving what they want to see or hear. It remains unclear whether this reflects an actual change in what people perceive or merely a bias in their responding. Here we manipulated the percept that participants wanted to see as they performed a visual categorization task. Even though the reward-maximizing strategy was to perform the task accurately, the manipulation biased participants' perceptual judgements. Motivation increased neural activity selective for the motivationally relevant category, indicating a bias in participants' neural representation of the presented image. Using a drift diffusion model, we decomposed motivated seeing into response and perceptual components. Response bias was associated with anticipatory activity in the nucleus accumbens, whereas perceptual bias tracked category-selective neural activity. Our results provide a computational description of how the drive for reward leads to inaccurate representations of the world.**

People tend to think of their perception as a veridical representation of the external world, but this view has long been challenged by psychological research<sup>1,2</sup>. Instead, people often report percepts that they are motivated to perceive, a phenomenon we term motivated perception. In one classic example in the visual domain, Dartmouth and Princeton students watched the same football game. Fans of each team subsequently reported seeing the other team commit more fouls<sup>3</sup>. Likewise, participants presented with ambiguous line drawings were more likely to report seeing the interpretation associated with desirable outcomes<sup>4</sup>.

One interpretation of these findings is that motivational factors, such as desires and wants, exert top-down influence over perceptual processing, such that people become biased towards seeing what they want to see<sup>5</sup>. We refer to the bias in perceptual processing as a perceptual bias. Alternatively, these effects could instead reflect a response bias: a bias not in what participants see, but merely in what they report seeing<sup>6,7</sup>. Although these two interpretations appear at odds with each other, they are not mutually exclusive; motivation could simultaneously bias both perception and responses. Computational models offer a promising analytical approach to dissociate these two sources of bias and identify their independent contributions to perceptual judgements.

Drift diffusion models (DDMs) assume that perceptual judgements arise from the accumulation of noisy sensory evidence towards one of two decision thresholds<sup>8,9</sup>. When the level of evidence exceeds the threshold associated with a particular percept, the corresponding response is made. Within this framework, a response bias can be modelled as a bias in the starting point of evidence accumulation. This reduces the amount of evidence needed to make a response, but assumes no effect on perceptual processing. Conversely, a perceptual bias can be modelled as a bias in the rate of evidence accumulation. This in turn reflects sensory information accumulating faster for one percept than for another, implying that perceptual processes are biased towards seeing that percept. The extent to which each bias influences behaviour can then be estimated from empirical data.

Neuroimaging offers a second, complementary approach for dissociating response from perceptual biases. The neural mechanisms underlying motivational effects on perceptual judgements are not well understood, but separate literatures on the neuroscience of motivation and perception suggest distinct neural processes that could be related to different components of bias. In particular, both functional magnetic resonance imaging (fMRI) and electrophysiology studies have identified the nucleus accumbens (NAcc) as a key structure in mediating motivational processes<sup>10,11</sup>. One putative role of the NAcc is that it biases response selection in favour of actions associated with higher reward<sup>12–14</sup>. Notably, greater NAcc activity precedes approach behaviour to reward-predicting stimuli<sup>15</sup>, whereas inactivating the NAcc reduces the preference for responses associated with larger rewards<sup>16</sup>. Thus, we predict that the NAcc would play a role in response biases by increasing the readiness to make motivationally desirable judgements.

Conversely, previous work suggests that perceptual judgements are determined by comparing the activity of neurons selective to different perceptual features<sup>17,18</sup>. For example, monkeys in a direction-of-motion task were more likely to categorize a cloud of dots as moving upward when activity was higher in sensory neurons preferring upward motion than in sensory neurons preferring downward motion<sup>19</sup>. Similarly, Heekeren and colleagues demonstrated in humans that perceptual judgements on a face/scene categorization task were computed by comparing activity in areas in the ventral temporal cortex selective to each category<sup>20</sup>. Motivation could potentially bias this comparison process by driving attention towards the features associated with a motivationally desirable percept<sup>21</sup>. This enhances the neural response to those features, thus giving rise to a perceptual bias.

The goal of the present study was twofold: (1) to decompose motivational influences on perceptual judgements into a response bias and a perceptual bias, and (2) to examine the neurocomputational mechanisms underlying motivational biases on perceptual judgements. To do this, we presented human participants visually ambiguous images created by morphing a face image and a scene image together, and rewarded them for correctly categorizing

<sup>1</sup>Department of Psychology, Stanford University, Stanford, CA, USA. <sup>2</sup>Department of Psychology, University of California, Riverside, CA, USA.

<sup>3</sup>Department of Psychology, Northeastern University, Boston, MA, USA. \*e-mail: [ytleong@stanford.edu](mailto:ytleong@stanford.edu)

whether the face or scene was of higher intensity. We manipulated participants' motivation by instructing them on each trial that they would win or lose extra money if the upcoming stimulus was of a particular category. Crucially, participants would gain or lose this additional money based only on the actual category of the stimulus, not what they reported seeing. As such, even though participants were motivated to see one category over the other, they would earn the most money on the task if they reported the stimulus category accurately.

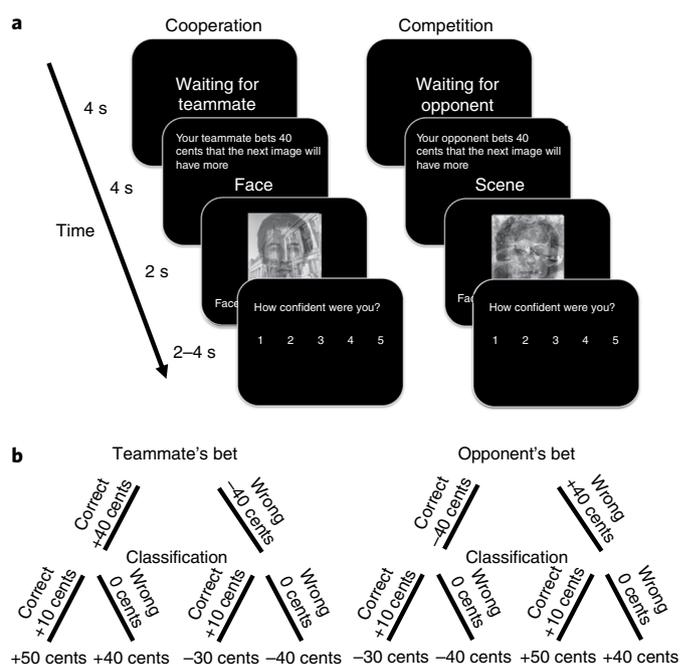
We estimated the magnitude of response and perceptual biases exhibited by our participants by fitting a DDM to choice and reaction time data. Using fMRI, we searched for distinct neural processes associated with each bias. Furthermore, as the perception of faces and scenes is associated with distinct patterns of activity in the ventral occipito-temporal cortex<sup>20,22,23</sup>, we used multivoxel pattern analysis to measure the level of face-selective and scene-selective activity as a correlate of perception. If the motivation to see one category increases the level of neural activity selective for that category, it would provide additional evidence that motivation modulates perceptual processing. By combining the neural measures with computational modelling, our approach provides a mechanistic account of motivational influences on perceptual judgements.

## Results

Thirty participants were scanned using fMRI while they performed a categorization task with visually ambiguous images comprising a mixture of a face and a scene (Fig. 1a). For each image, participants were rewarded for correctly indicating which category was of higher intensity (that is, 'more face' or 'more scene'). To motivate participants to see one category over another, we informed them that they would be performing the task with a teammate or an opponent. This other 'player' would bet on whether the upcoming image would be an image with more face or more scene. Participants were told that neither the teammate nor opponent had seen the upcoming image and their bets provided no informational value. Participants won a monetary bonus if the teammate's bet was correct, and lost money if the teammate's bet was wrong (cooperation condition). By contrast, participants lost money if the opponent's bet was correct, and won a bonus if the opponent's bet was wrong (competition condition). The competition condition allowed us to assess the effect of motivation above and beyond that of semantic priming due to having seen the words 'face' and 'scene'. Crucially, the outcome of the teammate's and opponent's bets were determined by the objective face/scene proportion of the presented image, and not by participants' subjective categorizations. To earn the most money, participants should ignore the bets and make their categorizations accurately (Fig. 1b). Additional details on participants' demographics and the experimental task are reported in the Methods section: see 'Participants' and 'Experimental task', respectively.

**Motivation biases visual categorization.** For each condition, we estimated the psychometric function describing the relationship between participants' categorizations and the relative proportions of face and scene in an image. Not surprisingly, as the proportion of scene in an image increases, participants were more likely to categorize the image as having more scene (generalized linear mixed-effects model (GLME):  $z = 16.9$ ,  $P < 0.001$ ,  $b = 2.19$ , 95% CI = 1.93–2.44).

To examine the effect of motivation, we estimated separate psychometric functions depending on the teammate's or opponent's bet (Fig. 2a). In the cooperation condition, participants were more likely to report seeing more scene when the teammate bet on scene than when the teammate bet on face (GLME:  $z = 2.52$ ,  $P = 0.012$ ,  $b = 0.33$ , 95% CI = 0.07–0.59); that is, participants were more likely to report seeing the category that the bet motivated them to see.

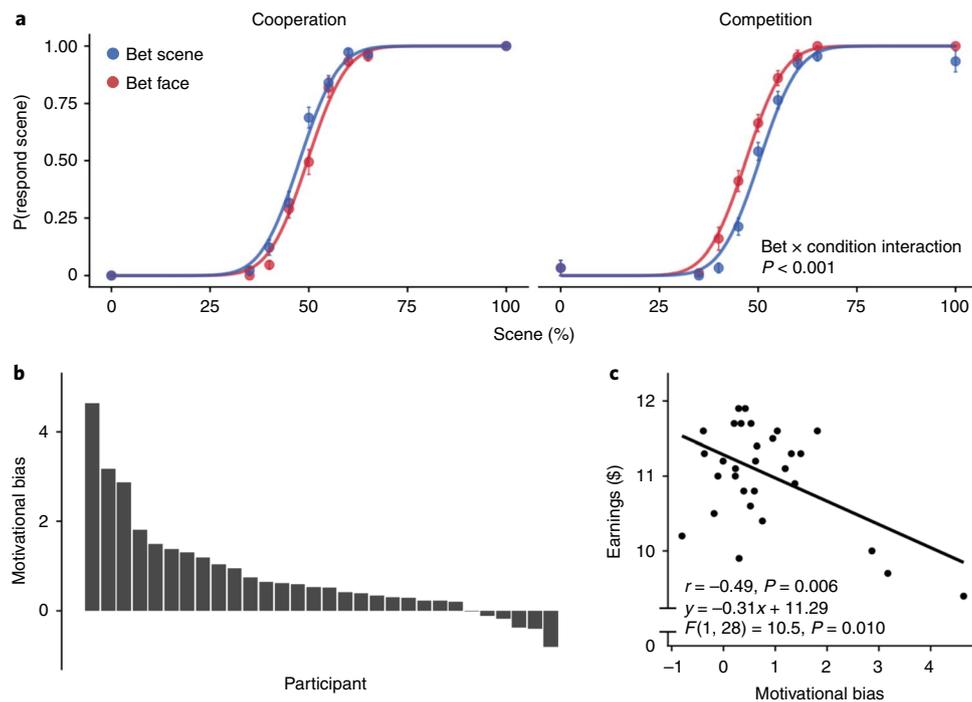


**Fig. 1 | Experimental design.** **a**, Participants were presented with composite face–scene images. In the cooperation condition, a teammate first makes a bet on whether the face or scene will be of higher intensity (that is, more face or more scene). Participants are then presented with the composite image and have to categorize whether it comprises mostly face or mostly scene. They then rated how confident they were in their categorization. In the competition condition, an opponent makes the bet instead. Face images were adapted with permission from ref. 55, Springer US; images licensed by CC BY 4.0. **b**, Payoff structure. Participants won an extra 40 cents if the teammate's bet was correct, but lost 40 cents if the teammate's bet was wrong. Conversely, they lost 40 cents if the opponent's bet was correct, but won 40 cents if the opponent's bet was wrong. Participants earned 10 cents for each correct categorization. As the outcome of the bets was determined by the objective face-to-scene ratio of the presented image and not by participants' subjective categorizations, the reward maximizing strategy was to ignore the bets and perform the categorizations accurately.

The bias in participants' perceptual judgements could also be due to semantic priming. For example, when the teammate bet that the upcoming image would have more face, participants might be more likely to report seeing more face because they were semantically primed by having just seen the word 'face', and not because they were motivated to see more face. The competition condition allows us to directly test this competing account.

In the competition condition, participants were motivated to see the category that was inconsistent with the opponent's bet. For example, if the opponent bet that the upcoming image would have more scene, participants would be motivated to see more face. If the bias in participants' judgements resulted from semantic priming, participants would instead be more likely to report seeing the category consistent with the opponent's bet. Consistent with a motivational account, participants were less likely to categorize an image as having more scene when the opponent bet scene than when the opponent bet face (GLME:  $z = -4.11$ ,  $P = 0.012$ ,  $b = -0.47$ , 95% CI = -0.69 to -0.24). These results also highlight the flexible nature of the motivational bias, as participants were able to remap the relationship between the word presented to them and the percept they were motivated to see based on the experimental context.

To quantify the magnitude of motivational bias across the two conditions, we computed the condition  $\times$  bet interaction on



**Fig. 2 | Motivation biases visual categorization.** All panels include data from 30 participants. **a**, Participants were more likely to categorize the ambiguous image as what they wanted to see. In the cooperation condition, participants' psychometric function was shifted left when the teammate bet on more scene relative to when the teammate bet on more face, indicating that less scene evidence was needed to categorize an image as having more scene. In the competition condition, participants' psychometric function was shifted right when the opponent bet on more scene relative to when the opponent bet on more face, indicating that more scene evidence was needed to categorize an image as having more scene. Statistical significance was assessed using GLME (see 'Psychometric functions' in Methods). Error bars indicate s.e.m. **b**, The magnitude of bias in each participant, defined as the individual effect of the bet  $\times$  condition interaction. Higher values indicate stronger bias. **c**, Participants with greater motivational bias performed worse on the task and received lower earnings. See Supplementary Fig. 2 for replication with an independent group of participants.

participants' categorizations. This interaction was significant (GLME:  $z = 3.35, P < 0.001, b = 0.81, 95\% \text{ CI} = 0.34\text{--}1.28$ ), such that participants were more likely to make categorizations consistent with the teammate's bet and inconsistent with the opponent's bet. Taken together, these results indicate that participants' categorizations were biased by what they were motivated to see.

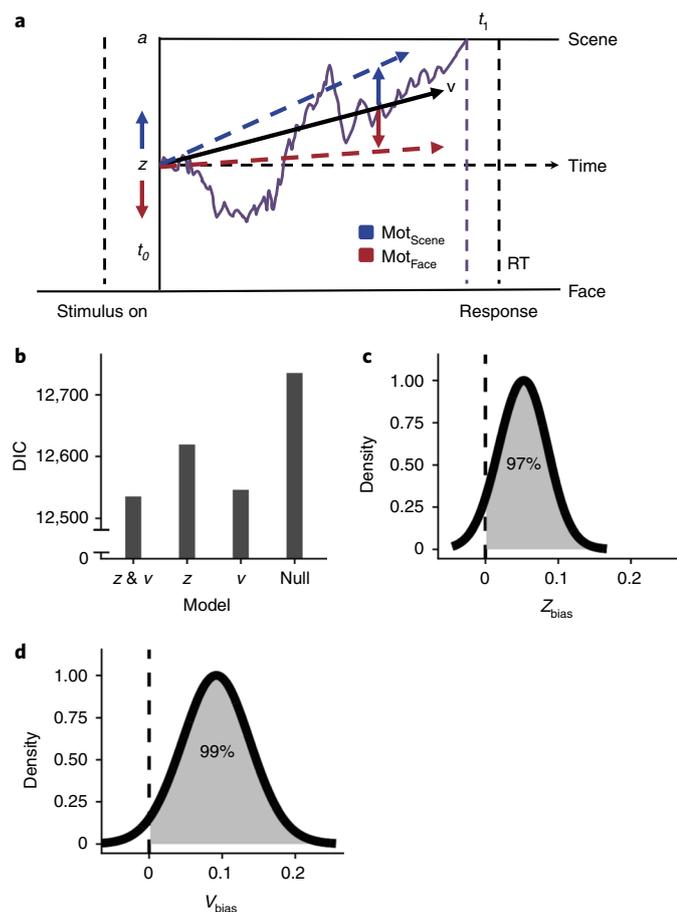
We estimated each participant's motivational bias as the individual effect of the condition  $\times$  bet interaction. Although the majority of participants exhibited motivational bias, the degree of bias varied across individuals (Fig. 2b). Participants who exhibited stronger motivational bias made fewer correct categorizations, indicating that the motivational bias impaired performance on the task and led to decreased earnings (Pearson's  $r = -0.49, P = 0.006$ ; robust regression:  $F_{(1,28)} = 10.5, P = 0.010, b = -0.31, 95\% \text{ CI} = -0.49 \text{ to } -0.12$ ; Fig. 2c). Additional analyses suggest that the performance impairment was not due to general differences in attention or perceptual sensitivity, but rather because some participants were more affected by the motivation manipulation than others (Supplementary Notes 1 and 2 and Supplementary Fig. 1). All behavioural findings were replicated in a separate group of 28 participants who performed the task without undergoing fMRI (Supplementary Fig. 2).

**Motivation biases both starting point and drift rate.** Having established that participants' categorizations were biased by what they wanted to see, we proceeded to examine how motivation biased the decision process. To this end, we fit a DDM to participants' choice and reaction time data. The DDM is a model of the cognitive processes involved in two-choice decisions<sup>9</sup>, and assumes that choice results from the accumulation of noisy sensory evidence towards one of two decision thresholds. The starting point of the accumulation

process is determined by a free parameter,  $z$ , and the decision threshold is determined by a free parameter,  $a$ . The rate of evidence accumulation is determined by the drift rate,  $\nu$ , which depends on the sensory information on each trial. In the case of our task, an image with a high scene proportion would be associated with a highly positive  $\nu$ , whereas an image with a high face proportion would be associated with a highly negative  $\nu$ . When the accumulation process reaches one of the two thresholds (the top threshold for scene and the bottom threshold for face), a response is initiated.

From a DDM perspective, our participants' motivational bias could reflect either or both of two mechanisms (Fig. 3a). First, a shift in the starting point,  $z$ , could result in an a priori bias to make motivationally consistent judgements. In particular, shifting the starting point towards the decision threshold of the motivationally consistent category reduces the amount of evidence needed to make the motivationally consistent response, thus creating a response bias. Second, a bias in the drift rate,  $\nu$ , could favour evidence accumulation in favour of the motivationally consistent category. This results in sensory evidence accumulating faster for the motivationally consistent category, thus creating a perceptual bias. Both biases increase the proportion of motivationally consistent judgements, but have distinguishable effects on the shape of reaction time distributions<sup>9,24</sup> (Supplementary Fig. 3).

To examine whether either or both of these processes explained the bias observed in our task, we fit three different DDMs to participants' data<sup>25</sup> (see 'DDM' in Methods): (1) a model in which motivation biases the starting point ( $z$  model), (2) a model in which motivation biases the drift rate ( $\nu$  model), (3) and a model in which motivation biases both the starting point and drift rate ( $z$  &  $\nu$  model). For comparison, we also fit an unbiased model in which neither



**Fig. 3 | Modeling results.** **a**, Schematic diagram of the DDM. On each trial, choice depends on the accumulation of noisy sensory evidence towards one of two decision thresholds. Motivation biases categorizations by modulating both the starting point and the drift rate.  $t_0$  and  $t_1$  refer to the non-decision time related to stimulus encoding and response execution, respectively.  $a$ , decision threshold;  $z$ , starting point;  $v$ , drift rate; purple line, decision variable (i.e., accumulated evidence); solid black line, average drift;  $\text{Mot}_{\text{Face}}$ , motivated to see more face;  $\text{Mot}_{\text{Scene}}$ , motivated to see more scene; RT, reaction time. **b**, The  $z$  &  $v$  model has the lowest DIC score, indicating that it provides the best fit to participants' data. See Supplementary Table 1 for best-fit values and summary statistics of all model parameters. **c**, Posterior distribution of the starting point bias ( $z_{\text{bias}}$ ). The dashed line indicates 0 (no bias). More than 95% of the distribution was greater than 0, indicating strong evidence of a starting point bias. **d**, Posterior distribution of the drift bias ( $v_{\text{bias}}$ ). More than 95% of the distribution was greater than 0, indicating strong evidence of a drift bias.

the starting point nor the drift rate was biased by motivation (null model). As the pattern of reaction times was comparable across the cooperation and competition conditions (Supplementary Fig. 4), all models were fit to data pooled over both conditions to take advantage of the larger number of trials for more reliable estimates.

We compared the model fits based on the deviance information criterion<sup>26</sup> (DIC; a common metric of model comparison for hierarchical models that penalizes for model complexity, with lower values indicating better fit). To verify that DIC is an accurate metric for model comparison, we fit the models to simulated data and demonstrated that DIC reliably identifies the true model used to generate each dataset (Supplementary Note 3 and Supplementary Fig. 5). When the models were fit to experimental data, DIC identified the  $z$  &  $v$  model as the model that provided the best fit to

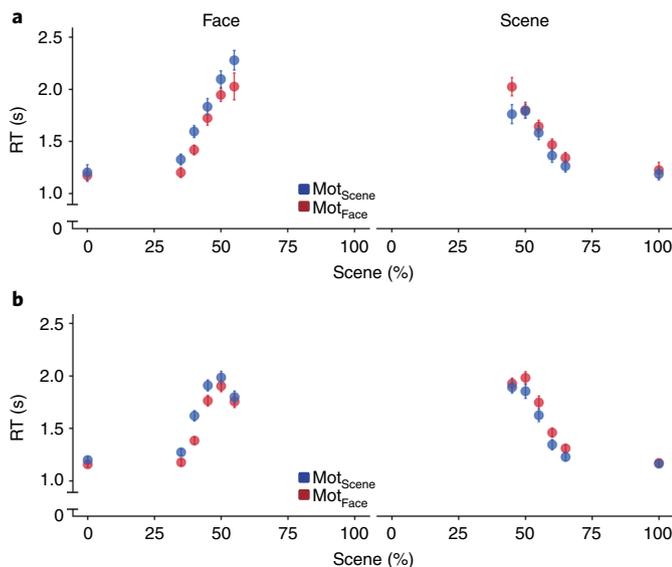
participants' data (DIC:  $z$  &  $v$ : 12,535;  $z$ : 12,619;  $v$ : 12,546, null: 12,735; Fig. 3b), suggesting that motivation biased both the starting point and the drift rate of evidence accumulation. The best-fit values and summary statistics of all model parameters are reported in Supplementary Table 1.

Next, we examined how the starting point and the drift rate were affected by motivation. We extracted the posterior distribution of the starting point bias estimated by the  $z$  &  $v$  model (Fig. 3c). Positive values indicate a positive motivational bias, such that the starting point is biased towards the scene threshold when participants were motivated to see more scene, and biased towards the face threshold when participants were motivated to see more face. If more than 95% of the posterior distribution were to be greater than 0, we consider it to be strong evidence that the estimate was positive; 96.8% of the posterior distribution of the starting point bias was greater than 0 ( $P(z_{\text{bias}} > 0) = 0.968$ , mean = 0.051, 95% credible interval =  $-0.003$  to  $0.105$ ), indicating strong evidence that motivation biased the starting point of evidence accumulation towards the threshold of the motivationally consistent category.

Similarly, there was strong evidence that the bias in the drift rate was greater than 0 ( $P(v_{\text{bias}} > 0) = 0.991$ , mean = 0.092, 95% credible interval =  $0.015$ – $0.168$ ; Fig. 3d), indicating that evidence accumulation was biased towards the face category when participants were motivated to see more face, and towards the scene category when participants were motivated to see more scene. Taken together, the modelling results suggest that motivation biased perceptual judgements by increasing the predisposition to respond in a motivationally consistent manner, as well as by biasing sensory processing in favour of the motivationally consistent category.

**DDM accounts for asymmetries in reaction time.** Participants were faster at making categorizations that were consistent with their motivations (linear mixed-effects model (LME): two-tailed, one-sampled  $t_{(28)} = -2.80$ ,  $P = 0.009$ ,  $b = -0.05$ , 95% CI =  $-0.98$  to  $-0.02$ ). In particular, participants were faster to categorize an image as a face when motivated to see more face (LME: two-tailed, one-sampled  $t_{(28)} = -3.12$ ,  $P = 0.004$ ,  $b = -0.07$ , 95% CI =  $-0.11$  to  $-0.02$ ), and faster to categorize an image as a scene when motivated to see more scene (LME: two-tailed, one-sampled  $t_{(28)} = -2.11$ ,  $P = 0.04$ ,  $b = -0.04$ , 95% CI =  $-0.09$  to  $-0.01$ ; Fig. 4a). We examined whether the  $z$  &  $v$  model would account for this feature of the data. We simulated choice and reaction time data using the  $z$  &  $v$  model (see 'Model simulations' in Methods), and showed that the simulated data reproduced the pattern of reaction times where motivationally consistent responses were faster than motivationally inconsistent responses (LME: two-tailed, one-sampled  $t_{(28)} = -3.65$ ,  $P = 0.001$ ,  $b = -0.04$ , 95% CI =  $-0.06$  to  $-0.02$ ; Fig. 4b). This was true for both face responses (LME: two-tailed, one-sampled  $t_{(28)} = -3.79$ ,  $P < 0.001$ ,  $b = -0.05$ , 95% CI =  $-0.07$  to  $-0.02$ ) and scene responses (LME: two-tailed, one-sampled  $t_{(29)} = -2.86$ ,  $P = 0.007$ ,  $b = -0.03$ , 95% CI =  $-0.06$  to  $-0.01$ ). The model simulations also predict a non-monotonic relationship between mean reaction time and percentage scene, with a turning point at 50% scene. This pattern was not observed in the data (Fig. 4a,b). At present, it is unclear whether this reflects a deviation from the model or an unreliable estimate from the data. As our experiment did not sufficiently and evenly sample the space of percentage scene to accurately estimate the relationship between percentage scene and mean reaction times, we do not explore this further.

We compared the model simulations from the  $z$  &  $v$  model to that of the other models, and show that the  $z$  &  $v$  model provides a closer match to the empirical data than the other models (Supplementary Fig. 6 and Supplementary Note 4). Finally, we show that the model simulations from the  $z$  &  $v$  model reproduced each participant's choice and reaction time distributions (Supplementary Fig. 7). Altogether, these results indicate that model fits of the  $z$  &  $v$  model align well with participants' data.



**Fig. 4 | DDM accounts for asymmetries in reaction times.** **a**, Reaction times for face (left) and scene (right) responses at each percentage scene, separately for when participants were motivated to see faces and scenes. Participants were faster to categorize an image as the category that they were motivated to see. Trial types with fewer than 48 trials (that is, 1% of the total number of trials) were excluded from the plot because there were too few trials for reliable estimates and they tend to come from a small number of participants. Error bars indicate between-subjects s.e.m. **b**, Model-predicted reaction times. Reaction times were simulated using the  $z$  &  $v$  model with parameter values sampled from the posterior distribution. Five hundred datasets were simulated, each with the same number of participants and trials as the original data in **a**. Reaction times were first averaged over simulations to obtain an average reaction time for each trial type for each participant, and then averaged over participants to obtain the mean reaction time for each type. Error bars denote between-subjects s.e.m.

**Motivationally consistent categorizations are associated with activity in the salience network and dorsal attention network.** To identify the brain areas associated with motivational biases in perceptual judgements, we first performed a whole-brain contrast to identify voxels that responded differently on trials on which participants categorized an image as the category that they were motivated to see (motivation consistent trials) than on trials on which they categorized an image as the category that they were motivated to not see (motivation inconsistent trials). This contrast revealed activations in two networks of brain regions: (1) the salience network, which includes the NAcc, the insula, the dorsal anterior cingulate (dACC), and (2) the dorsal attention network, including the intraparietal sulcus (IPS) and frontal eye fields (FEFs) (Fig. 5).

**NAcc activation is associated with response bias.** The NAcc is thought to be crucial in mediating the effects of motivation on actions and has been previously implicated in biasing responses towards actions associated with larger rewards<sup>12–14</sup>. Thus, we predicted that the NAcc would be associated with the response bias when making motivated perceptual judgements. We defined a NAcc region of interest (ROI) using the Harvard–Oxford Cortical Structural Atlas, and computed the NAcc response of each participant as the average  $z$ -statistic of the motivation consistent > motivation inconsistent contrast in the ROI (Fig. 5). This value reflects the extent to which the NAcc of a particular participant was more active when they made motivationally consistent categorizations than when they made motivationally inconsistent categorizations.

We then examined whether the NAcc response would be associated with either participants' response bias, their perceptual bias or both biases.

For each participant, we computed response bias as the posterior mean estimate of that participant's starting point bias ( $z_{\text{bias}}$ ; Fig. 3c), and perceptual bias as the posterior mean estimate of their drift bias ( $v_{\text{bias}}$ ; Fig. 3d). Estimates of the two biases were not significantly correlated (Pearson  $r=0.290$ ,  $P=0.120$ ; robust regression:  $F_{(1,28)}=1.58$ ,  $P=0.219$ ,  $b=0.58$ , 95% CI =  $-0.33$  to  $1.49$ ). When both biases were entered as predictors in the same linear regression model, participants' response bias was associated with the NAcc response (linear regression: two-tailed, one-sampled  $t_{(27)}=2.71$ ,  $P=0.012$ ,  $\beta_z=0.47$ , 95% CI =  $0.12$ – $0.83$ ), but their perceptual bias was not (linear regression: two-tailed, one-sampled  $t_{(27)}=0.37$ ,  $P=0.711$ ,  $\beta_v=0.07$ , 95% CI =  $-0.29$  to  $0.42$ ; Fig. 6a).

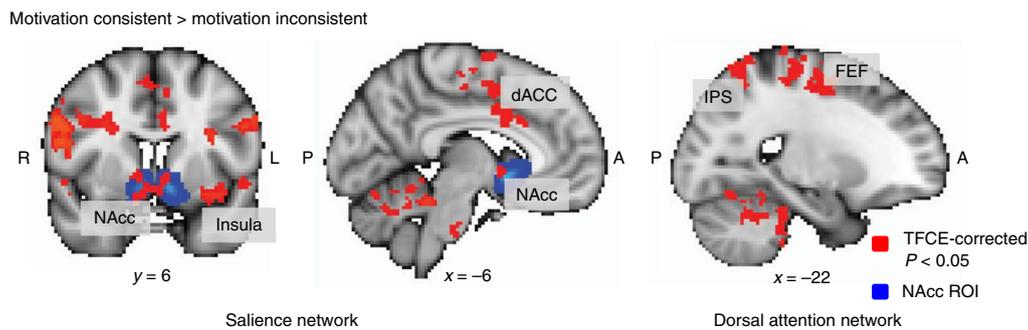
NAcc activity can lead to a response bias by increasing the readiness to make a particular response. This account would predict that the increase in NAcc activity was preparatory in nature and would precede the onset of the image. To test this hypothesis, we examined the average activity in the NAcc as a trial unfolded, separately for trials on which participants made motivationally consistent categorizations (motivation consistent trials) and for trials on which participants made motivationally inconsistent categorizations (motivation inconsistent trials). Consistent with a preparatory account, NAcc activity was significantly higher on motivation consistent trials prior to the image appearing on the screen and remained significantly higher until image offset (Fig. 6b,c and Supplementary Table 2).

Notably, the results also provide evidence against the alternative account that NAcc activation reflects the reward participants' experience on seeing the category that they were motivated to see, as the increase in NAcc activity occurred before participants saw the image. Instead, the results suggest that NAcc activity predisposes participants to categorize an image as the category that they were motivated to see, and sustained NAcc activation increases the likelihood of making motivationally consistent responses.

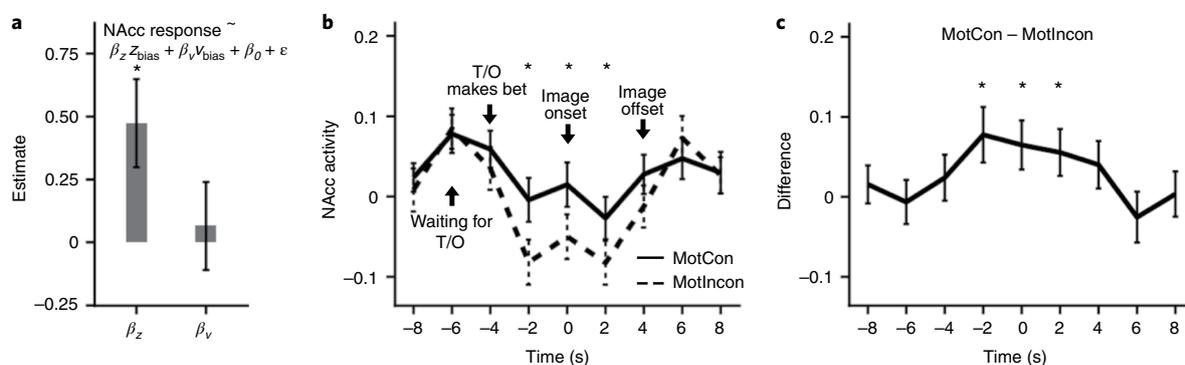
One implication of these results is that trial-by-trial NAcc activity might be related to trial-by-trial variability in the starting point, such that the starting point is more biased towards the motivationally consistent category when NAcc activity is higher. As we did not explicitly model the trial-by-trial variability in the model parameters, we cannot test this hypothesis in the current study. We opted not to model the trial-by-trial variability in model parameters because doing so can compromise the reliability and accuracy of the estimates of the individual parameters, especially when the number of trials per participant is small<sup>27</sup>. Nevertheless, we believe that this is an interesting question that warrants further investigation in a future study with a larger dataset.

**Face-selective and scene-selective neural activity is associated with perceptual bias.** Face-selective and scene-selective activity in the ventral occipito-temporal cortex provides a proxy measure of participants' perception. Thus, we examined whether motivation affected perception by assessing whether the motivation to see faces or scenes modulated this activity. We applied multivariate pattern analysis to the blood-oxygen-level-dependent (BOLD) data to quantify the level of face-selective and scene-selective activity on each trial. Specifically, we trained a logistic regression classifier to estimate the probability that participants were seeing a scene rather than a face based on the pattern of activity in the ventral occipito-temporal cortex (see 'Multivoxel pattern analyses' in Methods).

As the proportion of scene in an image increased, the classifier predicted that the participants were seeing a scene with higher probability, indicating that the classifier tracked the amount of scene in the presented image (LME: two-tailed, one-sampled  $t_{(4,756)}=25.8$ ,



**Fig. 5 | Neural correlates of motivational bias.** Activity in the salience network and the dorsal attention network was higher when participants made motivationally consistent categorizations than when participants made motivationally inconsistent categorizations. Correction for multiple comparisons was performed using threshold-free cluster enhancement (TFCE) with an alpha level of 0.05 (ref. 62). The NAcc ROI was defined using the Harvard-Oxford Cortical Structural Atlas. A, anterior; L, left; P, posterior; R, right.



**Fig. 6 | NAcc activation is associated with response bias.** **a**, Linear regression predicting participants' NAcc response from their starting point bias ( $z_{\text{bias}}$ ) and drift bias ( $v_{\text{bias}}$ ). The regression coefficient for  $z_{\text{bias}}$  was significant, but that for  $v_{\text{bias}}$  was not. Error bars indicate s.e. **b**, NAcc time course time-locked to image onset, corrected for haemodynamic lag by shifting the BOLD data by 4 s. The trial starts with the 'waiting for teammate or opponent (T/O)' screen at  $-6$  s. The teammate or opponent makes a bet at  $-4$  s, which remains on the screen for 4 s. The image is presented at 0 s and stays on the screen for 4 s. NAcc activity was significantly higher on motivation-consistent trials (MotCon) than on motivation-inconsistent trials (MotIncon) from 2 s before image onset until image offset. **c**, Difference in activity between motivation-consistent and motivation-inconsistent trials peaked before image offset. Two-tailed, one-sampled  $t$ -tests,  $*P < 0.05$ ; the full statistics are reported in Supplementary Table 2. Error bars indicate between-subjects s.e.m.

$P < 0.001$ ,  $b = 0.12$ , 95% CI = 0.11–0.13). There was a significant bet  $\times$  condition interaction on classifier probability, such that the classifier was more likely to predict that participants were seeing a scene when they were motivated to see a scene than when they were motivated to see a face (LME: two-tailed, one-sampled  $t_{(4,756)} = 2.05$ ,  $P = 0.040$ ,  $b = 0.038$ , 95% CI = 0.002–0.075; Supplementary Fig. 8), indicating that the motivation to see a category increased the level of sensory evidence for that category in the visual pathway. In other words, motivation not only biased participants' categorization of an image but it also biased their neural representation of the image.

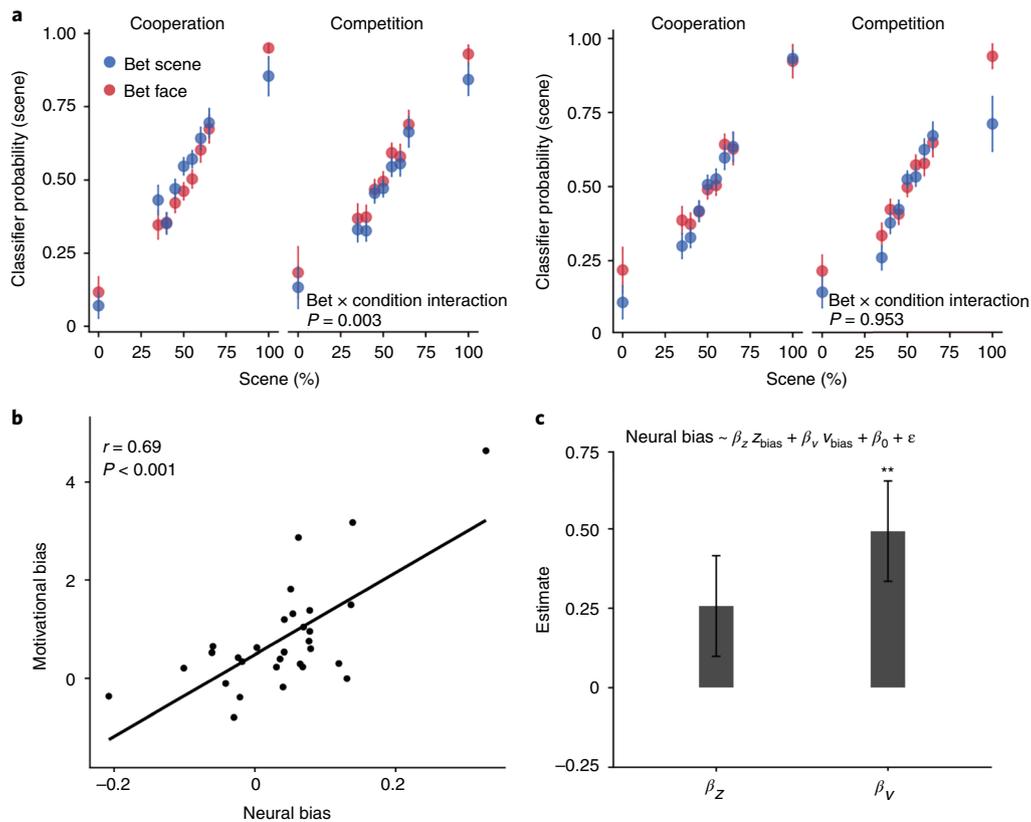
Next, we examined how the bias in category-selective activity relates to the bias in participants' categorical judgements (that is, the motivational effect on a participant's psychometric function). There was a significant triple interaction between behavioural bias, condition and bet on the level of category-selective activity (LME: two-tailed, one-sampled  $t_{(4,752)} = 3.37$ ,  $P = 0.001$ ,  $b = 0.06$ , 95% CI = 0.02–0.09). To better interpret the directionality of the interaction, we performed a median split to divide participants into those with higher behavioural bias and those with lower behavioural bias. Motivation biased the classifier probability of high-bias participants (LME: two-tailed, one-sampled  $t_{(2,378)} = 2.96$ ,  $P = 0.003$ ,  $b = 0.07$ , 95% CI = 0.03–0.13), but not of low-bias participants (LME: two-tailed, one-sampled  $t_{(2,373)} = -0.06$ ,  $P = 0.953$ ,  $b = -0.002$ , 95% CI =  $-0.053$

to 0.050; interaction by group: two-tailed, one-sampled  $t_{(4,752)} = 2.14$ ,  $P = 0.033$ ,  $b = 0.08$ , 95% CI = 0.02–0.14; Fig. 7a).

We then extracted each participant's individual effect of the bet  $\times$  condition interaction on classifier probability to obtain a measure of the extent to which motivation biased face-selective and scene-selective activity in the participant. The bias in participant's face-selective and scene-selective activity correlated strongly with their behavioural bias (Pearson  $r = 0.69$ ,  $P < 0.001$ ; robust regression:  $F_{(1,28)} = 15.1$ ,  $P < 0.001$ ,  $b = 6.94$ , 95% CI = 3.73–10.15; Fig. 7b), indicating that participants who were more biased in their categorizations were also more biased in their neural representation of the presented image.

We then sought to relate the bias in face-selective and scene-selective activity to response and perceptual biases more specifically. When model estimates of the two biases were entered as predictors in the same linear regression model, participants' perceptual bias was associated with the bias in face-selective and scene-selective neural activity (linear regression: two-tailed, one-sampled  $t_{(27)} = 3.12$ ,  $P = 0.004$ ,  $\beta_v = 0.49$ , 95% CI = 0.17–0.82), but their response bias was not (linear regression: two-tailed, one-sampled  $t_{(27)} = 1.63$ ,  $P = 0.115$ ,  $\beta_z = 0.26$ , 95% CI =  $-0.07$  to 0.58; Fig. 7c).

Together with our earlier analyses on NAcc activity, these results suggest distinct neural contributions to participants' biased categorizations. While the NAcc was associated with a response bias, the



**Fig. 7 | Motivation biases face-selective and scene-selective neural activity during visual categorization. a**, Classifier probability that the presented image was a scene rather than a face, based on the BOLD response in the ventral visual stream, separately for participants with high (left) and low (right) behavioural bias. For high-bias participants, the scene probability was higher when participants were motivated to see a scene (that is, a teammate bets scene or an opponent bets face) than when participants were motivated to see a face (that is, a teammate bets face or an opponent bets scene). There was no effect of motivation in low-bias participants. Error bars denote between-subjects s.e.m. **b**, The effect of motivation on classifier probability (neural bias) was correlated with the extent to which a participant was biased in their categorizations (motivational bias). **c**, Regression coefficients of the response bias ( $\beta_z$ ) and perceptual bias ( $\beta_v$ ) when both were entered into the same model to predict participants' neural bias. Only the perceptual bias was significantly associated with participants' neural bias. Two-tailed, one-sampled  $t_{(2,378)} = 2.96$ ,  $**P = 0.003$ ,  $b = 0.07$ , 95% CI = 0.03–0.13. Error bars denote s.e.

modulation of face-selective and scene-selective activity in visual areas was associated with a perceptual bias. By combining computational modelling with neuroimaging, we identified two dissociable neurocomputational components underlying motivational biases in perceptual judgements.

## Discussion

This study combines computational modelling of behaviour and fMRI to examine whether and why people exhibit biases towards seeing what they want to see. We developed a behavioural paradigm that allowed us to find that people indeed make biased perceptual judgements, more often labelling ambiguous images as corresponding to a reward-associated category. This is true even though participants were incentivized to accurately report their perceptual experience, and thus earned less money in the experiment when making biased judgements. Evidence from computational modelling suggests that motivational effects on perceptual judgements could be attributed to both a response bias and a bias in perceptual processing. While the response bias was associated with anticipatory activity in the NAcc, the bias in perceptual processing was associated with the modulation of category-selective neural activity in the ventral visual stream. These results provide converging evidence for two distinct contributions to motivational influences on perceptual judgements and shed light on the neurocomputational mechanisms underlying each bias.

The claim that perceptual processes are influenced by motivational factors can be traced back to the 'New Look' movement in psychology, which argued that the perception of external stimuli is subject to the constant influence of a perceiver's internal goals and states<sup>28,29</sup>. Recent evidence supporting this view includes studies demonstrating that perceptually ambiguous stimuli are more likely to be seen as the percept associated with favourable outcomes<sup>4,30</sup>, desirable objects are judged nearer than undesirable objects<sup>31</sup>, and desirable food items are judged as larger by dieters than non-dieters<sup>32</sup>. Whether these results reflect a bias in subjective reports or a bias in perception remains a topic of intense debate (see the open peer commentary for ref. 7). In particular, as these studies rely primarily on subjective reports, and participants often have an incentive to report seeing what they want to see, there is reason to suspect that subjective reports might not reflect one's underlying perceptual experience.

Our work builds on earlier efforts to understand motivated perception using computational models. An earlier study<sup>33</sup> similarly decomposed motivational biases into response and perceptual components by mapping them respectively onto biases in the starting point and drift rate of a DDM. However, there are non-perceptual explanations for a bias in drift rate. For example, a response bias that becomes stronger over the course of a trial (that is, a dynamic response bias) could also result in a biased drift rate<sup>34,35</sup>. As such, a bias in the drift rate alone is insufficient evidence of a bias in perception.

A strength of the current work is that, in addition to modelling the behavioural data using a DDM, we used functional neuroimaging to measure the level of sensory evidence in visual areas of the brain. We found that motivation enhances the sensory evidence of the motivationally consistent category in the ventral visual stream, and the degree of sensory enhancement was associated with the bias in drift rate but not with the bias in starting point. In relating the neural data to the modelling results, we achieve two complementary goals: (1) the neural results provide convergent validity of our interpretation of the bias in drift rate as a perceptual bias, and (2) the modelling results provide a computational description of how the bias in sensory activity contributes to biases in perceptual judgements.

Perceptual judgements are thought to be computed by comparing the activity of neurons selective to different perceptual features<sup>19,20</sup>. Within this framework, the nervous system ‘reads out’ the activity of face-selective and scene-selective neurons as sensory evidence for faces and scenes, respectively. A perceptual judgement can then be determined by comparing the activity of face-selective and scene-selective neurons. Our results indicate that motivation biases this comparison by enhancing the activity of the neurons selective to the category that participants were motivated to see. This enhancement could in turn reflect the biased processing of incoming sensory information, with the biasing signal originating from frontoparietal attention regions<sup>36</sup>.

Indeed, we found that the IPS and FEFs were more active when participants made motivationally consistent judgements. The IPS and FEFs are part of the dorsal attention network associated with the top-down control of attention<sup>37,38</sup>. Their involvement in our task suggests that the bias in perceptual processing might be in part mediated by dynamic changes in the focus of attention<sup>39</sup>. In addition to the frontoparietal activations, the dACC and the insula were also more active on motivation consistent trials. The dACC and the insula are part of a salience network involved in the detection of motivationally salient stimuli<sup>40,41</sup>, and the dACC has been recently implicated in determining what stimulus feature to attend to in a perceptual decision-making task<sup>42</sup>. The increased activity in the salience network on motivation consistent trials might be responsible for the selection of motivationally relevant features for enhanced processing. However, this interpretation is speculative, and future studies will be needed to clarify the role of each region in biasing perceptual judgements.

Conversely, participants’ response bias was associated with activity in the NAcc. This is consistent with behavioural neuroscience work suggesting that dopaminergic projections to the NAcc biases animals towards responses associated with greater reward<sup>12–14,16</sup>. Both human neuroimaging and animal physiology studies have also shown that the NAcc is activated in anticipation of reward<sup>10,11</sup>. Our results suggest a functional role for this anticipatory activity. In particular, they suggest that the NAcc increases participants’ readiness to respond in a motivation consistent manner. When the motivation consistent response is aligned with task demands (for example, pressing a lever for reward), this preparatory response facilitates faster responding for reward<sup>15,43</sup>. However, when the motivationally consistent response conflicts with task demands, as was the case in our task, the preparatory response is maladaptive and impairs performance on the task.

Our results add to the rich literature on perceptual decision-making in cognitive neuroscience by dissociating motivation (‘wanting to see’) from optimal task performance (‘reward maximization’). Previous studies have manipulated the reward associated with different perceptual alternatives and found that reward biases responses but not perceptual processing<sup>44–47</sup>. We speculate that these results differ markedly from ours because our paradigm tapped into distinct biasing mechanisms. In these earlier studies, participants would earn the additional reward if they correctly categorized the stimulus as the rewarded category. Under this payoff scheme,

biasing responses towards the option associated with larger reward would result in greater cumulative reward over the course of the experiment<sup>44,48</sup>. Hence, the bias in these earlier experiments probably reflects a strategic shift in responses to maximize reward on the task, and thus not affect sensory processing.

By contrast, in our task, the additional reward associated with the motivationally consistent category was independent of participants’ responses. For example, if the teammate bet that the next image would have more face, participants would receive the bonus if the upcoming image indeed had more face, regardless of how they responded on the trial. In this case, a bias towards the motivationally consistent category would lower participants’ earnings by hurting their accuracy on the categorization task. Thus, the biases observed in our task cannot be explained by existing normative models of judgement and decision-making that assume organisms adjust their choice strategies to maximize expected reward. Instead, they highlight a motivational component to perceptual judgements—wanting an outcome to be true can impinge on one’s perceptual judgement, even when doing so could lead to lower rewards in the long run. Our results suggest that this bias reflects not only a response bias but also a perceptual bias.

At a broader level, this work provides a novel bridge between social psychology and cognitive neuroscience. Using tools and analytical techniques from cognitive neuroscience, we examine the neurocomputational mechanisms underlying an age-old phenomenon of interest in social psychology. In doing so, we offer a fresh perspective on a classic debate. Unlike earlier work that assesses whether motivation biases perception, we provide a neurocomputational account of how motivation biases perception. The results also complement the existing literature on motivated person perception, which has focused primarily on the neural and computational mechanisms by which people form overly positive evaluations of themselves and close others<sup>49–51</sup>. Our work extends the phenomenon beyond the domain of social attributions, and show that motivated visual perception can be similarly characterized as a change in initial beliefs (that is, starting point) and information updating (that is, drift rate).

Desires and wants exert a powerful influence over how people make sense of the world. Recent studies have examined the neural mechanisms underlying motivational biases across various human reasoning and evaluative processes<sup>52</sup>, including how the brain learns more from positive outcomes than from negative outcomes<sup>53</sup> and why people form unrealistically optimistic expectations about future events<sup>54</sup>. Here, we demonstrate that motivation biases human cognition as early as visual perception, and provide a neurocomputational account of this effect. The current work extends our understanding of motivational biases and provides a starting point to explore how motivation acts on different neural systems at different stages of information processing to influence human cognition.

## Methods

**Participants.** Thirty-three participants were recruited from the Stanford community and provided written, informed consent before the start of the study. All experimental procedures were approved by the Stanford University Institutional Review Board. Participants were paid between US\$30 and \$50 depending on their performance on the task. Data from 3 participants were discarded because of excessive head motion (>3 mm) during  $\geq 1$  scanning sessions, yielding an effective sample size of 30 participants (17 male, 13 female, 18–43 years of age, mean age = 22.3 years).

**Stimuli.** For each participant, seven sets (one for the practice task and six for the experimental task) of composite stimuli were created. Each stimulus set consists of 40 greyscale images, each comprising a mixture of a face image and a scene image in varying proportions (1 × 100% scene, 3 × 65% scene, 5 × 60% scene, 7 × 55% scene, 8 × 50% scene, 7 × 45% scene, 5 × 40% scene, 3 × 35% scene and 1 × 0% scene). Scene images comprised half indoor scenes and half outdoor scenes, whereas face images comprised half male faces and half female faces. All faces were frontal photographs posing a neutral expression, and were taken from the Chicago

Face Database<sup>55</sup>. Stimuli were presented using MATLAB software (MathWorks) and the Psychophysics Toolbox<sup>56</sup>.

**Practice task.** Participants first performed 40 practice trials in which they were presented with composite face/scene images (see ‘Stimuli’). Each image was presented for 4 s, during which participants had to judge whether the image contained a greater proportion of face (more face) or a greater proportion of scene (more scene). Participants earned 10 cents for each correct categorization. They then indicated how confident they were in their classification on a 1–5 scale. If they did not respond within 4 s, the trial timed out and they would not earn a bonus on that trial. After a variable inter-trial interval (ITI; 2–4 s), they moved on to the next trial. We collected participants’ anatomical scans while they performed the practice task.

**Experimental task.** The experimental task consists of four fMRI runs, each approximately 8-min long. Participants performed two runs of the cooperation condition and two runs of the competition condition (interleaved order, counterbalanced across participants; Fig. 1a). Each run consisted of 40 trials. In the cooperation condition, participants were told that they would perform a visual categorization task with a teammate. At the start of each trial, their teammate would make a bet on the image type of the upcoming image (more face or more scene, presented for 4 s). Participants were then presented with a composite image created by averaging a face image and a scene image in different proportions (see ‘Stimuli’). If the teammate’s bet was correct, both the teammate and the participant would earn 40 cents. If the teammate’s bet was wrong, both the teammate and the participant would lose 40 cents.

Participants then had 4 s to make a categorization on whether the image contained more face or more scene. Participants earned 10 cents for each correct categorization. They then indicated how confident they were in their classification on a 1–5 scale. If they did not respond within 4 s, the trial timed out and they would not earn a bonus on that trial (although the bet would still be implemented). After a variable ITI (2–4 s), they moved on to the next trial. In the competition condition, participants performed the task with an opponent. The trial structure was identical to the cooperation condition, except that if the opponent’s bet was correct, the opponent would earn 40 cents, whereas participants would lose 40 cents. If the opponent’s bet was wrong, the opponent loses 40 cents, whereas participants would earn 40 cents. As such, participants were motivated to see the image type their teammate bet on and to see the image type opposite of what their opponent bet on. Participants were told that the teammate and the opponent would not be informed about their responses, and would thus not be motivated by a desire to appease or defy their teammate or opponent.

Crucially, the outcome of the bets was contingent on whether the image objectively contained more face or more scene, and was not contingent on participants’ subjective categorization. Hence, the reward-maximizing strategy was to ignore the bets and categorize the images as accurately as possible (Fig. 1b). Bets by both the teammate and the opponent were pseudo-randomized such that they were accurate on exactly 50% of the trials. As such, participants’ earnings in the experiment depended solely on their performance on the categorization task. We computed participants’ performance as the average number of correct categorizations.

**Localizer task.** To identify BOLD activation associated with viewing faces or scenes, we had participants perform a localizer task at the end of the experiment. Participants viewed 5 blocks of 15 unambiguous faces and 5 blocks of 15 unambiguous scenes (blocks were interleaved, and the order was counterbalanced across participants). In the face blocks, participants were sequentially presented with face images and had to indicate whether each face was male or female. In the scene blocks, participants were sequentially presented with scene images and had to indicate whether each scene was indoors or outdoors. Each image was presented for 2 s, with a 2-s ITI. Participants took a self-timed break between blocks. The localizer task was split into two scans.

**fMRI data acquisition and preprocessing.** MRI data were collected using a 3 T General Electric MRI scanner. Functional images were acquired in interleaved order using a T2\*-weighted echo planar imaging pulse sequence (46 transverse slices, repetition time = 2 s, echo time = 25 ms, flip angle = 77° and voxel size = 2.9 mm<sup>3</sup>). Anatomical images were acquired at the start of the session with a T1-weighted pulse sequence (repetition time = 7.2 ms, echo time = 2.8 ms, flip angle = 12° and voxel size = 1 mm<sup>3</sup>). Image volumes were preprocessed using FSL/FEAT v.5.98 (FMRIB software library, FMRIB, Oxford, UK). Preprocessing included motion correction, slice-timing correction, removal of low-frequency drifts using a temporal high-pass filter (100-ms cut-off) and spatial smoothing (4-mm full-width at half-maximum). For multivoxel classification analyses, we trained and tested our classifier in each participant’s native space. For all other analyses, functional volumes were first registered to participants’ anatomical image (using boundary-based registration) and then to a template brain in the Montreal Neurological Institute (MNI) space (affine transformation with 12 d.f.).

**Psychometric functions.** We modelled participants’ behavioural data using GLME, which allows for the modelling of all of the data in one step rather than

fitting a separate model for each participant<sup>57</sup>. Three separate GLME were fit to participants’ data to estimate the effect of the motivation manipulation (that is, the ‘bet’ by the teammate or opponent) on participants’ categorizations. The first two models (M1 and M2; see Supplementary Table 3 for full model specification) were fit to data from the cooperation and competition condition to estimate the effect of the teammate’s and opponent’s bet, respectively, while the third model (M3) was fit to data from both conditions to estimate the bet × condition interaction. All models included the percentage scene in the image as a covariate. All models also included random intercepts and random slopes for the effects of the teammate’s or opponent’s bet (scene/face) to account for random variability across participants. The third model also included random intercepts and slopes for the condition and the condition × bet interaction. To obtain a measure of individual participants’ motivational bias, we added the estimate of each participant’s random effect on the condition × bet interaction to the estimate of the corresponding fixed effect. This individual effect reflects the extent to which a participant’s categorizations were biased by the motivation manipulation. Models were estimated using the `glmer` function in the `lme4` package in R<sup>58</sup>, with *P* values computed from *t*-tests with Satterthwaite approximation for the degrees of freedom as implemented in the `lmerTest` package<sup>59</sup>.

**Reaction time analyses.** We ran a series of LME to examine the effect of motivation on the pattern of reaction times. Reaction times were log-transformed before being entered into the model. We first examined whether reaction times were faster when making motivationally consistent responses than when making motivationally inconsistent responses, controlling for the absolute difference between percentage scene and percentage face as a measure of stimulus uncertainty (M4; see Supplementary Table 4 for full model specification). Next, we examined motivational effects on reaction times separately for trials in which participants responded face (M5) and those in which they responded scene (M6). For each group, we tested whether the reaction times differed depending on the category that participants were motivated to see, again controlling for stimulus uncertainty. To visualize these results, we plot the average reaction time for face and scene responses at each level of percentage scene, separately for when participants were motivated to see more face and when participants were motivated to see more scene. We then repeated the analyses with simulated reaction time data generated from the *z* & *v* model (M7–M9).

**Robust regression analysis.** We ran robust regression models to supplement our correlational tests. Robust regression has been shown to be less sensitive to outliers<sup>60</sup>. We fit three linear models by robust regression. The first model regressed participants’ earnings on their motivational bias (Fig. 2c), the second model regressed participants’ drift bias on their starting point bias, and the third model regressed participants’ motivational bias on their neural bias (Fig. 7b). Model fitting was performed using the `rlm` function from the ‘MASS’ package in R. Statistical significance of the regression coefficient was assessed by performing a robust *F*-test.

**DDM.** The DDM assumes that decisions are made by accumulating evidence over time until it crosses one of two decision bounds<sup>9</sup> (Fig. 3a). The starting point and the rate of evidence accumulation were determined by the free parameters *z* and *v*, respectively. The distance between the two boundaries depended on the free parameter *a*, while time not related to decision process (for example, stimulus encoding or motor response) was modelled by the free parameter *t*.

Model parameters were estimated from participants’ categorizations and reaction time distributions using hierarchical Bayesian estimation as implemented by the HDDM toolbox<sup>25</sup>. Parameters for individual participants were assumed to be randomly drawn from a group-level distribution. In the fitting procedure, each participant’s parameters both contributed to and were constrained by the estimates of group-level parameters. Markov chain Monte Carlo sampling methods were used to estimate the joint posterior distribution of all model parameters (100,000 samples; burn-in = 10,000 samples; thinning = 2). We estimated both group-level parameters as well as parameters for each individual participant, which allowed us to examine biases in both the entire sample and in each individual participant. To account for outliers generated by a process other than that assumed by the model (for example, lapses in attention and accidental button press), we estimated a mixture model in which 5% of trials were assumed to be distributed according to a uniform distribution.

In the *z* & *v* model, we modelled *z* as a function of the motivationally consistent category and an intercept term. The HDDM package implements *z* as the relative starting point, bound between 0 and 1, with 0.5 reflecting an unbiased starting point. As such, we used the inverse logit link function to restrict *z* to values between 0 and 1:

$$z = \frac{1}{1 + \exp(-(\beta_{z1} \text{Motivation} + \beta_{z0}))}$$

where Motivation denotes the motivationally consistent category, and was coded as +1 when participants were motivated to see more scene, and –1 when participants were motivated to see more face. Thus,  $\beta_{z1}$  reflects the effect of motivation on the starting point ( $z_{\text{bias}}$ ), while  $\beta_{z0}$  denotes an intercept term.

We modelled  $v$  as a linear combination of the category that participants were motivated to see, the level of percentage scene and an intercept term:

$$v = \beta_{v1} \text{Motivation} + \beta_{v2} \% \text{scene} + \beta_{v0}$$

where Motivation was again coded as +1 (motivated to see more scene) and -1 (motivated to see more face).  $\beta_{v1}$  reflects the effect of motivation on the drift rate ( $v_{\text{bias}}$ ) and  $\beta_{v2}$  reflects the effect of percentage scene on the drift rate. We demeaned percentage scene before entering it into the model such that the intercept term,  $\beta_{v0}$ , would also reflect the intrinsic drift bias. For each of the bias parameters ( $z_{\text{bias}}$  and  $v_{\text{bias}}$ ), we computed the proportion of posterior samples that were greater than 0 (Fig. 3c,d). If more than 95% of the posterior distribution were to be greater than 0, we consider it to be strong evidence that the estimate was positive.

To examine whether either of the biases were sufficient for explaining the data, we fit two additional comparison models in which only  $z$  ( $z$  model) or only  $v$  ( $v$  model) was biased by motivation. In the  $z$  model,  $\beta_{z1}$  was fixed at 0, and in the  $v$  model,  $\beta_{v1}$  was fixed to 0. As a baseline for comparison, we also fit a null model in which neither the starting point nor the drift rate was biased by motivation. We then compared the four models using DIC, which is a measure of model performance that appropriately penalizes for model complexity in hierarchical models<sup>26</sup>. To verify that DIC is an accurate metric for model comparison, we also ran a model recovery study to examine whether DIC correctly recovers the model used to generate simulated data (Supplementary Note 3)

Model convergence of all models was formally assessed using the Gelman–Rubin  $\hat{R}$  statistic<sup>61</sup>, which runs multiple Markov chains to compare within-chain and between-chain variances. Large differences ( $\hat{R} > 1.1$ ) between these variances would signal non-convergence. In addition, we examined each trace to check that there were no drifts or large jumps, which would also suggest non-convergence. We report model convergence metrics, posterior means and 95% credible intervals of all parameters in Supplementary Table 1.

**Model simulations.** We generated 500 simulated datasets, each comprising the same number of participants performing the same number of trials as the real dataset. Each dataset was generated with parameter values sampled from the posterior distribution estimated by the  $z$  &  $v$  model. These datasets reflect the pattern of choice and reaction time data if participant's behaviour was perfectly described by the model. To compare the simulations to real data, we averaged over the 500 simulations to obtain the average reaction time of face and scene responses at each level of percentage scene, separately for when participants were motivated to see more face and when participants were motivated to see more scene.

Given that the DDM was fitted to reaction time distributions (rather than the mean), we also assessed how well the model accounts for the shape of participants' reaction time distributions. For each participant, we overlay the distribution of simulated reaction times with the true reaction time distributions, separately for face and scene responses. These plots serve as posterior predictive checks to assess how well the model aligns with participants' data (Supplementary Fig. 7).

**GLM.** We implemented a linear model (GLM 1) to contrast BOLD activity on motivation consistent trials and that on motivation inconsistent trials. A motivation consistent trial was defined as a trial on which participants categorized an image as the category they were motivated to see. Thus, this contrast would identify voxels in the brain that were significantly more active when participants reported seeing what they wanted to see, versus what they did not want to see. Stimulus onset, the objective percentage scene, reaction time and head movement parameters were included as nuisance regressors. With the exception of head movement parameters, all regressors were convolved with a haemodynamic response function. The GLM was estimated throughout the whole brain using FSL/FEAT v.5.98, which is available as part of the FMRIB software library. Correction for multiple comparisons was performed using threshold-free cluster enhancement with an alpha level of 0.05, as implemented by the randomize tool in FSL<sup>62</sup>.

We implemented a second linear model (GLM 2) in which the onset of each trial was modelled as a separate regressor. This model allowed us to estimate a separate statistical map for each trial (that is, single-trial activation patterns). We then used these maps as inputs to the multivoxel pattern analyses (see below). As was the case in GLM 1, reaction time and head movement parameters were included as nuisance regressors.

**NAcc ROI analyses.** We defined an independent NAcc ROI using the Harvard–Oxford subcortical structural atlas (available for download at <https://neurovault.org/collections/EAAXGDRJ/>). For each participant, we extracted the average  $z$ -statistic of the motivation consistent > motivation inconsistent contrast (GLM 1) within the NAcc ROI. This average  $z$ -statistic reflects the extent to which an ROI is more reliably active on motivation consistent trials than on motivation inconsistent trials for that participant, and was taken as the participant's NAcc response. The NAcc response was then regressed against estimates of starting point and drift bias estimated by the  $z$  &  $v$  model (see 'Relating model parameters to behaviour and neural measures').

To examine how the NAcc response unfolded across a trial, we extracted and  $z$ -scored the mean time course in the NAcc ROI from each run. Each time course was shifted by two repetition times (4 s) to correct for haemodynamic lag. We extracted the data from 8 s before stimulus onset to 8 s after stimulus onset to obtain the time course of a single trial, and computed the average time course of activity separately for motivation consistent trials and motivation inconsistent trials. At each time point, we assessed whether activity was higher on motivation consistent trials than on motivation inconsistent trials using a paired sample  $t$ -test.

**Multivoxel pattern analyses.** Multivoxel pattern analyses were performed using tools available as part of the Nilearn Python module<sup>63</sup>. An L1-regularized logistic regression model ( $C=1$ ) was trained on BOLD data from the localizer task to classify the image category that participants were seeing on each localizer trial. Notably, the localizer task was unrelated to face versus house discrimination (see 'Localizer task'). Hence, the category-selective patterns identified by the classifier were specific to seeing faces or scenes, and not confounded by response-related information.

Analysis was restricted to voxels in a ventral visual stream mask consisting of the bilateral occipital lobe and ventral temporal cortex. The ventral occipito-temporal regions of the brain are thought to be important in perceiving object categories such as faces and scenes<sup>22</sup>. The mask was created in MNI space using anatomical masks defined by the Harvard–Oxford Cortical Structural Atlas. The mask was then transformed into each participant's native space using FSL's FLIRT implementation, and classification was performed in participants' native space.

The trained model was then applied to the single-trial activation patterns in the experimental task (GLM 2). On each trial, the classifier returned the probability that the participant was seeing a scene rather than a face based on activity in the ventral visual stream mask. We then modelled classifier probability on each trial using a LME with the percentage scene of an image, the task condition (cooperation/competition), the teammate's or opponent's bet (face/scene) and the interaction between condition and bet as predictor variables (M10; see Supplementary Table 5 for full model specification). The models included random intercepts and random slopes for each of the predictor variables to account for the random variability across participants. We computed each participant's individual effect on the condition  $\times$  bet interaction by adding the estimate of a participant's random effect to the estimate of the fixed effect. This individual effect reflects the extent to which classifier probability was biased by the motivation manipulation for that particular participant and was taken as a measure of neural bias.

To examine the relationship between motivation, participants' behavioural bias and classifier probability, we ran a second model to test for the triple interaction of condition, bet and behavioural bias on classifier probability (M11). To better understand the directionality of this interaction, we performed a median split on participants' behavioural bias, and tested for the condition  $\times$  bet interaction on classifier probability separately for high-bias (M12) and low-bias (M13) participants. Finally, we tested for the triple interaction of group (high bias or low bias)  $\times$  condition  $\times$  bet (M14) on classifier probability.

**Relating model parameters to behaviour and neural measures.** We used linear regression to examine the relationship between model parameters and neural activity. We entered participant-level estimates of the starting point bias ( $z_{\text{bias}}$ ) and the drift bias ( $v_{\text{bias}}$ ) as predictor variables in regression models. The first model was used to predict participants' NAcc response to the motivation consistent–motivation inconsistent contrast (GLM 1), and assessed the extent to which each bias was associated with NAcc activity. The second model was used to predict participants' neural bias and assessed the extent to which each bias contributed to the modulation of category-selective activity in the ventral visual stream.

**Statement on statistics and reproducibility.** All statistical tests were two-tailed and used an alpha level of 0.05. Data distribution was assumed to be normal but this was not formally tested. Sample size was determined to be comparable to other fMRI studies of motivational biases<sup>44</sup> and perceptual decision-making<sup>45</sup>. The study employed a within-subject design and participants were not assigned to different experimental conditions.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The data that support the findings of this study are available from the corresponding author on request. Behavioural data of both the reported experiment and the in-lab replication are available at: <https://github.com/yicleong/MotivatedPerception>. The unthresholded  $p$ -map of the motivation consistent–motivation inconsistent contrast is available at: <https://neurovault.org/collections/EAAXGDRJ/images/62743/>.

## Code availability

The custom code for the modelling and neuroimaging analyses is included in the Supplementary Software. The live version of the code is available at <https://github.com/yicleong/MotivatedPerception>.

Received: 10 July 2018; Accepted: 20 May 2019;  
Published online: 1 July 2019

## References

- Bruner, J. S. & Goodman, C. C. Value and need as organizing factors in perception. *J. Abnorm. Soc. Psychol.* **42**, 33–44 (1947).
- Dunning, D. & Balcells, E. Wishful seeing: how preferences shape visual perception. *Curr. Dir. Psychol. Sci.* **22**, 33–37 (2013).
- Hastorf, A. H. & Cantril, H. They saw a game; a case study. *J. Abnorm. Soc. Psychol.* **49**, 129–134 (1954).
- Balcells, E. & Dunning, D. See what you want to see: motivational influences on visual perception. *J. Pers. Soc. Psychol.* **91**, 612–625 (2006).
- Kunda, Z. The case for motivated reasoning. *Psychol. Bull.* **108**, 480–498 (1990).
- Goldiamond, I. Indicators of perception: I. Subliminal perception, subception, unconscious perception: an analysis in terms of psychophysical indicator methodology. *Psychol. Bull.* **55**, 373–411 (1958).
- Firestone, C. & Scholl, B. J. Cognition does not affect perception: evaluating the evidence for ‘top-down’ effects. *Behav. Brain Sci.* **39**, e229 (2016).
- Forstmann, B. U., Ratcliff, R. & Wagenmakers, E.-J. Sequential sampling models in cognitive neuroscience: advantages, applications, and extensions. *Annu. Rev. Psychol.* **67**, 641–666 (2016).
- Ratcliff, R. & McKoon, G. The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* **20**, 873–922 (2008).
- Berridge, K. C. The debate over dopamine’s role in reward: the case for incentive salience. *Psychopharmacology* **191**, 391–431 (2007).
- Knutson, B., Adams, C. M., Fong, G. W. & Hommer, D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J. Neurosci.* **21**, RC159 (2001).
- Floresco, S. B. The nucleus accumbens: an interface between cognition, emotion, and action. *Annu. Rev. Psychol.* **66**, 25–52 (2015).
- Ikemoto, S. & Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Res. Rev.* **31**, 6–41 (1999).
- Nicola, S. M. The nucleus accumbens as part of a basal ganglia action selection circuit. *Psychopharmacology* **191**, 521–550 (2007).
- McGinty, V. B., Lardeux, S., Taha, S. A., Kim, J. J. & Nicola, S. M. Invigoration of reward seeking by cue and proximity encoding in the nucleus accumbens. *Neuron* **78**, 910–922 (2013).
- Stopper, C. M. & Floresco, S. B. Contributions of the nucleus accumbens and its subregions to different aspects of risk-based decision making. *Cogn. Affect. Behav. Neurosci.* **11**, 97–112 (2011).
- Gold, J. I. & Shadlen, M. N. The neural basis of decision making. *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
- Heekeren, H. R., Marrett, S. & Ungerleider, L. G. The neural systems that mediate human perceptual decision making. *Nat. Rev. Neurosci.* **9**, 467–479 (2008).
- Shadlen, M. N., Britten, K. H., Newsome, W. T. & Movshon, J. A. A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J. Neurosci.* **16**, 1486–1510 (1996).
- Heekeren, H. R., Marrett, S., Bandettini, P. A. & Ungerleider, L. G. A general mechanism for perceptual decision-making in the human brain. *Nature* **431**, 859–862 (2004).
- Summerfield, C. & Egner, T. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* **13**, 403–409 (2009).
- Grill-Spector, K. The neural basis of object perception. *Curr. Opin. Neurobiol.* **13**, 159–166 (2003).
- Hasson, U., Hendler, T., Bashat, D. B. & Malach, R. Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *J. Cogn. Neurosci.* **13**, 744–753 (2001).
- White, C. N. & Poldrack, R. A. Decomposing bias in different types of simple decisions. *J. Exp. Psychol. Learn. Mem. Cogn.* **40**, 385–398 (2014).
- Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in Python. *Front. Neuroinform.* **7**, 14 (2013).
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64**, 583–639 (2002).
- Boehm, U. et al. Estimating across-trial variability parameters of the diffusion decision model: expert advice and recommendations. *J. Math. Psychol.* **87**, 46–75 (2018).
- Allport, F. H. *Theories of Perception and the Concept of Structure: A Review and Critical Analysis with an Introduction to a Dynamic-Structural Theory of Behavior* (John Wiley & Sons, 1955).
- Bruner, J. S. On perceptual readiness. *Psychol. Rev.* **64**, 123–152 (1957).
- Balcells, E., Dunning, D. & Granot, Y. Subjective value determines initial dominance in binocular rivalry. *J. Exp. Soc. Psychol.* **48**, 122–129 (2012).
- Balcells, E. & Dunning, D. Wishful seeing: more desired objects are seen as closer. *Psychol. Sci.* **21**, 147–152 (2010).
- van Koningsbruggen, G. M., Stroebe, W. & Aarts, H. Through the eyes of dieters: biased size perception of food following tempting food primes. *J. Exp. Soc. Psychol.* **47**, 293–299 (2011).
- Voss, A., Rothermund, K. & Brandtstädter, J. Interpreting ambiguous stimuli: separating perceptual and judgmental biases. *J. Exp. Soc. Psychol.* **44**, 1048–1056 (2008).
- Moran, R. Optimal decision making in heterogeneous and biased environments. *Psychon. Bull. Rev.* **22**, 38–53 (2015).
- Hanks, T. D., Mazurek, M. E., Kiani, R., Hopp, E. & Shadlen, M. N. Elapsed decision time affects the weighting of prior probability in a perceptual decision task. *J. Neurosci.* **31**, 6339–6352 (2011).
- Serences, J. T. Value-based modulations in human visual cortex. *Neuron* **60**, 1169–1181 (2008).
- Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).
- Petersen, S. E. & Posner, M. I. The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* **35**, 73–89 (2012).
- Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V. & Niv, Y. Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron* **93**, 451–463 (2017).
- Menon, V. in *Brain Mapping* (ed. Toga, A. W.) 597–611 (Academic Press, 2015).
- Sridharan, D., Levitin, D. J. & Menon, V. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc. Natl Acad. Sci. USA* **105**, 12569–12574 (2008).
- Shenhav, A., Straccia, M. A., Musslick, S., Cohen, J. D. & Botvinick, M. M. Dissociable neural mechanisms track evidence accumulation for selection of attention versus action. *Nat. Commun.* **9**, 2485 (2018).
- Niv, Y., Daw, N. D., Joel, D. & Dayan, P. Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* **191**, 507–520 (2007).
- Feng, S., Holmes, P., Rorie, A. & Newsome, W. T. Can monkeys choose optimally when faced with noisy stimuli and unequal rewards? *PLoS Comput. Biol.* **5**, e1000284 (2009).
- Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W. & Forstmann, B. U. Bias in the brain: a diffusion model analysis of prior probability and potential payoff. *J. Neurosci.* **32**, 2335–2343 (2012).
- Rorie, A. E., Gao, J., McClelland, J. L. & Newsome, W. T. Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (LIP) of the macaque monkey. *PLoS One* **5**, e9308 (2010).
- Summerfield, C. & Koechlin, E. Economic value biases uncertain perceptual choices in the parietal and prefrontal cortices. *Front. Hum. Neurosci.* **4**, 208 (2010).
- Bogacz, R., Brown, E., Moehlis, J., Holmes, P. & Cohen, J. D. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol. Rev.* **113**, 700–765 (2006).
- Flagan, T., Mumford, J. A. & Beer, J. S. How do you see me? The neural basis of motivated meta-perception. *J. Cogn. Neurosci.* **29**, 1908–1917 (2017).
- Hughes, B. L. & Beer, J. S. Orbitofrontal cortex and anterior cingulate cortex are modulated by motivated social cognition. *Cereb. Cortex* **22**, 1372–1381 (2012).
- Korn, C. W., Prehn, K., Park, S. Q., Walter, H. & Heekeren, H. R. Positively biased processing of self-relevant social feedback. *J. Neurosci.* **32**, 16832–16844 (2012).
- Hughes, B. L. & Zaki, J. The neuroscience of motivated cognition. *Trends Cogn. Sci.* **19**, 62–64 (2015).
- Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S. & Palminteri, S. Behavioural and neural characterization of optimistic reinforcement learning. *Nat. Hum. Behav.* **1**, 0067 (2017).
- Sharot, T., Korn, C. W. & Dolan, R. J. How unrealistic optimism is maintained in the face of reality. *Nat. Neurosci.* **14**, 1475–1479 (2011).
- Ma, D. S., Correll, J. & Wittenbrink, B. The Chicago face database: a free stimulus set of faces and norming data. *Behav. Res. Methods* **47**, 1122–1135 (2015).
- Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
- Knoblauch, K. & Maloney, L. T. *Modeling Psychophysical Data in R* (Springer, 2012).
- Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
- Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
- Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S* (Springer, 2003).
- Gelman, A. & Rubin, D. B. Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992).
- Smith, S. M. & Nichols, T. E. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* **44**, 83–98 (2009).
- Abraham, A. et al. Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* **8**, 14 (2014).

64. Hughes, B. L., Zaki, J. & Ambady, N. Motivation alters impression formation and related neural systems. *Soc. Cogn. Affect. Neurosci.* **12**, 49–60 (2017).

### Acknowledgements

We thank I. Ballard and members of the Stanford Social Neuroscience Laboratory for scientific discussions and helpful comments on earlier versions of the manuscript. The research was supported by the Wu Tsai Neuroscience Institute NeuroChoice Initiative. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

### Author contributions

Y.C.L., B.L.H. and J.Z. designed the study. Y.C.L. and Y.W. collected and analysed the data. Y.C.L. and J.Z. wrote the manuscript, with revisions from Y.W. and B.L.H.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41562-019-0637-z>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to Y.C.L.

**Peer review information:** Primary Handling Editor: Mary Elizabeth Sutherland.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Behavioral data were collected using custom code written in MATLAB (2016b) and the Psychophysics Toolbox (Version 3).

Data analysis

Choice, reaction time and confidence data were analyzed using custom code written in R (version 3.3.2). Drift diffusion models were implemented using the HDDM toolbox (version 0.6.0) with custom code written in python 2.7.13. Preprocessing and general linear modeling of fMRI data was performed using FSL (version 5.98). Multivariate classification analyses were implemented using tools available as part of the Nilearn (version 0.3.1) module with custom code written in python 2.7.13. All custom code are available at: <https://github.com/ycleong/MotivatedPerception>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon request. Behavioral data of both the reported experiment as well as the in-lab replication are available at: <https://github.com/ycleong/MotivatedPerception>. Unthresholded p-map of the Motivation Consistent – Motivation Inconsistent contrast is available at: <https://neurovault.org/collections/EAAXGDRJ/images/62743/>.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	This is a quantitative study. We collected choice and reaction time data on a two-alternative-forced-choice task, while also measuring the blood-oxygenation-level-dependent response using fMRI.
Research sample	A convenience sample of thirty-three participants were recruited from the Stanford community. Participants provided written, informed consent prior to the start of the study. All experimental procedures were approved by the Stanford Institutional Review Board. Data from three participants were discarded because of excessive head motion (> 3mm) during one or more scanning sessions, yielding an effective sample size of thirty participants (17 male, 13 female, ages 18-43, mean age = 22.3).
Sampling strategy	Convenience sampling. Participants signed up for the study on an online platform. All participants who passed the safety screening for MRI were enrolled in the study. Sample size was determined to be comparable to other fMRI studies of motivational biases (e.g., Hughes, Zaki & Ambady, 2016, SCAN) and perceptual decision-making (e.g., Mulder et al., 2012, J. Neuro)
Data collection	Participants performed the task in the MRI scanner while researchers operated the scanner in a separate room. Data were collected using button presses with custom code written in MATLAB (2016b) and the Psychophysics Toolbox (Version 3). The study employed a within-subject design, so participants were not assigned to experimental condition. Researchers were not blind to the study hypothesis, but had minimal contact with participants beyond safety screening for MRI as instructions were delivered on screen while participants were in the scanner.
Timing	May 10th 2016 - July 1st 2016
Data exclusions	Data from three participants were discarded because of excessive head motion (> 3mm) during one or more scanning sessions. Exclusion criteria were predetermined.
Non-participation	None.
Randomization	The study employed a within-subject design, so participants were not assigned to groups.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

### Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

## Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	A convenience sample of thirty-three participants were recruited from the Stanford community. Participants provided written, informed consent prior to the start of the study. All experimental procedures were approved by the Stanford Institutional Review Board. Data from three participants were discarded because of excessive head motion (> 3mm) during one or more scanning sessions, yielding an effective sample size of thirty participants (17 male, 13 female, ages 18-43, mean age = 22.3).
Recruitment	Convenience sample. Participants signed up for the study on an online platform. All participants who passed the safety screening for MRI were enrolled in the study.

Ethics oversight

Stanford University Institutional Review Board

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Magnetic resonance imaging

### Experimental design

Design type

Event-related task design

Design specifications

Participants performed 4 blocks (2 of each condition) of 40 trials each. Each block was approximately 8 minutes long. Each trial was 8 seconds long (4 seconds of motivation manipulation + 4 seconds for response). There was a variable inter-trial interval of 2-4seconds.

Behavioral performance measures

We recorded both button presses (face or scene) and response times. All participants were more likely to indicate that they saw more scene with increasing proportion of scene in an image, indicating that they were appropriately performing the perceptual categorization task.

### Acquisition

Imaging type(s)

Functional and structural

Field strength

3T

Sequence &amp; imaging parameters

Functional images were acquired in interleaved order using a T2\*-weighted echo planar imaging (EPI) pulse sequence (46 transverse slices, TR=2s, TE=25ms, flip angle=77°, voxel size 2.9 mm<sup>3</sup>). Anatomical images were acquired at the start of the session with a T1-weighted pulse sequence (TR = 7.2ms, TE = 2.8ms, flip angle=12°, voxel size 1 mm<sup>3</sup>).

Area of acquisition

Whole brain

Diffusion MRI

Used

Not used

### Preprocessing

Preprocessing software

Preprocessing was performed using FSL (v. 5.98). Brain extraction was performed using the BET routine available as part of the FSL software. Preprocessing consists of motion correction, slice-timing correction, removal of low-frequency drifts using a temporal high-pass filter (100ms cutoff), and spatial smoothing with a 4-mm FWHM kernel.

Normalization

For multivariate classification analyses, we trained and tested our classifier in each participant's native space. For all other analyses, functional volumes were first registered to participants' anatomical image (Boundary-Based Registration) and then to a template brain in Montreal Neurological Institute (MNI) space (affine transformation with 12 degrees of freedom).

Normalization template

MNI152

Noise and artifact removal

Stimulus onset, reaction time and head movement parameters were entered into the general linear models as nuisance regressors. With the exception of head movement parameters, all regressors were convolved with a hemodynamic response function.

Volume censoring

N.A.

### Statistical modeling & inference

Model type and settings

We implemented two whole-brain mass univariate models using FSL/FEAT v.5.98. At the first level (i.e. runs from the same participant), we estimated fixed effects models. At the second level (i.e. across participants), we estimated mixed effects models. We also estimated a multivariate model to classify whether participants were seeing a face or a scene based on BOLD activity (see below).

Effect(s) tested

We implemented a linear model (GLM 1) to contrast BOLD activity on Motivation Consistent trials and that on Motivation Inconsistent trials. A Motivation Consistent trial was defined as a trial on which participant categorized an image as the category they were motivated to see. We implemented a second linear model (GLM 2) in which the onset of each trial was modeled as a separate regressor. This model allowed us estimate a separate statistical map for each trial (i.e. single trial activation patterns).

Specify type of analysis:

Whole brain

ROI-based

Both

Anatomical location(s)

The nucleus accumbens region of interest (ROI) was defined using the Harvard-Oxford subcortical structural atlas. The ROI can also be downloaded from the following NeuroVault collection: <https://neurovault.org/collections/EAAXGDRJ/>.

Statistic type for inference  
(See [Eklund et al. 2016](#))

Threshold-free-cluster-enhancement

## Models & analysis

- |                                     |  |
|-------------------------------------|--|
| n/a                                 | Involvement in the study   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Functional and/or effective connectivity                |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Graph analysis  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> Multivariate modeling or predictive analysis |

### Multivariate modeling and predictive analysis

Multivoxel pattern analyses were performed using tools available as part of the Nilearn Python module. An L1-regularized logistic regression model ( $C = 1$ ) was trained on BOLD data from the localizer task to classify the image category (face or scene) participants were seeing on each localizer trial. Analysis was restricted to voxels in a ventral visual stream mask consisting of the bilateral occipital lobe and ventral temporal cortex. The mask was created in MNI space using anatomical masks defined by the Harvard-Oxford Cortical Structural Atlas. The mask was then transformed into each participant's native space using FSL's FLIRT implementation, and classification was performed in participants' native space. The trained model was then applied to the single trial activation patterns in the experimental task (GLM 2). On each trial, the classifier returned the probability that the participant was seeing a scene rather than a face based on activity in the ventral visual stream mask.