nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0584-8

Gaze bias differences capture individual choice behaviour

Armin W. Thomas[1,2,3,7], Felix Molter[2,3,4,5,7], Ian Krajbich[6], Hauke R. Heekeren[2,3] and Peter N. C. Mohr[2,3,4,5*]

How do we make simple choices such as deciding between an apple and an orange? Recent empirical evidence suggests that choice behaviour and gaze allocation are closely linked at the group level, whereby items looked at longer during the decision-making process are more likely to be chosen. However, it is unclear how variable this gaze bias effect is between individuals. Here we investigate this question across four different simple choice experiments and using a computational model that can be easily applied to individuals. We show that an association between gaze and choice is present for most individuals, but differs considerably in strength. Generally, individuals with a strong association between gaze and choice behaviour are worse at choosing the best item from a choice set compared with individuals with a weak association. Accounting for individuals' variability in gaze bias in the model can explain and accurately predict individual differences in choice behaviour.

In everyday life, we are constantly confronted with simple consumer choices such as whether to have an apple or a banana for breakfast or which bottle of juice to buy at the supermarket. Traditional models describing this type of consumer choice assume that people assign a utility (or value) to each available option and make utility-maximizing choices[1]. Notably, choices are assumed to be based solely on the attributes of the option, and are therefore independent of information search processes during the decision-making process[2]. This assumption has recently been challenged by a variety of empirical findings showing that the allocation of gaze during the decision-making process also plays a substantial role, as a longer gaze towards one option is regularly associated with a higher choice probability for that option (independent of its value)[3–17]. Similarly, stimulus salience has been shown to influence decision behaviour[18–20]. Furthermore, external manipulation of gaze allocation leads to changes in choice probabilities[3,12,14]. Similar effects have recently been demonstrated in perceptual decision-making, whereby participants judge perceptual attributes of stimuli based on the available sensory information (for example, the orientation of line segments[21]).

These findings led to the development of computational models that integrate eye movement data into the choice process and formalize the empirically observed association between gaze and choice[4,9–11,20,22–24]. These models are based on classical evidence-accumulation models[25,26], but make the additional assumption that the momentary rate of evidence accumulation depends on the eye movements of the decision-maker. Evidence accumulation for an option is assumed to be discounted by a constant factor while another item is fixated on. Accounting for this gaze bias, these models provide a precise account of many aspects of simple choice behaviour at the group level[4,9–11,20,22–24].

While group-level statistics are informative for some research questions (for example, to specifically address differences between groups or experimental conditions or to forecast product sales in economic research), they can be unsuitable for understanding the choice behaviour of an individual. Aggregate models can lead to false conclusions about true underlying individual processes[27,28]. In a learning task, for example, the group-level average learning curve would appear as a gradual, smooth function over time, even if all individuals showed abrupt, step-like learning curves (much like an epiphany), but with variable learning onsets across individuals[29]. In this case, the group-level model would not accurately describe any individual of the group, and the deduction that individual learning occurs smoothly would be false. Similarly, using a single model parameter set to describe the choice behaviour of a group could lead to false conclusions about the behaviour of the underlying individuals. Therefore, it is crucial to study choice behaviour at the level of the individual.

Previously reported group-level models that quantify the association between gaze and choice specified a constant gaze bias for all individuals without rigorously testing the performance of the model at the level of the individual[9,10]. A rigorous test of gaze bias effects at the level of the individual should ideally be based on non-restricted individual model fits, include comparisons to models without gaze bias, establish that the model provides an accurate account of individually observed data and test how individuals' gaze biases relate to their response behaviour. If, for example, people's decisions were affected differently by gaze behaviour, we would find that the choices of some individuals were more biased by gaze than others and possibly be more inconsistent with the values of the items. Imagine, for example, a choice between two bottles of juice at the supermarket: one has a slightly higher value for the decision-maker than the other, but it is also less visually salient[20,30]. If a person's association of gaze and choice behaviour was strong, their choice would be biased towards the more visually salient bottle that attracts more of their gaze, even though it has a lower value. Conversely, if the person's association was weak, they would be able to select the higher valued option, despite their gaze

[1]Department of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany. [2]Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany. [3]Center for Cognitive Neuroscience Berlin, Freie Universität Berlin, Berlin, Germany. [4]School of Business and Economics, Freie Universität Berlin, Berlin, Germany. [5]WZB Berlin Social Science Center, Berlin, Germany. [6]Department of Psychology and Department of Economics, The Ohio State University, Columbus, OH, USA. [7]These authors contributed equally: Armin W. Thomas, Felix Molter.
*e-mail: peter.mohr@neuroeconomics-lab.de

being attracted more towards the visually salient but lower valued option. Accordingly, if the strength of this association is variable across individuals, it is necessary to account for these differences to accurately predict individual choice behaviour.

Here, we investigated whether the previously reported link between gaze and choice behaviour is variable across individuals. We analysed four previously published choice datasets[6,9,10,21], in total including 118 individuals, two choice set sizes (two- and three-alternative) and two choice domains (value-based and perceptual). For the analysis, we developed a computational model that can be easily applied to individual participant and multialternative choice data. With this model, we reaffirmed that an association between gaze and choice is present for most individuals. The strength of this association, however, varied substantially. This variability was directly linked to an individual's ability of choosing the best item from a choice set, such that stronger associations of gaze and choice were linked to lower probabilities of choosing the best item. Accounting for the variability in individuals' gaze biases, we were able to explain and accurately predict observed differences in choice behaviour.

Results

Datasets and task overview. We investigated individual differences in the influence of gaze allocation on simple choice behaviour across four previously published datasets[6,9,10,21]. In each dataset, healthy participants made repeated decisions between multiple options while their eye movements were recorded (for additional details, see Fig. 1 and Supplementary Methods 1).

The first dataset is from Krajbich and colleagues[10] (henceforth referred to as Krajbich 2010). In the corresponding experiment, hungry participants made choices between two snack food items without any time restrictions (Fig. 1a). Participants also gave a liking rating for each of the 70 snack food items that were used in the experiment. This dataset includes 39 participants, each of whom performed 100 trials.

The second dataset from Krajbich and Rangel[9] (henceforth referred to as Krajbich 2011) is similar to that from Krajbich 2010. In Krajbich 2011, participants chose between three snack food items (Fig. 1b). As in Krajbich 2010, participants provided liking ratings for all available items in a separate task. This dataset includes 30 participants, each of whom performed 100 trials.

The third dataset consists of experiment 2 from the study by Folke and colleagues[6] (henceforth referred to as Folke 2016). In this experiment, 24 hungry participants performed 144 trials of a task that closely resembled the Krajbich 2011 three-alternative forced-choice snack food task (Fig. 1c). Unlike in Krajbich 2011, however, the choice task used a gaze-contingent presentation, whereby items were only revealed when the participant's gaze was directed to an item's location on the screen. In addition, after each choice, the participants provided confidence ratings (which we did not use in this study). Similar to Krajbich 2010 and Krajbich 2011, item values were estimated in a separate task, in which a Becker–DeGroot–Marschak auction procedure was used to elicit willingness-to-pay estimates[31].

The fourth dataset is the first experiment from the study by Tavares and colleagues[21] (henceforth referred to as Tavares 2017). This dataset is qualitatively different from the other three value-based choice datasets. Participants made perceptual judgments about the orientations of two line segments and were asked to decide which is closer to a target (Fig. 1d). In this case, we define the value of an item by its angular distance to the target (with higher values for smaller distances). This dataset includes 25 participants, each of whom performed 1,344 trials across four sessions.

In sum, our analyses span a total of four datasets ($n=118$) that contain data from two- and three-alternative forced-choice tasks in two different choice domains (value-based and perceptual) and from two different laboratories.

Individual differences in the data. We analysed the following three metrics for individual differences: participants' mean response time (RT); mean probability of choosing the best item (we define the best item either as the item with the highest liking rating or willingness-to-pay in the value-based choice tasks, or the item with the smaller angular distance to the target in the perceptual choice task); and influence of gaze allocation on choice probability (defined as the mean increase in choice probability for an item that was fixated on longer than the others, after correcting for the influence of the item value; see below). We found that participants differed considerably in all metrics (Fig. 2). The participants' ($n=118$) overall mean RTs ranged from 958 to 7,577 ms, with a mean ± s.d. of 2,844 ± 1,676 ms (Fig. 2a), while their probabilities of choosing the best item in a trial ranged from 45% to 97%, with a mean ± s.d. of 77% ± 12% (Fig. 2c).

We also probed the relationship between individual allocation of gaze and choice. Previous studies of simple choice tasks have shown that individuals are more likely to choose an option when they spent more time fixating on it compared with the other opti ons[3,4,6,9,10]. Here, we devised a measure to quantify this relationship for each individual. Following previous work[9,10], we first estimated a participant's probability of choosing an item in a choice set using logistic regression, based on its relative item value (the difference between the item's value and the mean value of all other items in that trial) and the range between the other items' value (this regressor was omitted in all two-item datasets). We then subtracted this estimated probability from the empirically observed choice (either 1 if the item was chosen or 0 otherwise). Finally, we averaged the resulting 'residual' choice probability for trials in which the item had a positive and negative final gaze advantage (computed as the difference in the fraction of the total fixation time that the participants spent fixating on the item and the average fraction that they spent fixating on the others). The difference between these two described the average difference in choice probability for the items with a positive versus negative final gaze advantage, when corrected for the influence of the values of the items. We found that individual scores on this measure varied substantially and ranged from −11% to 72%, with a mean ± s.d. of 24% ± 15% (Fig. 2b). Notably, 98% of the participants showed positive scores, indicating an overall positive relationship between gaze allocation and choice.

The four datasets differed strongly in the three behavioural metrics (Table 1). Differences between datasets, however, cannot be attributed to the effect of choice domain (perceptual versus value-based) or set size (two versus three items) alone, as original tasks also differed in other aspects (for example, different stimuli in value-based versus perceptual tasks, different number of trials and different presentation format). However, when comparing the two- and three-alternative choice data, individuals exhibited shorter response times (Mann–Whitney $U=821$, $P<0.001$, Pearson's $r=0.52$) and a higher probability of choosing the best item ($U=444$, $P<0.001$, $r=0.74$) when making choices between two alternatives. Furthermore, individuals exhibited a weaker association of gaze and choice behaviour, indicated by the gaze influence measure ($U=1{,}001$, $P<0.001$, $r=0.42$), when making choices between two alternatives. Comparing behavioural data from value-based and perceptual choice tasks, we found that response times were shorter ($U=632$, $P<0.001$, $r=0.46$) and individuals had a higher probability of choosing the best item ($U=342$, $P<0.001$, $r=0.71$) in the perceptual choice task. The average strength of the association of gaze and choice behaviour in the perceptual choice task, however, was similar to that measured in the two-alternative value-based choice experiment ($U=419$, $P=0.17$, $r=0.14$).

All regression coefficients reported throughout represent fixed effects from Bayesian mixed-effects linear (for continuous-dependent variables) or logistic (for binary-dependent variables) regression models, including a random intercept and slope for each dataset on each predictor. For each fixed effect, we report the

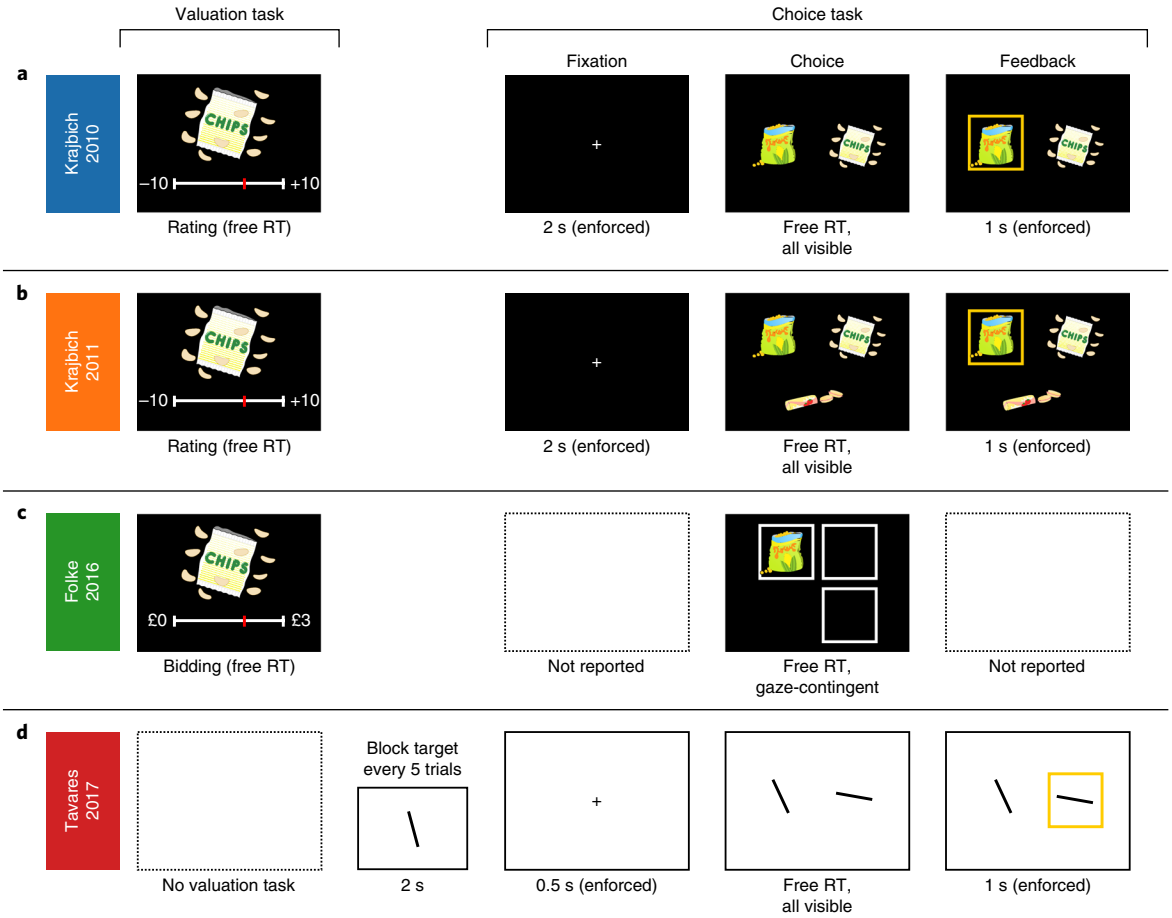

Fig. 1 | Experimental paradigms. a–d, We included four datasets in our analyses. These include three value-based experiments (**a** Krajbich 2010[10], **b** Krajbich 2011[9], **c** Folke 2016[6]) and one perceptual choice experiment (**d** Tavares 2017[21]). In all experiments, participants were instructed to choose the best out of two (**a**, **d**) or three (**b**, **c**) items (that is, the item they would like to eat most in value-based tasks or the item most similar to a target stimulus that was presented every five trials in the perceptual task). Value-based experiments included a valuation task before the main choice task, whereby participants either rated each item (**a**, **b**) or indicated their willingness-to-pay in a Becker–DeGroot–Marschak procedure (**c**). All choices were made without time restrictions. The choice task in **c** used a gaze-contingent presentation, whereby items were only revealed when the participant's gaze was directed to an item's location on the screen. Experiments used real snack food items instead of illustrations. For additional details, see Supplementary Methods 1.

coefficient's posterior mean (β) and the associated 95% highest posterior density interval (HDI) values (for further details, see the "Mixed effects modelling" section).

To further probe the relationship between the three behavioural metrics, we computed pairwise mixed-effects regression models between them (Fig. 2d–f). We did not find any association between participants' probability of choosing the best item and their RTs (Fig. 2d; $\beta = -0.19\%$, 95% HDI = [−3.08%, 2.87%] per second increase in RT). Similarly, participants' gaze influence was not related to their RTs (Fig. 2e; $\beta = -1$ ms, 95% HDI = [−33 ms, 33 ms] per percentage increase in the gaze influence measure). However, participants' probability of choosing the best item from a choice set decreased with increasing individual gaze influence measures (Fig. 2f; $\beta = -0.34\%$, 95% HDI = [−0.71%, 0.08%] per percentage increase in gaze influence, 95.2% of posterior density below 0).

Modelling individual differences in gaze influence on simple choice. The behavioural and eye-tracking data suggested that there is substantial variability in the extent to which gaze affects participants' choice behaviour (Fig. 2b). However, conclusive quantitative evidence for or against the presence of a mechanism that biases choices depending on the distribution of gaze has yet to be provided at the level of the individual (for example, by means of a formal model comparison). We therefore adopted a computational modelling approach to investigate whether a formalized gaze bias mechanism, in conjunction with individual gaze patterns, can improve model predictions of individual choice and RT data when compared with a model without gaze bias.

We propose a model called the Gaze-weighted Linear Accumulator Model (GLAM; Fig. 3) that we view as an analytical tool to study gaze bias effects at the level of the individual and that is inspired by the multialternative attentional Drift Diffusion Model (aDDM)[9]. Similar to the aDDM, the GLAM assumes accumulation of evidence in favour of each item that is modulated by gaze behaviour. While an item is not fixated on, accumulation occurs at a rate discounted by the gaze bias parameter γ (Fig. 3a). A choice is made as soon as evidence in favour of one item reaches a decision threshold. In contrast to the aDDM, which focuses on the effect of individual trial fixation sequences on the decision-making process, the GLAM focuses on gaze bias effects at the level of the trial. Therefore, it can average over the observed sequence of fixations within a trial (Fig. 3b). The resulting gaze-weighted decision signals are then fed through a logistic transform into a linear stochastic race[32,33] (Fig. 3c,d). Race models are generalizable to choice scenarios with more than two items and remain analytically tractable, allowing for more complex applications (for example, Bayesian parameter estimation,

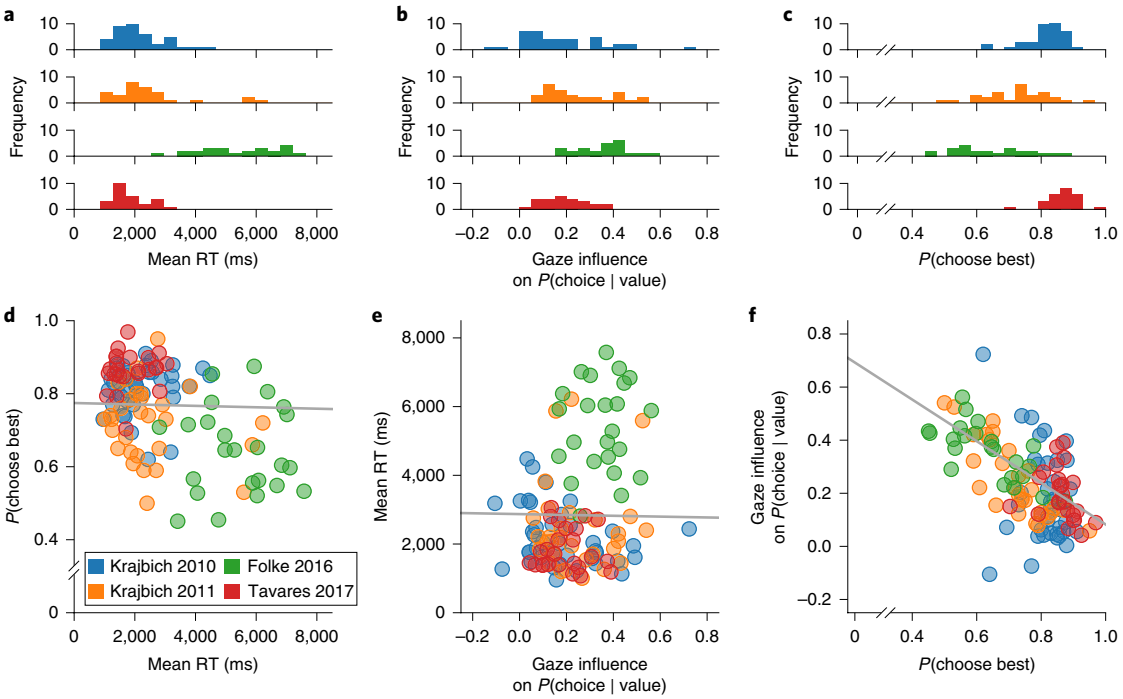

Fig. 2 | Individual differences in the three studied behavioural metrics and their associations. a–**c**, Distributions of individuals' mean RT (**a**), gaze influence (mean increase in choice probability for an item that is fixated longer than the others, after correcting for the influence of item value) (**b**) and probability of choosing the best item (**c**) per dataset. **d**, There is no association between mean RTs and the individual probability of choosing the best item ($\beta = -0.19\%$, 95% HDI = [−3.08%, 2.87%] per second increase in RT). **e**, There is no association between gaze influence and the mean RT ($\beta = -1$ ms, 95% HDI = [−33 ms, 33 ms]) per percentage increase in the gaze influence measure. **f**, An individual's probability of choosing the best item decreases with increasing gaze influence ($\beta = -0.34\%$, 95% HDI = [−0.71%, 0.08%] per percentage increase in the gaze influence measure). Each circle represents one individual participant. Grey lines represent the fixed effect from mixed-effects regression models with random slopes and intercepts for each dataset. The key in **d** is applicable to all panels.

Table 1 | Description of the behavioural metrics of the included datasets

Dataset	Krajbich 2010	Krajbich 2011	Folke 2016	Tavares 2017	Overall
N	39	30	24	25	118
Set size	2	3	3	2	-
Choice domain	Value-based	Value-based	Value-based	Perceptual	-
Mean RT (ms)	2,192 ms (851 ms)	2,462 ms (1,298 ms)	5,414 ms (1,284 ms)	1,849 ms (601 ms)	2,844 ms (1,676 ms)
***P*(choose best)**	81% (6%)	72% (10%)	66% (12%)	86% (5%)	77% (12%)
Gaze influence	19% (17%)	25% (14%)	35% (11%)	19% (9%)	24% (15%)

Means are given, with respective standard deviations in parentheses.

embedding in a hierarchical Bayesian framework). In addition to the gaze bias parameter γ, the GLAM includes a velocity parameter v, a noise parameter σ and a scaling parameter τ (for further details, see the "GLAM details" section).

Even though both models are closely related, the GLAM has a practical advantage over the aDDM in that it sidesteps the complex problem of modelling and simulating the fixation process in a given task. The development of such fixation models for individual participants, particularly in more complex choice scenarios with more than two items or multiple item attributes, is often not trivial (for example, see ref. [20]) or not of main interest to the researcher. The GLAM solely uses the observed distribution of gaze to the items over the course of the trial. In contrast, the aDDM is fitted to empirical data using model simulations, which themselves rely on an accurate simulation of the fixation trajectories. As a side effect, this allows the application of the GLAM in situations in which only limited trial-level data are available (for example, the Folke 2016 dataset included here, which only contains trial-averaged gaze data). In theory, a similar simplification of the multialternative aDDM[9] would be possible, but would result in a model highly similar to the GLAM. Furthermore, fitting such a simplified aDDM variant would still rely on simulations of the decision-making process. These simulations are particularly costly in the case of the aDDM, whereby every trial represents a unique condition owing to the incorporation of trial-specific eye movement data.

Testing the presence of gaze biases in individuals. We fitted and compared two GLAM variants to the RT and choice data of each participant to gauge the evidence in favour of the previously described gaze bias mechanism and to quantify its strength on an individual

Fig. 3 | The GLAM. a–d, The GLAM describes the influence of gaze allocation on the decision-making process in the form of a linear stochastic race. While participants look at the available choice options (**a**), an absolute evidence signal A_i for each option in the choice set is computed (**b**). The magnitude of this signal is dependent on the allocation of gaze, with lower magnitudes for options that are momentarily not fixated on. Absolute evidence signals are transformed into relative decision signals (indicating relative item preferences) by computing the average absolute evidence signal for each item in the trial (broken lines in **b**) and computing the difference between each of these averages and the maximum of the other two. The GLAM assumes an adaptive representation of these relative evidence signals that is maximally sensitive to small differences in the relative decision signals. To this end, a logistic transform is applied (**c**). The resulting scaled relative evidence signals determine the drift terms R_i of the relative evidence accumulators E_i in the stochastic race (**d**). A choice for an option is made as soon as the accumulated relative evidence E_i reaches a choice threshold. The stochastic race provides first-passage time distributions p_i, describing the likelihood of each item being chosen at each time point. For a more detailed model description, see the "GLAM details" section. Colours indicate choice alternatives.

level. The first is a full GLAM variant (with gaze bias) with free parameters ν, γ, σ and τ. This model allowed the gaze bias parameter γ to vary freely between individuals. The second is a no-gaze-bias GLAM variant, whereby the gaze bias parameter γ was fixed to 1 (resulting in no influence of gaze on the accumulation process).

The two models differ in their complexity. The full model has one more free parameter and can therefore be expected to provide a better absolute fit to the data. We used the widely applicable information criterion (WAIC)[34] to perform model comparisons at the level of the individual, as it includes a penalty for model complexity. Lower WAIC scores indicate a better model fit, accounting for differences in model complexity.

Overall, the full GLAM fitted 109 out of 118 (92%) participants better than the no-gaze-bias variant. Within each dataset, the data of 79% (Krajbich 2010), 97% (Krajbich 2011), 100% (Folke 2016) and 100% (Tavares 2017) of the participants were better described by the full GLAM (Fig. 4a).

This analysis also suggested a categorical distinction between the perceptual and value-based choice datasets (such that the comparison more clearly favours the model with gaze bias in perceptual decisions). However, these more extreme differences in relative model fit (difference in WAIC scores; Fig. 4b) could be driven by the fact that the perceptual dataset contains approximately nine times more trials per participant than the other datasets, allowing the comparison to be more decisive. Consequently, they cannot necessarily be attributed to the difference between perceptual and value-based decision processing alone.

Individual estimates of the gaze bias parameter γ in the full model ranged from −1.03 (strong gaze bias) to 0.97 (almost no gaze bias), with a mean ± s.d. of 0.15 ± 0.39 (n = 118) (Supplementary Fig. 1). Importantly, sizeable gaze biases were present for all datasets (across choice domains and set sizes), with mean ± s.d. values of 0.26 ± 0.48 (Krajbich 2010, n = 39), 0.18 ± 0.41 (Krajbich 2011, n = 30), −0.017 ± 0.28 (Folke 2016, n = 24) and 0.08 ± 0.23 (Tavares 2017, n = 25). Note, however, that the order of datasets according to γ estimates differs from their order based on the behavioural gaze influence measure (Fig. 2b and Table 1). This result demonstrates the conceptual difference between the behavioural measure and the model's estimates of the latent gaze bias variable. While the behavioural measure aggregates instances of observed behaviour, the latent gaze bias γ describes the assumed underlying generative mechanism.

We further probed the relationship between the difference in the models' WAIC scores (which describe how much better the data of an individual is described by the full GLAM relative to the no-gaze-bias variant) and the three behavioural metrics. We did not find any association between the differences in WAIC scores and mean RTs (β = 1 ms, 95% HDI = [−17 ms, 10 ms] per unit increase in WAIC difference). However, both the probability of choosing the best item in a trial (Fig. 2c; β = 0.10%, 95% HDI = [−0.01%, 0.22%] per unit increase in WAIC difference; 95.9% of posterior density above 0) and the strength of participants' influence of gaze on choice (Fig. 2b; β = −0.21%, 95% HDI = [−0.42%, 0.01%] per unit increase in WAIC difference, 96.4% of the posterior density below 0) varied systematically with the WAIC differences. The full GLAM therefore outperformed the no-bias variant, particularly for individuals with low choice accuracies and a strong influence of gaze on their choices.

Taken together, these findings provide strong empirical evidence to indicate that a gaze bias mechanism is present for most participants. Importantly, the extent to which the accumulation process was influenced by gaze, as captured by individual gaze bias (γ) estimates, showed nontrivial individual differences.

Probing the functional form of individuals' gaze bias. We also compared the gaze bias mechanism implemented in the GLAM against another variant that included an additive effect of gaze on choice behaviour (see also ref. [4]; for a detailed description of this variant, see Supplementary Methods 2). This comparison revealed that similar proportions of participants were better described by either model variant, with an additional group of participants whose choice behaviour was described similarly well by both variants (Supplementary Fig. 2a). We therefore concluded that there is not a 'winning' gaze bias mechanism that we can identify across individuals. Model simulations also revealed that both variants described participants' response behaviour similarly well and mimicked each other considerably in their predictions, both on the individual (Supplementary Fig. 2b–d) and group-averaged level (Supplementary Fig. 3). The variants' gaze bias estimates were also highly correlated (Spearman's $\rho(117) = -0.86$, $P < 0.001$). Therefore, we decided to continue using the original multiplicative variant, as multiplicative effects have received more empirical support in the literature[9–11,20,21,35–37]. Importantly, however, the results and conclusions of our study would not change if we had used the additive variant instead (Supplementary Fig. 2). Further distinguishing these

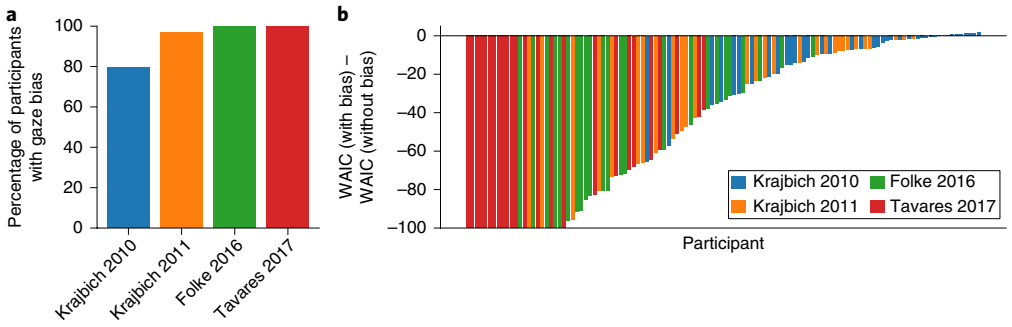

Fig. 4 | Individual relative model comparison between the full GLAM and a restricted no-gaze-bias GLAM variant. a, Across datasets, the response behaviour of most individuals is better described by the full model with a gaze bias (given by the lowest score on the WAIC). **b**, Individual WAIC differences between the full and restricted GLAM variant. Negative differences indicate better fits of the full model. Note that the *y* axis in **b** is truncated to better show small differences. The lowest WAIC difference was −400.64.

variants would ideally require choice datasets that include both appetitive and aversive choice options (positive and negative values, respectively), as the two mechanisms' predictions diverge more clearly in these choice settings (for example, a multiplicative effect would predict that a longer gaze towards an aversive item should reduce its probability of being chosen, whereas an additive effect would predict the opposite).

Predicting individual choice behaviour. We found that in a relative model comparison, the data of most participants were better described by the full model with a gaze bias mechanism compared with a restricted variant with no gaze bias ($\gamma = 1$; Fig. 4). However, this analysis did not take into account whether the model with bias also accurately predicts individuals' behaviour on an absolute level. To test this, we used both model variants to simulate response data for each individual. This time, however, we split the data into even- and odd-numbered trials. We then used all even trials to estimate individual model parameters. Subsequently, we predicted choices and RTs for all odd-numbered trials, thereby comparing model predictions to data that did not inform the parameter estimates. We note, however, that even- and odd-numbered trials from the same participant are not fully independent from one another.

To assess the quality of the fit of both models' predictions to the empirically observed data across datasets, we performed the following test. For each model and each behavioural measure, we computed a mixed-effects regression, regressing the respective measure onto a binary variable, which indicates whether each value on this measure comes from the empirically observed data or from the model simulations. If the fixed-effects estimate of the indicator variable differed from 0, model predictions deviate meaningfully from observed data across datasets. Overall, the full model accurately predicted participants' RTs (Fig. 5a; $\beta = -9$ ms, 95% HDI = [−410 ms, 344 ms] difference between the observed and predicted data), the probability of choosing the best item (Fig. 5b; $\beta = -2.22\%$, 95% HDI = [−7.03%, 2.30%] difference between the observed and predicted data) as well as the strength of their gaze influence (Fig. 5c; $\beta = -2.20\%$, 95% HDI = [−8.25%, 4.21%] difference between the observed and predicted data). The full model also accurately recovered the observed associations between the three behavioural metrics (for a comparison, see Fig. 2 and Supplementary Fig. 4). The no-gaze-bias variant predicted the participants' individual mean RTs (Fig. 5d; $\beta = 15$ ms, 95% HDI = [−355 ms, 413 ms] difference between the observed and predicted data) and the probability of choosing the best item (Fig. 5e; $\beta = 0.13\%$, 95% HDI = [−4.43%, 4.83%] difference between the observed and predicted data) similarly well. However, the restricted model by design cannot predict the influence of gaze on the participants' choices (Fig. 5f; $\beta = -22.72\%$, 95% HDI = [−30.81%, −13.20%] difference between the observed and predicted data), resulting in no association between the predicted and empirical data in our gaze influence measure. This illustrates the inferiority of the restricted model. The full model further accurately captured the distribution of RTs within and across individuals (Supplementary Figs. 5 and 6).

Overall, these results show that the full model with gaze bias outperformed the restricted model in accurately predicting the participants' empirical choices, as it also captured empirical choice patterns that are driven by gaze and not solely by the values of the items.

Model parameters explain individual choice behaviour. We found that the full model with gaze bias accurately predicted individuals' response behaviour. Next, we tested whether the model's parameters are able to explain variability in participants' choice behaviour. Again, we used mixed-effects models to predict the three behavioural metrics in the odd-numbered trials from the parameters estimated from the even-numbered trials (Fig. 6). We found that v (velocity parameter) scaled logarithmically with the participants' mean RT (Fig. 6a; $\beta = -0.79$ log(ms), 95% HDI = [−0.85 log(ms), −0.71 log(ms)] per unit increase in log(v)). We did not find a meaningful relationship between σ estimates and the probability of choosing the best item ($\beta = -0.23\%$, 95% HDI = [−2.32%, 1.25%] per 0.001 increase in σ), even though the σ parameter determines the magnitude of noise in the accumulation process. We also found that γ estimates predicted the strength of participants' gaze influence on choice probability (Fig. 6b; $\beta = -26.59\%$, 95% HDI = −37.10%, −17.24% per unit increase in γ).

Additionally, we found that γ (gaze bias) estimates relate to participants' probabilities of choosing the best item (Fig. 6c; $\beta = 12.65\%$, 95% HDI = [−3.19%, 28.84%] per unit increase in γ; 94.5% of the posterior density estimates were greater than 0). Thereby, stronger gaze biases (smaller γ) were associated with more choices inconsistent with the value of the items. This relationship can be explained as follows: the gaze bias parameter lets the model bias the choice process according to the distribution of gaze between items. That is, with a strong gaze bias, the model's predictions are strongly dependent on the distribution of gaze, and a gaze distribution that is random with respect to the value of the items then leads to more random choices. Conversely, the model's predictions are independent of gaze when no gaze bias is present. The model then neglects gaze and predicts choices solely driven by the values of the item.

To further probe the robustness of this association between gaze bias strength and probability of choosing the best item, we performed three additional tests. First, we tested whether the correlation of individual gaze bias estimates and individuals' probability of choosing the best item was also present and statistically meaningful

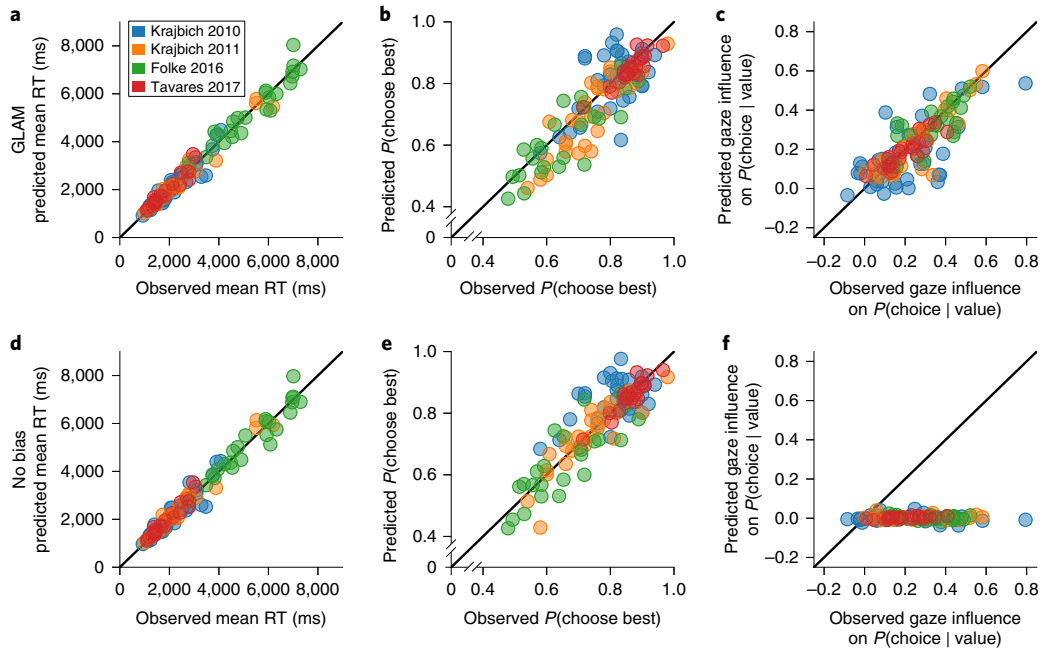

Fig. 5 | Individual out-of-sample predictions of behavioural metrics for all odd-numbered trials. a–**c**, The full model with gaze-bias variant accurately predicts individuals' mean RT (**a**), probability of choosing the best item (**b**) and influence of gaze on choice probability (**c**). **d**–**f**, While the no-gaze-bias variant also accurately predicts individuals' mean RT (**d**) and probability of choosing the best item (**e**), it fails to accurately capture individuals' influence of gaze on choice probability (**f**). See the "Predicting individual choice behaviour" section for the corresponding statistical analysis. Model predictions are simulated using parameter estimates obtained from individual fits on even-numbered trials (for more details on the simulation procedures, see the "Model simulations" section).

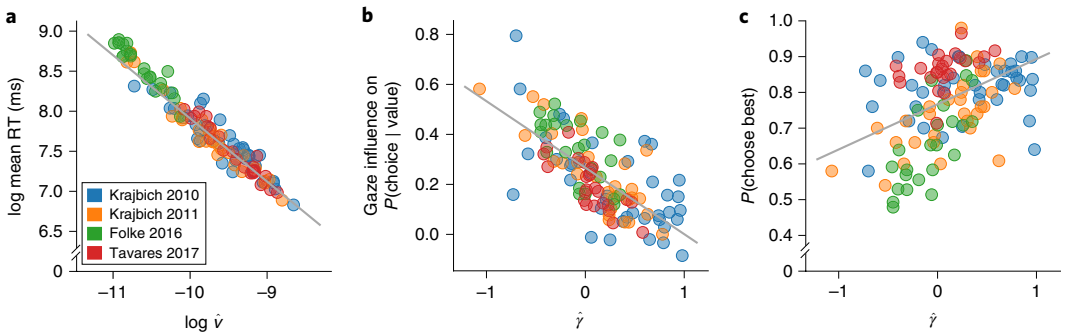

Fig. 6 | Associations between individuals' response behaviour in the odd-numbered trials and the model parameters estimated from the even-numbered trials. a, log-transformed mean RTs decrease with increasing log-transformed v estimates ($\beta = -0.79$ log(ms), 95% HDI = [−0.85 log(ms), −0.71 log(ms)]). **b**, Behavioural gaze influence decreases with increasing γ estimates ($\beta = -26.59\%$, 95% HDI = [−37.10%, −17.24%]). **c**, Individuals' probability of choosing the best item increases with increasing γ estimates ($\beta = 12.65\%$, 95% HDI = [−3.19%, 28.84%] per unit increase in γ; 94.5% of the posterior density estimate were greater than 0).

within each of the four datasets. The relationship is present in three out of the four included datasets (Krajbich 2010: Spearman's $\rho(38) = 0.15$, $P = 0.36$; Krajbich 2011: Pearson's $r(29) = 0.62$, $P < 0.001$; Folke 2016: Pearson's $r(23) = 0.75$, $P < 0.001$; Tavares 2017: Pearson's $r(24) = 0.41$, $P = 0.04$). Second, we ascertained that this relationship is not determined by first trial fixations only. First trial fixations have been shown to be less influenced by item value[9,10] and more driven by other factors, such as screen position (for example, first fixations are often directed towards items in the upper left portion of the screen, irrespective of the value of that item). One could therefore hypothesize that individuals with a strong association of gaze and choice are more prone to choosing the first item seen in a trial (and thereby less likely to choose the best item) compared with individuals with a weak association. This effect would then diminish, however, in the later stages of the trial (in which fixations are driven more by item value). To establish whether this relationship also holds later in the trial, we repeated our analyses after discarding first fixations in each trial (and trials in which only one fixation was made). Note that this analysis did not include the Folke 2016 dataset, which only contains trial-aggregated gaze data, and thus did not allow us to remove first fixation data. Importantly, the relationship between individuals' gaze bias strength and their probability of choosing the best item was still present and statistically meaningful ($\beta = 8.78\%$, 95% HDI = [0.14%; 17.08%] increase in probability of choosing the best item per unit increase in γ; 97.3% of the posterior density estimate was greater than 0). Third, to ensure that the

relationship between the estimated gaze bias strength (γ) and individuals' probability of choosing the best item is not dependent on the specific formulation of the gaze bias mechanism in the GLAM, we also established that this association remains statistically meaningful when using γ estimates from the additive variant instead ($\beta = -0.54\%$, 95% HDI = [−0.98%, −0.06%] per unit increase in γ_{additive}; 97.8% of the posterior density estimate were greater than 0).

Discussion

We investigated individual differences in the influence of gaze allocation on choice behaviour by analysing four previously published datasets[6,9,10,21], in total including 118 individuals, two choice set sizes (two- and three-alternative) and two choice domains (value-based and perceptual). Across datasets, we found an overall positive behavioural relationship between gaze and choice (with a longer gaze increasing choice probability). The strength of this relationship, however, was highly variable across individuals. To better understand the underlying computational mechanism, we proposed a model called GLAM, which is inspired by the multialternative aDDM[9] and can be used to study gaze biases at the level of the individual. The GLAM assumes that individuals accumulate evidence in favour of each available item and make a choice as soon as the cumulative evidence for one item reaches a choice threshold. Importantly, the accumulation process is biased by gaze behaviour, with discounted accumulation rates for unattended items. The model is statistically and computationally tractable, making it readily extendable to novel choice tasks and research questions. Generally, the GLAM can be seen as a way to sidestep the complex problem of simulating individual fixation trajectories. Although researchers have started to explore generative fixation models in simple decision-making tasks (for example, see ref. [20]), this is often not feasible or not of main interest to researchers trying to understand the influence of gaze allocation on the decision-making process. Here, the GLAM provides a tractable, but simplified, alternative to the aDDM that solely requires trial-level statistics, namely overall gaze proportions (next to the RTs, choices and item values).

We then used the GLAM to perform three rigorous tests of gaze bias effects at the level of the individual. First, we formally tested whether individuals' behaviour was better described by a model with or without gaze bias. In this comparison, a large majority of participants (109 out of 118) were better described by the full model with gaze bias than by a restricted variant without. Second, we established that the full model accurately predicts observed behavioural differences between individuals, namely, in the RT, the probability of choosing the best item and the observed influence of gaze allocation on choice behaviour. Third, we tested how individuals' gaze bias estimates relate to their response behaviour. The strength of individuals' gaze biases was predictive of both the strength of individuals' association of gaze and choice and individuals' probability of choosing the best item (stronger gaze biases were associated with more choices that were inconsistent with item values). This identifies another source of variability among individuals' ability to choose the best item from a choice set. Previously, these differences were mostly attributed to differences in generic accumulation noise parameters[38,39], obscuring further insight into the mechanisms driving these individual differences.

Thereby, our approach goes beyond previous analyses of individuals' gaze biases. For example, the Krajbich 2010 study[10] reported individual gaze bias estimates in supplementary figure 11 of their paper. These estimates, however, do not result from non-restricted model fits and also leave open the possibility that individuals' behaviour might be better described by a model without gaze bias.

One reason for the superior performance of models with a gaze bias is their use of individual trial gaze data, which allows them to make different predictions across otherwise identical choice sets. Leveraging a gaze bias mechanism lets these models make trial-specific predictions, which will have higher predictive accuracy owing to the positive relationship between gaze and choice. Conversely, a stochastic choice model without a gaze bias mechanism will make probabilistic, but identical, predictions for two such trials. Previous work has also shown a higher influence of gaze on choice in trials in which individuals did not have strong preferences among the alternatives[19], suggesting that decision models with a gaze bias mechanism will be particularly useful in these situations.

Our analyses also confirmed the need to account for individual variability in model parameters, as we found substantial variability across individuals in the influence of gaze on choice that was hidden in the group-level analyses. Given that the influence of gaze on choice is variable among individuals, a single gaze bias parameter γ for the whole group would not fit all individuals well, and therefore result in inferior predictive performance of the model. On the one hand, individuals whose link between gaze allocation and choice behaviour is weaker than the group average would falsely be predicted to make choices less consistent with item values, and driven more by looking behaviour. On the other hand, predictions for individuals' choices with a stronger link than the group average would not contain enough influence of gaze. Accounting for individual differences in the link between gaze allocation and choice behaviour opens important avenues for future research that focus on the specific determinants of these differences. For example, are these differences best characterized as a trait (stable within a person, but variable between persons, as suggested in ref. [40]), a state (variable within a person, between different situations or contexts) or both (variable between persons and contexts)?

Despite a wealth of evidence exploring the computational mechanisms underlying simple choice behaviour and its link to gaze allocation[3,4,9–11,14,20,23], most of this work, and the associated computational frameworks (for example, see ref. [26]), is difficult to extend to choice scenarios involving more than two choice alternatives. For example, a previous study[4] explored the link between gaze behaviour and the choice mechanism underlying binary value-based choices. To model the decision process, the authors used a hierarchical variant of the DDM, combined with a trial-averaged gaze-weighting mechanism similar to the GLAM. This model is strictly limited to binary decisions, as it describes the decision-making process as a single accumulator that diffuses between two decision bounds (each bound representing one of the two choice alternatives). Therefore, this model cannot be used to study choices between more than two alternatives. Furthermore, their hierarchical estimation of the model's parameters is solely focused on obtaining better group-level estimates and not on better capturing individuals' choice behaviour. As a result, they neither analyse nor report individual parameter estimates or any associations of such estimates with individuals' response behaviour. Our work thereby expands on their findings in three meaningful ways. First, our modelling approach can be extended to an unlimited amount of choice alternatives. Second, our analyses, decision modelling and findings focus explicitly on the individual level, leading to insights into the association of individuals' association of gaze allocation and choice behaviour. In particular, we found that individuals' choice behaviour is better described by a model with a gaze bias compared with one without, and that individuals' gaze bias estimates predict their behaviour on several measures (Fig. 6). Third, our analyses span both multiple choice set sizes and choice domains.

While we have shown that the GLAM accurately captures individuals' choice behaviour in choice situations with two and three choice alternatives, it also extends to choices involving many more options, scenarios that we mostly encounter in our everyday lives. Vending machines, for example, can easily display up to 20 items. We assume that in these multialternative choice situations, both gaze and individual differences will play a prominent role. That is, individuals when confronted with large choice sets do not always look at

all available items (for example, see ref. [41]). A choice model that only considers item values will therefore fail to accurately predict individuals' choice behaviour. Conversely, a model that includes information about individuals' gaze distribution during decision formation will outperform such naive models, because it will better account for the set of items that individuals actually consider for a choice. In addition, we assume behavioural differences between individuals to increase with increasing choice set size. Some individuals, for example, may look at only a few of the available items before making a choice, while others may spend a long time searching for the most highly valued option (as indicated in a previous study[41]). To understand whether there is a common choice mechanism underlying these different types of choice behaviour, it is necessary to test the ability of a model to capture individual choice patterns.

Recently, it was also shown that single-trial electroencephalography components reflecting attention in simple perceptual decision-making tasks explain the variance in single-trial evidence accumulation rates of the decision-making process[42] and that variability in these components can explain behavioural differences between individuals[43]. Two recent studies also provided empirical evidence that value-driven activity in the orbitofrontal cortex of monkeys is modulated by fixation location when they viewed reward-associated visual cues in a free-viewing paradigm[44,45]. Together, these studies provide neurobiological evidence of the influence of visual fixations on the process of decision formation. Ultimately, a better understanding of these computations will be central to building holistic models of the choice process and for advancing existing choice frameworks. In addition, it might also help us to better understand the origin of the behavioural variability that we observe within and between individuals as well as the specific functional form of the underlying computational mechanisms linking gaze allocation and choice behaviour.

Methods

Datasets, tasks, procedure and preprocessing. We reanalysed four datasets previously published by Krajbich and colleagues[10], Krajbich and Rangel[9], Folke and colleagues[6] and Tavares and colleagues[21]. An overview of the corresponding tasks and procedures is given in Fig. 1 and in the "Datasets and task overview" section. A more detailed description of the datasets can be found in Supplementary Methods 2 or in the original publications.

Additional processing. The original studies used different scales of item value (that is, liking rating between −10 and +10, willingness-to-pay, angular distance to target line segment). We linearly rescaled all values to a common scale from 1 to 10 so that model parameters are comparable across datasets. For the Tavares 2017 data, values were rescaled so that higher values indicate lower angular distance to the target. Furthermore, for each trial, we computed relative gaze g_i as the sum of gazes towards this item, divided by the total sum of gazes to all items in that trial for each item.

GLAM details. The GLAM belongs to the class of linear stochastic race models[32,33]. It assumes accumulation of noisy evidence in favour of each alternative i, and that choices are determined by the first accumulator that reaches a common boundary b (which we set to 1). In particular, we define the accumulated relative evidence E_i in favour of alternative i, as a stochastic process that changes at each time point t according to equation (1):

$$E_i(t) = E_i(t-1) + vR_i + N(0,\sigma^2), \text{ with } E_i(0) = 0 \quad (1)$$

E_i consists of two separate components: a drift term R_i and zero-centred normally distributed noise with standard deviation σ. The overall speed of the accumulation is governed by the velocity parameter v. The drift term R_i describes the average amount of relative evidence for item i that is accumulated at each time point t. We define the relative evidence R_i^* as the difference in the stationary absolute evidence signal A_i of item i and the maximum absolute evidence of all other items J according to equation (2):

$$R_i^* = A_i - \max_J(A_J) \quad (2)$$

The gaze bias mechanism is implemented in the absolute evidence signal A_i. Similar to the aDDM, absolute evidence signals are assumed to be proportional to the item value r_i, and, crucially, switch between two different states during the trial: an unbiased state, when an item is currently looked at, and a biased state, when gaze is directed towards a different item. Therefore, on average, A_i is a linear combination of two terms weighted by relative gaze g_i according to equation (3):

$$A_i = g_i r_i + (1-g_i)\gamma r_i \quad (3)$$

Here, γ ($\gamma \leq 1$) is the gaze bias parameter that determines the strength of the downweighting during the biased state. If $\gamma = 1$, there is no difference between the biased and unbiased state, producing no gaze bias. If $\gamma < 1$, the absolute evidence signal is discounted, resulting in a gaze bias. If $\gamma < 0$, the sign of the evidence signal changes, thereby leaking evidence, when the item is not fixated on. This leakage mechanism is supported by a recent empirical study[22].

Importantly, by computing an average absolute evidence signal over the two states (equation (3)), each accumulator E_i has constant drift, allowing the use of an analytical solution for its first passage time density (equation (6)). Thereby, the GLAM is statistically and computationally tractable.

Note that the range of possible R_i^* (equation (2)) depends on the participants' use of the item value scale. That is, if the item values r_i only cover a narrow range of possible values on the given scale, relative evidence values R_i^* will likewise be small, whereas they will be large if the participant utilizes the entire range of the value scale. The GLAM assumes an adaptive representation of the relative evidence signals that is compensating for the participants' use of the value scale and thereby sensitive to marginal differences in the relative evidence, particularly to values close to 0 (where the absolute evidence signal for one item is only marginally different to the maximum of all others). To this end, a logistic transform $s(x)$, with scaling parameter τ is applied as follows:

$$s(x) = \frac{1}{1+\exp(-\tau x)} \quad (4)$$

$$R_i = s(R_i^*) \quad (5)$$

The first passage time density $f_i(t)$ of a single linear stochastic accumulator E_i, with decision boundary b, is given by the inverse Gaussian distribution[46] as follows:

$$f_i(t) = \left[\frac{\lambda}{2\pi t^3}\right]^{\frac{1}{2}} \exp\left\{\frac{-\lambda(t-\mu)^2}{2\mu^2 t}\right\}, \text{ with } \mu = \frac{b}{vR_i} \text{ and } \lambda = \frac{b^2}{\sigma^2} \quad (6)$$

However, this density does not take into account that there are multiple accumulators in each trial racing towards the same boundary. As soon as any of these accumulators crosses the boundary, a choice is made and the trial ends. For this reason, $f_i(t)$ must be corrected for the probability that any other accumulator crosses the boundary first. The probability that a single accumulator crosses the boundary before t is given by its cumulative distribution function $F_i(t)$ as follows:

$$F_i(t) = \Phi\left(\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}-1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}+1\right)\right) \quad (7)$$

where $\Phi(x)$ is the standard normal cumulative distribution function. Hence, the joint probability $p_i(t)$ that accumulator E_i crosses b at time t, and that no other accumulator E_i has reached b first, is given by equation (8):

$$p_i(t) = f_i(t)\prod_J (1-F_i(t)) \quad (8)$$

Importantly, all of the GLAM's parameters could be recovered to a satisfying degree without bias (see Supplementary Methods 3 for detailed a procedure and results).

GLAM parameter estimation. All models were implemented in a Bayesian framework using the Python library PyMC3[47]. The full GLAM has four parameters (v, γ, σ, τ). We placed uninformative, uniform priors between sensible limits on all parameters as follows:

$$v \sim \text{Uniform}(1^{-10}, 0.01)$$
$$\gamma \sim \text{Uniform}(-10, 1)$$
$$\sigma \sim \text{Uniform}(1^{-10}, 0.02)$$
$$\tau \sim \text{Uniform}(0, 5)$$

The γ parameter has a natural upper bound at 1 (no gaze bias). The τ parameter has a natural lower bound at 0 (no sensitivity to differences in relative evidence R_i^*).

The GLAM variant without gaze bias used the specification of the absolute evidence signal A_i from the full variant (equation (3)) and fixed γ at a value of 1, resulting in no influence of gaze on the drift term.

To reduce the influence of erroneous responses (for example, when the participant presses a button by accident or has a lapse of attention during the task) on parameter estimation, we explicitly included a model of contaminant processes

in all estimation procedures. In line with existing DDM modelling toolboxes[48], we assumed a fixed 5% rate of erroneous responses, which we model as a participant-specific uniform likelihood distribution $u_s(t)$. This contaminant likelihood describes the probability of a random choice for any of the N available choice items at a random time point in the interval of empirically observed RTs as follows (see also refs. [48,49]):

$$u_s(t) = \frac{1}{N(\max(rt_s) - \min(rt_s))} \quad (9)$$

The resulting choice likelihood is then given by the following:

$$l_i(t) = 0.95 \cdot p_i(t) + 0.05 \cdot u_s(t) \quad (10)$$

Models were fit using Markov–Chain–Monte–Carlo sampling. We used the default implementation of the No-U-Turn-Sampler (NUTS[50]) in PyMC 3.4.1. We sampled two chains with 500 tuning samples that were discarded, and 2,000 posterior samples to estimate the model parameters. If the sampler did not converge as indicated by the Gelman–Rubin statistic ($|\hat{R} - 1| < 0.05$) or if the number of effective samples was low (<100), all of this participant's models were re-estimated using more robust but less efficient Metropolis sampling (two chains, with 10,000 samples each). Again, convergence was diagnosed using the Gelman–Rubin statistic. Convergence was achieved for all models. Reported parameter estimates are maximum a posteriori estimates.

Model simulations. Choice and RT data were simulated from the GLAM according to the following procedures. Each trial in the left-out dataset (all odd-numbered trials) was repeated 50 times. For every trial, the model used the observed item values and gaze distributions. With a fixed rate of 5% the simulation produced a random choice and RT between the participant's minimum and maximum observed RT (see equations (9) and (10)). With a rate of 95% the choice and RT were simulated from the GLAM.

Parametric statistical tests. Assumptions of normality and homoscedasticity were tested for all reported parametric tests. If the normality assumption was violated, we report results from non-parametric tests (Spearman's ρ, Mann–Whitney U). We report rank-biserial correlations r as effect size measures for U-tests. If the homoscedasticity assumption was violated, we report results from the more robust Welch's t-test. All tests were two-tailed.

Mixed-effects modelling. All mixed-effects models reported in the manuscript across datasets were implemented and estimated using the bambi Python library[51]. Bambi automatically generates weakly informative priors for all model terms by default[52]. We sampled two chains, with 20,000 samples each, using NUTS. Convergence was diagnosed using the Gelman–Rubin criterion ($|\hat{R} - 1| < 0.05$) for all analyses. We declare fixed effects as statistically meaningful either when the 95% HDI excludes zero or when 95% of the posterior density is above (below) zero (see also ref. [4]). In the latter case, we also report the proportion of the posterior mass above (below) zero, directly indicating the posterior probability of the effect being larger (smaller) than zero (see also ref. [53]).

Software. All analyses were performed in Python, using the NumPy and SciPy[54], Pandas[55], Statsmodels[56], PyMC3[47], bambi[51] and Theano[57] libraries. We used Matplotlib[58] for visualization.

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

Data availability

All datasets are available at http://www.github.com/glamlab/gaze-bias-differences. The Folke 2016 dataset[59] is originally available at figshare: https://doi.org/10.6084/m9.figshare.3756144.v2.

Code availability

All analyses and figures can be reproduced using the datasets, scripts and GLAM resources that are available at http://www.github.com/glamlab/gaze-bias-differences.

Received: 4 December 2017; Accepted: 6 March 2019;
Published online: 15 April 2019

References

1. Von Neumann, J. & Morgenstern, O. *Theory of Games and Economic Behavior* (Princeton Univ. Press, 1944).
2. Luce, R. D. & Raiffa, H. *Games and Decisions: Introduction and Critical Survey* (Wiley, 1957).
3. Armel, K. C., Beaumel, A. & Rangel, A. Biasing simple choices by manipulating relative visual attention. *Judgm. Decis. Mak.* **3**, 396–403 (2008).
4. Cavanagh, J. F., Wiecki, T. V., Kochar, A. & Frank, M. J. Eye tracking and pupillometry are indicators of dissociable latent decision processes. *J. Exp. Psychol. Gen.* **143**, 1476–1488 (2014).
5. Fiedler, S. & Glöckner, A. The dynamics of decision making in risky choice: an eye-tracking analysis. *Front. Psychol.* **3**, 335 (2012).
6. Folke, T., Jacobsen, C., Fleming, S. M. & De Martino, B. Explicit representation of confidence informs future value-based decisions. *Nat. Hum. Behav.* **1**, 0002 (2017).
7. Glöckner, A. & Herbold, A.-K. An eye-tracking study on information processing in risky decisions: evidence for compensatory strategies based on automatic processes. *J. Behav. Decis. Mak.* **24**, 71–98 (2011).
8. Konovalov, A. & Krajbich, I. Gaze data reveal distinct choice processes underlying model-based and model-free reinforcement learning. *Nat. Commun.* **7**, 12438 (2016).
9. Krajbich, I. & Rangel, A. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proc. Natl Acad. Sci. USA* **108**, 13852–13857 (2011).
10. Krajbich, I., Armel, C. & Rangel, A. Visual fixations and the computation and comparison of value in simple choice. *Nat. Neurosci.* **13**, 1292–1298 (2010).
11. Krajbich, I., Lu, D., Camerer, C. & Rangel, A. The attentional drift-diffusion model extends to simple purchasing decisions. *Front. Pyschol.* **3**, 193 (2012).
12. Pärnamets, P. et al. Biasing moral decisions by exploiting the dynamics of eye gaze. *Proc. Natl Acad. Sci. USA* **112**, 4170–4175 (2015).
13. Roe, R. M., Busemeyer, J. R. & Townsend, J. T. Multialternative decision field theory: a dynamic connectionist model of decision making. *Psychol. Rev.* **108**, 370 (2001).
14. Shimojo, S., Simion, C., Shimojo, E. & Scheier, C. Gaze bias both reflects and influences preference. *Nat. Neurosci.* **6**, 1317–1322 (2003).
15. Stewart, N., Hermens, F. & Matthews, W. J. Eye movements in risky choice. *J. Behav. Decis. Mak.* **29**, 116–136 (2016).
16. Stewart, N., Gächter, S., Noguchi, T. & Mullett, T. L. Eye movements in strategic choice. *J. Behav. Decis. Mak.* **29**, 137–156 (2016).
17. Vaidya, A. R. & Fellows, L. K. Testing necessary regional frontal contributions to value assessment and fixation-based updating. *Nat. Commun.* **6**, 10120 (2015).
18. Tsetsos, K., Chater, N. & Usher, M. Salience driven value integration explains decision biases and preference reversal. *Proc. Natl Acad. Sci. USA* **109**, 9659–9664 (2012).
19. Milosavljevic, M., Navalpakkam, V., Koch, C. & Rangel, A. Relative visual saliency differences induce sizable bias in consumer choice. *J. Consum. Psychol.* **22**, 67–74 (2012).
20. Towal, R. B., Mormann, M. & Koch, C. Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proc. Natl Acad. Sci. USA* **110**, E3858–E3867 (2013).
21. Tavares, G., Perona, P. & Rangel, A. The attentional drift diffusion model of simple perceptual decision-making. *Front. Neurosci.* **11**, 468 (2017).
22. Ashby, N. J. S., Jekel, M., Dickert, S. & Glöckner, A. Finding the right fit: a comparison of process assumptions underlying popular drift-diffusion models. *J. Exp. Psychol. Learn. Mem. Cogn.* **42**, 1982–1993 (2016).
23. Fisher, G. An attentional drift diffusion model over binary-attribute choice. *Cognition* **168**, 34–45 (2017).
24. Gluth, S., Spektor, M. S. & Rieskamp, J. Value-based attentional capture affects multi-alternative decision making. *eLife* **7**, e39659 (2018).
25. Ratcliff, R. A theory of memory retrieval. *Psychol. Rev.* **85**, 59–108 (1978).
26. Ratcliff, R., Smith, P. L., Brown, S. D. & McKoon, G. Diffusion decision model: current issues and history. *Trends Cogn. Sci.* **20**, 260–281 (2016).
27. Grandy, T. H., Lindenberger, U. & Werkle-Bergner, M. When group means fail: can one size fit all? Preprint at *bioRxiv* https://doi.org/10.1101/126490 (2017).
28. Lewandowsky, S. & Farrell, S. *Computational Modeling in Cognition: Principles and Practice* (SAGE Publications, 2010).
29. Hayes, K. J. The backward curve: a method for the study of learning. *Psychol. Rev.* **60**, 269–275 (1953).
30. Itti, L. & Koch, C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* **40**, 1489–1506 (2000).
31. Becker, G. M., DeGroot, M. H. & Marschak, J. Measuring utility by a single-response sequential method. *Behav. Sci.* **9**, 226–232 (1964).
32. Tillman, G. The racing diffusion model of speeded decision making. Preprint at *PsyArXiv* https://doi.org/10.31234/osf.io/xuwbk (2017).
33. Usher, M., Olami, Z. & McClelland, J. L. Hick's Law in a stochastic race model with speed–accuracy tradeoff. *J. Math. Psychol.* **46**, 704–715 (2002).
34. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
35. Lopez-Persem, A., Domenech, P. & Pessiglione, M. How prior preferences determine decision-making frames and biases in the human brain. *eLife* **5**, e20317 (2016).
36. Krajbich, I. Accounting for attention in sequential sampling models of decision making. *Curr. Opin. Psychol.* **29**, 6–11 (2019).

37. Smith, S. M. & Krajbich, I. Gaze amplifies value in decision making. *Psychol. Sci.* **30**, 116–128 (2019).
38. Ratcliff, R., Thapar, A. & McKoon, G. Individual differences, aging, and IQ in two-choice tasks. *Cognit. Psychol.* **60**, 127–157 (2010).
39. Ratcliff, R., Thapar, A. & McKoon, G. Aging and individual differences in rapid two-choice decisions. *Psychon. Bull. Rev.* **13**, 626–635 (2006).
40. Smith, S. M. & Krajbich, I. Attention and choice across domains. *J. Exp. Psychol. Gen.* **147**, 1810–1826 (2018).
41. Reutskaja, E., Nagel, R., Camerer, C. F. & Rangel, A. Search dynamics in consumer choice under time pressure: an eye-tracking study. *Am. Econ. Rev.* **101**, 900–926 (2011).
42. Nunez, M. D., Srinivasan, R. & Vandekerckhove, J. Individual differences in attention influence perceptual decision making. *Front. Psychol.* **8**, 18 (2015).
43. Nunez, M. D., Vandekerckhove, J. & Srinivasan, R. How attention influences perceptual decision making: single-trial EEG correlates of drift-diffusion model parameters. *J. Math. Psychol.* **76**, 117–130 (2017).
44. Hunt, L. T. et al. Triple dissociation of attention and decision computations across prefrontal cortex. *Nat. Neurosci.* **21**, 1471 (2018).
45. McGinty, V. B., Rangel, A. & Newsome, W. T. Orbitofrontal cortex value signals depend on fixation location during free viewing. *Neuron* **90**, 1299–1311 (2016).
46. Wald, A. *Sequential Analysis* (Courier Corp., 1973).
47. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2**, e55 (2016).
48. Wiecki, T. V., Sofer, I. & Frank, M. J. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in Python. *Front. Neuroinform.* **7**, 14 (2013).
49. Ratcliff, R. & Tuerlinckx, F. Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.* **9**, 438–481 (2002).
50. Hoffman, M. D. & Gelman, A. The No-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* **15**, 1593–1623 (2014).
51. Yarkoni, T. & Westfall, J. Bambi: a simple interface for fitting Bayesian mixed effects models. Preprint at *OSF Preprints* https://doi.org/10.31219/osf.io/rv7sn (2016).
52. Westfall, J. Statistical details of the default priors in the Bambi library. Preprint at *arXiv* https://arxiv.org/abs/1702.01201 (2017).
53. Kruschke, J. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan* (Academic Press, 2014).
54. Oliphant, T. E. Python for scientific computing. *Comput. Sci. Eng.* **9**, 10–20 (2007).
55. McKinney, W. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (O'Reilly Media, 2012).
56. Seabold, S. & Perktold, J. Statsmodels: econometric and statistical modeling with Python. In *Proc. 9th Python in Science Conference* (Eds van der Walt, S. & Millman, J.) 57–61 (SciPy, 2010).
57. The Theano Development Team. Theano: a Python framework for fast computation of mathematical expressions. Preprint at *arXiv* https://arxiv.org/abs/1605.02688 (2016).
58. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
59. Folke, T. Explicit representations of confidence informs future value-based decisions. *Figshare* https://doi.org/10.6084/m9.figshare.3756144.v2(2016).

Acknowledgements

The Junior Professorship of P.N.C.M. as well as the associated Dahlem International Network Junior Research Group Neuroeconomics is supported by Freie Universität Berlin within the Excellence Initiative of the German Research Foundation (DFG). Further support for P.N.C.M. is provided by the WZB Berlin Social Science Center. F.M. is supported by the International Max Planck Research School on the Life Course (LIFE). I.K. is funded by the National Science Foundation Career Award 1554837. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.W.T. and F.M. contributed equally to the manuscript and share first authorship. A.W.T. and F.M. conceived of the GLAM, implemented all visualizations of the experimental procedures and performed all formal data analyses. A.W.T. and F.M. also co-wrote all software that was used in the data analyses that underlies the GLAM. A.W.T. and F.M. wrote the original draft of the manuscript, and I.K., H.R.H. and P.N.C.M. reviewed and edited the manuscript. Funding for this work was acquired by P.N.C.M. The work was supervised by H.R.H. and P.N.C.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0584-8.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.N.C.M.

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

nature research

Corresponding author(s): Peter N. C. Mohr

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No new data was collected in this study. Therefore, no software was used in the collection of the data.
Data analysis	All analyses were performed in Python 3.5.2, using the NumPy 1.13 and SciPy 1.0 (Van der Walt, Colbert & Varoquaux, 2011), Pandas 0.21 (McKinney, 2010), Statsmodels 0.8 (Skipper & Perktold, 2010), PyMC 3.4.1 (Salvatier et al., 2016) and Theano 1.0.1 (Theano Development Team, 2016) libraries. We used Matplotlib 2.1.1 (Hunter, 2007) for visualization. For our modeling we used a custom code (http://www.github.com/glamlab/gaze-bias-differences).

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 upon request. 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

All analyses and figures can be reproduced using the dataset, scripts and GLAM resources that are available at http://www.github.com/glamlab/gaze-bias-differences. The dataset from Folke et al. (Nat Hum Beh, 2016) is licensed under CC BY 4.0 and originally available under https://doi.org/10.6084/m9.figshare.3756144.v2.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Behavioural & social sciences study design

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

Study description	Our study includes choice, reaction time and eye tracking data collected from human participants in four previously published datasets.
Research sample	The dataset from Krajbich et al. (2010) contains data from 39 Caltech students. Only participants who self-reportedly regularly ate the snack foods used in the experiment and without food restrictions could participate. No information about participants' age or gender is provided in the original study. The dataset from Krajbich and Rangel (2011) includes data from 30 Caltech students. Only participants who self-reportedly regularly ate the snack foods used in the participants and without food restrictions could participate. No information about participants' age or gender is provided in the original study. The dataset from Folke et al. (2016) includes data from 24 participants (17 females, age 21-38). Participants were required to fast for 4 hours before taking part in the experiment. The dataset from Tavares et al. (2017) includes data from 25 participants (10 female, mean age 23) which included Caltech students and staff as well as members of the surrounding community.
Sampling strategy	We have no information about the sampling strategies used in the original studies.
Data collection	The original studies recorded participants' choices and eye movements using a task presented on a computer. Eye movements were recorded using an eye tracker. We do not have any information about the experimenters presence in the room or their blindness to the research hypotheses.
Timing	We do not have any information about start and end of data collection in the original datasets.
Data exclusions	We did not exclude any data from the dataset that we obtained from the original authors.
Non-participation	We do not have any information about drop-outs or declined participation in the original datasets.
Randomization	We allocated data into separate groups once in our analysis, when performing an out of sample prediction exercise. Here, we split the data into a training and test data set, respectively including the even and odd numbered experiment trials.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	The dataset from Krajbich et al. (2010) contains data from 39 Caltech students. Only participants who self-reportedly regularly ate the snack foods used in the experiment and without food restrictions could participate. No information about participants' age or gender is provided in the original study. The dataset from Krajbich and Rangel (2011) includes data from 30 Caltech students. Only participants who self-reportedly regularly ate the snack foods used in the participants and without food restrictions could participate. No information about participants' age or gender is provided in the original study. The dataset from Folke et al. (2016) includes data from 24 participants (17 females, age 21-38). Participants were required to fast for 4 hours before taking part in the experiment. The dataset from Tavares et al. (2017) includes data from 25 participants (10 female, mean age 23) which included Caltech students and staff as well as members of the surrounding community.
Recruitment	We have no information about the recruitment process in the original studies.