Approximate Number System and Incentive Motivation

INTRODUCTION

Value-based decision making and quantitative judgments have been studied separately in the past. This work aims to integrate the two. Researchers have been interested in exploring irrational value-based decision making in the past, which manifests through errors such as endowment effect and loss aversion. Separately, researchers have also studied number cognition, numerosity discrimination, approximate number system (ANS), ANS acuity and how it relates to math performance. Since ANS acuity is related to decision making in real-world situations, it is important to study the confluence of two.

The current study explores the impact of reward on attention and thereby on number judgements. Influence of reward on decision making has been extensively studied before. Platt and Glimcher (1999) found that decision-theoretic model better explains response selection as compared to the traditional neurophysiological models. In particular, both expected gain and response probability increase neuron activation in the posterior parietal cortex. Expected gain reflects the expected utility/gain derived out of exercising a chosen option, and response probability reflects the chances of such gain materializing. The product of these two quantities determines the economic value of each option that is part of the decision making process. Moreover, the influence of this gain expectation on choice and neuron activation is correlated. Platt (2002) reviews this study along with others to suggest that gradual evolution of decision signals in higher sensory-motor areas is what underlies choice of one option over others in decision making.

Gold and Shadlen (2007) present a theoretical framework for decision making that incorporates our prior knowledge, existing evidence, and the value we assign to different options. Authors utilize signal detection theory and sequential analysis to provide such a framework. Application of this framework to perceptual tasks has revealed the distinction between encoding of sensory input and accumulation of evidence over time to reach a decision.

Delving deeper into the decision making mechanism, Platt and Plassmann (2014), in their chapter, describe decision making as a process guided by underlying value representations and break down value-based decision making into a three-stage model. During stage 1, subjective value signals are formulated for each option and they take into account relevant features of each option. Predicted value signals are the output of this stage and estimate the value that the user can derive from exercising each option. In the second stage, these predicted value signals are passed to the motor system, which formulates action valuation signals. These action value signals lead to implementation of the intended choice. The third and final stage is the outcome stage, marked by experience value signals that capture the utility derived by exercising the chosen option.

Impact of rewards on decision accuracy and latency in macaque monkeys has been studied before (Bendiksby & Platt, 2006) and the results indicate that reward can indeed lead to faster reaction times and improved performance, indicating higher attention and motivation during rewarded trials. The role of the lateral intraparietal area (LIP) was found to be crucial in this task. In particular, activation of many LIP neurons to visual cues correlated with the reward size, and independently with the reaction times. Therefore, LIP can independently integrate motivation and attention per the reward size to drive decision making.

Examining the role of attention in the link between reward and decision has been explored through drift diffusion models (DDM) for a while now. For example, Krajbich et al. (2012) fitted the attention drift diffusion model (aDDM) on purchasing data to reveal that aDDM fits reasonably well on fixation, choice, and reaction time data in this case. Authors call for further investigation about the reduced accuracy that aDDM yields when fitted on purchasing data. Krajbich and Rangel (2011) also extend the two-choice DDM to trinary case with high accuracy. The three-choice data reveals that subjects pay attention to all three options but as the time progresses they narrow down attention to the choice of their preference. Even though fixated options are more likely to be chosen, the fixation process and the valuation process are not completely independent of each other.

Gottlieb et al. (2014) review in detail how saccades are the building blocks of gathering evidence that informs decision making. Subjects actively participate in this evidence accumulation process and select the options they deem worthy of attention and saccades. Moreover, saccades are goal-directed and rely on the priority of incoming information as decided by the brain. The different studies reviewed by authors suggest that saccade target selection can be guided by stimulus-reward associations, or by bidirectional relationships with the task at hand, or by uncertainty of the task at hand. For example, monkey studies implicate the role of LIP and the basal ganglia in learning stimulus-reward associations through value-driven attention. The bidirectional relationships, on the other hand, come into play when values on ongoing actions influence saccades, and saccades influence action through option/sensation selection. Finally, uncertainty reduction through gain in information and motivation enhancement through gain in reward influence attention and gaze patterns.

Reward-driven attention can lead to changes in low level salience of visual stimuli. Qin et al. (2021) showed that association with a higher (but not a lower) reward cue in the training phase increased the perceived contrast of the same-side stimulus later on presented in the test phase.

Combining the sequential sampling approach of the DDM with neuroscientific method of eye-tracking, Sheng and colleagues (2020) have shown that loss aversion can result from either a valuation bias that overweighs losses as compared to gains or from a response bias that encourages avoidance of choices with potential for losses. Moreover, there is strong physiological evidence to show this dissociation: gaze allocation selectively showed valuation bias and pupil dilation selectively showed response bias.

Gaze and pupil dilation have also been shown to be separate biomarkers for different biases that lead to the endowment effect (Sheng et al., 2023). The same computational approach that

corroborated both valuation-based and response-based bias accounts in the case of loss aversion, worked well here as well in the case of endowment effect.

DDMs have not only been used to model purchase and value decisions, but also number representations. For example, Ratcliff and McKoon (2018) ran a two-alternative forced choice experiment where subjects had to choose the pattern with a larger number of dots when two dot patterns were presented side-by-side. As expected, the accuracy decreased as task difficulty increased and response times increased. But these relationships and number representations are task dependent. DDMs showcase three independent components that dictate number cognition decisions: boundaries that suggest the amount of information required to reach the decision, speed of information accumulation as given by the drift rate, and non-decision time.

Ratcliff et al. (2015) suggest that when it comes to research on numerosity judgments, both accuracy and response time data should be used together to test numerosity representation theories. This is because the distributions of accuracy and response times are varied across subjects such that choosing only one measure provides an incomplete picture of the numerosity representations. Response times and accuracy do not necessarily have to be correlated negatively in numerosity representation data, and when these dependent variables are modeled to yield drift rate, boundary separation, and non-decision times through the DDM, we get three variables that are independent of each other, wherein accuracy is determined by drift rate and response times by boundary separation.

Park and Starns (2015) make the same argument: weber fraction is a problematic way of measuring ANS acuity since it relies solely on accuracy. The speed-accuracy trade-off during the numerosity judgment tasks ensures that individual differences in this trade-off influence the weber fraction. Fitting a DDM helps in resolving this issue and it provides drift rate which is considered a better measure of ANS acuity.

Malleability of ANS (as measured through Weber fraction) has been explored before (DeWind & Brannon, 2012). Rapid improvements in ANS acuity, that is, a decrease in Weber fraction was found when participants were given trial-by-trial feedback. This improvement plateaued after a session of feedback but persisted after the feedback was removed. DeWind and Brannon (2012) found that one reason for this improvement in ANS was a decreased reliance on surface area as a numerosity cue. Dramkin et al. (2022) extend this work to show that the malleability effects of surface area on ANS (that is, congruency effects between area and weber fraction) are due to response competition and not because of our prior naturalistic experiences.

Odic and colleagues (2014) showed that number judgements and ANS acuity is not just a function of within trial features such as feedback and surface area, but also of preceding trials. A confidence hysteresis effect was observed wherein repeated low-confidence (high-confidence) decisions decrease (increase) performance in the subsequent focal trials. Although ANS acuity improves as children age, it is no reason to think that age is the only factor determining ANS acuity. Temporary changes in weber fraction have been observed in infants as a result of a numerical hysteresis effect. Wang et al. (2018) found that infants are able to discriminate a 2:3

ratio if they receive trial-by-trial feedback and had earlier experience in progressing from easy to hard ratios (but not vice-versa).

Combining the impact of reward on attention and thereby on number judgments with DDM, Dix and Li (2020) explore whether incentive motivation can improve numerosity discrimination and the mechanism that underlie incentive benefits. Some possible mechanisms they explore are the incentive salience hypothesis: stronger perceptual salience of reward-related objects because of dopamine, learning to greater drift rates in the rewarded condition. Top-down attention regulation and lower decision threshold are some other mechanisms that authors suggest for how incentive motivation can impact numerosity judgments.

Findings from this study reveal that reward leads to more accurate numerosity discrimination as measured by drift rate. Higher drift rate implies faster evidence accumulation, which was corroborated by higher pupil dilation since reward cue should increase attention and perceptual salience per incentive salience hypothesis. Mechanism proposed behind higher pupil dilation was that the anterior cingulate cortex allocates higher attention on reward cue and optimizes task engagement through locus coeruleus.

Reward also led to a positive correlation between wins and anticipatory dilation, careful decision making because of higher boundary separation, increase in response time and non-decision time, and higher pupil dilation even during numerosity discrimination and feedback phase. Higher task difficulty, on the other hand, led to higher pupil dilation and longer response times since subjects exercised higher response caution in high conflict trials. Higher task difficulty also led to lower accuracy, higher non-decision time, and lower boundary separation.

The overarching goal of the current study is to check how incentive motivation and ANS interact. We introduce spatial asymmetry (larger reward on correct answer when more dots are on left vs right) to see whether numerosity judgements are susceptible to bias. If so, how much are those judgments driven by a calculation bias or a response bias? In other words, do participants do better in numerosity judgments because they see the dots more clearly, i.e., a perceptual change, or because there is a shift in the decision threshold. Previous work has similarly delineated effects such as loss aversion (Sheng et al., 2020) and endowment effect (Sheng et al., 2023) to show that some participants show the effect because of a valuation bias, as evidenced by gaze asymmetry, whereas other participants show the effect because of a response bias, as evidenced by pupil dilation.

METHODS

Participants

The study was approved by the Institutional Review Board at the University of Pennsylvania. Fifty-three adults participated in the experiment in four different cohorts (six subjects participated in January 2024, ten subjects participated in February 2024, twenty subjects participated in March 2024, and seventeen subjects participated in April 2024). Eye-tracking data was collected only from participants from the February, March, and April cohorts, but eye-tracking

data from one March and one April cohort participant went missing, so eye-tracking files from 45 participants were analyzed. All participants provided informed consent and had normal or corrected-to-normal vision. No participants were excluded for completing fewer than 80% trials or for showing a left- or right-side bias on more than 75% of trials.

Procedure

A total of 52 dot patterns were generated, controlling for dot size, total surface area, sparsity, and field area (DeWind et al., 2015). 13 different dot numerosities (8, 9, 10, 11, 13, 14, 16, 18, 20, 23, 25, 29, 32) for four categories: (constant dot size v/s constant total surface area) X (constant field area v/s constant sparsity). To create pairwise comparisons, each numerosity, n, was paired up with four possible numerosities: $n*2^{(1/1)}$, $n*2^{(1/2)}$, $n*2^{(1/3)}$, $n*2^{(1/6)}$. For example, 8 dots were paired with 16, 11, 10, and 9 dots, whereas 25 dots were paired with 32, and 29 dots, leading to a total of 40 pairwise combinations in each stimuli category. Therefore, 160 pairs in total and 320 after considering presentation order: left vs right.

The experiment utilized PsychoPy 2023.2.3 and an ABBA design. Participants were instructed that they will see two arrays of dots and had to select the pattern with a larger number of dots by pressing the ‘f’ (if more dots in left pattern) or ‘j’ (if more dots in right pattern) key on the keyboard (Fig. 1). Participants were also informed about a performance-dependent bonus of upto 10 dollars in the beginning of the experiment but each participant received a fixed amount of 8.75 dollars at the end of the experiment.

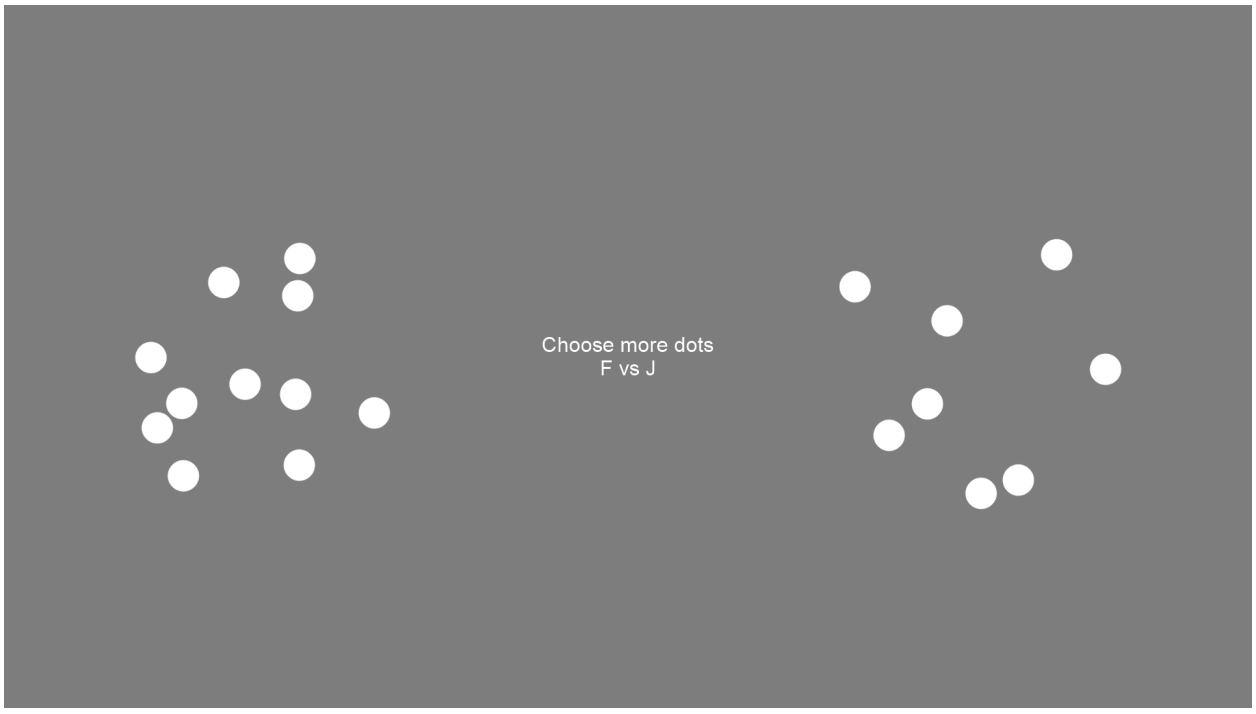

Fig. 1. The dot pattern selection task. The example displays a dot pattern comparison task where participants had to select the pattern with the higher number of dots. Participants chose the left vs. the right pattern by pressing one of two keys.

A and B blocks in the ABBA design corresponded to the reward conditions: participants accumulated 60 (30) points for correct answer when higher numerosity was on the left (right) in block A, and vice versa in block B. 320 stimuli pairs were evenly split between the two A blocks. Same for two B blocks, leading to a total of 640 (160*4) trials. Therefore, participants faced each comparison four times: (in block A vs in block B) X (when more dots were on the left vs on the right).

Each trial started with a 1-second presentation of fixation cross in the middle of the screen. Two stimuli images were then presented side-by-side with this prompt in the middle: “Choose more dots. F vs J”. Participants were given unlimited time to make a decision. After the key response, participants were given the trial feedback about the accuracy (correct vs incorrect) and the reward earned (30/60 vs 0). Participants were also given feedback after each of the four blocks about the total points they earned and could have earned in that block.

Participants in the March and April cohorts were also given a follow-up survey that consisted of the Subjective Numeracy Scale (Fagerlin et al., 2007) to measure perceived ability to perform various mathematical tasks, questions to measure the raw or math scores on SAT or ACT, and Lipkus numeracy scale (Lipkus et al., 2001) to measure actual performance on math questions.

Gaze and Pupil Data Acquisition

Participants were seated about 60 cm in front of the screen in a silent room. Gaze fixation and pupil diameter were sampled at 250 Hz using an eye tracker (Tobii). The eye tracker was synchronized with the experiment software (PsychoPy 2023.2.3).

Processing of gaze data

0.03% of the analysis time windows (from decision onset to decision made) contained fewer than 50% valid gaze position samples, wherein invalid samples were the ones populated with null or with values that situate the gaze outside the screen. These time windows were discarded. The boxes containing the left (150, 270 to 690, 810) and the (1230, 270 to 1770, 810) right dot patterns were identified as the regions of interest on the whole screen (0,0 to 1920, 1080). Total time spent looking at the left and right dot patterns in each analysis time window (from decision onset to decision made) was calculated and referred to as left gaze duration and right gaze duration, respectively. Gaze proportions were calculated for only the analysis time windows that had either the left gaze duration or the right gaze duration strictly greater than zero. Left (right) gaze proportion was calculated as the proportion of gaze directed at the left (right) dot pattern out of the total gaze directed at both dot patterns.

Processing of pupil data

The analysis time window spanned the presentation of fixation cross, decision time period, and presentation of feedback, i.e., 1 s before offer onset to 1 s after decision. There were no trials with fewer than 50% valid pupil samples (i.e., non-null and positive samples) in the analysis time window, so no trials were dropped. We considered the average of left and right diameter for analysis. Extreme pupil samples were excluded following Kret and Sjak-Shie (2019). Missing or

excluded samples were interpolated linearly. The pupil data was then processed with a third-order, low pass, 4 Hz Butterworth filter, and z-scored for each participant.

Baseline correction was done at trial level by subtracting mean pupil value in the baseline time window (spanning presentation of the fixation cross, i.e., 1 s before offer onset) from each value within the analysis time window. Decision-related pupil size was calculated as the mean pupil size in the analysis time window from 1 s before to 1 s after the decision.

DDM

Drift-diffusion models were fit on choice and response time data for each block (A: higher reward on left pattern vs. B: higher reward on right pattern) for each participant. The decision process was assumed to be a noisy evidence-accumulation process unfolding between the upper boundary that governed the choice for the right dot pattern (denoted by 1) and the bottom boundary that governed the choice for the left dot pattern (denoted by 0) with moment-to-moment fluctuations. The accumulation process ends as soon as one of the two response boundaries—either choosing the right dot pattern (upper) or choosing the left dot pattern (lower)—is reached. Through DDM, we jointly determine the choice and the response time by the rate at which evidence accumulates (drift rate, v) and the starting point of the accumulation process (z), relative to the two response boundaries, i.e., 0 and 1.

For each dot comparison trial with R dots in the right dot pattern and L dots in the left dot pattern, we assumed the drift rate, $v \sim w(\log(R/L)) + b + \varepsilon$. Here, w is the coefficient returned by the DDMs, which was assumed to be independent for block A (w(A)) and block B (w(B)). Calculation bias was then calculated as [w(B) - w(A)] / [w(B) + w(A)]. b is an intercept to account for a constant component in the drift rate, i.e., variance in choices and response times that was not related to the ratio of number of dots in the dot patterns. b was also allowed to differ between blocks (i.e., b(A) and b(B)). ε indicates random noise in the diffusion process with a standard normal distribution, i.e., $\varepsilon \sim N(0,1)$.

Evidence-accumulation starting point was denoted by z(A) and z(B) for blocks A and B, respectively. Response bias was calculated as [z(B) - z(A)] / [z(B) + z(A)]. A starting point without any bias towards any boundary would be right in the middle, where z = 0.5, whereas $z<0.5$ would indicate bias towards the left dot pattern and $z>0.5$ would indicate bias towards the right dot pattern. Parameter a, which was allowed to vary by blocks (i.e., a(A) and a(B)) was another output of the DDMs to scale boundary separation. Non-decision time was captured by a parameter t. We used the HSSM package in Python to conduct hierarchical Bayesian estimation for our DDM. We ran two separate chains for each model with 800 samples each and with the first 200 samples as burn-ins.

aDDM

Building on the DDM models, we fit aDDMs using the HSSM package by allowing the weight of the log(R/L) ratio to vary as a function of gaze allocation in both blocks A and B. We assumed that when gaze was on the left dot pattern, the weight of the log(R/L) ratio was w(A),(L) and

w(B),(L) in blocks A and B, respectively, and when gaze was on the right dot pattern, the weight was w(A),(R) and w(B),(R) in blocks A and B, respectively. When the left dot pattern was fixated, we assumed the drift rate, v ~ (w(A),(L))(log(R/L)) in block A and v ~ (w(B),(L))(log(R/L)) in block B. When the right dot pattern was fixated, we assumed the drift rate, v ~ (w(A),(R))(log(R/L)) in block A and v ~ (w(B),(R))(log(R/L)) in block B. Assuming that the proportion of time spent looking at the left dot pattern, or left gaze proportion, was GazeL, and that for the right dot pattern, or right gaze proportion, was GazeR (i.e., 1 − GazeL), the mean drift rate was v ~ GazeL · (w(A),(L))(log(R/L)) + GazeR · (w(A),(R))(log(R/L)) in block A and v ~ GazeL · (w(B),(L))(log(R/L)) + GazeR · (w(B),(R))(log(R/L)) in block B. We fit the aDDM in HSSM in a way similar to DDM except that only trials with valid eye positions were included (see the "Processing of gaze data" section)

RESULTS

No side bias in either block

A generalized linear model was fit to each participants' behavioral data within each block (A vs. B) to capture the side bias and the effect of ratio of number of dots on the right vs. the left. Specifically, we fit: p(chooseright) ~ Φ[$\beta_{(side)}$ + $\beta_{(ratio)}$*log(R/L)], where $\beta_{(side)}$ is the side bias irrespective of the number of dots in the dot patterns, $\beta_{(ratio)}$ is the regression coefficient for the log ratio of the number of dots on the right, R vs. the number of dots on the left, L, and Φ is the cumulative distribution function of the standard normal distribution.

We ran a one-sample t-test to determine whether $\beta_{(side)}$ and $\beta_{(ratio)}$ from each block were significantly different from zero across participants. $\beta_{(side)}$ for block A was not significantly different from zero and the same held true for block B. We expected participants to show a bias towards the right (left) dot pattern over the left (right) dot pattern in block B (A), given the reward associations so this result was unexpected (Block A: $\beta_{(side)}$ M = 0.089, SEM = 0.068, t_{52} = 1.318, p = .193; Block B: $\beta_{(side)}$ M = 0.093, SEM = 0.073, t_{52} = 1.261, p = .213; see Fig. 2). As expected, $\beta_{(ratio)}$ for each block was significantly different from zero, indicating that participants shifted to choosing the right dot pattern over the left dot pattern as the number of dots in the right pattern started to outweigh the number of dots in the left pattern (Block A: $\beta_{(ratio)}$ M = 9.435, SEM = 0.417, t_{52} = 22.611, p < .001; Block B: $\beta_{(ratio)}$ M = 9.538, SEM = 0.438, t_{52} = 21.798, p < .001). We also did paired t-tests to compare $\beta_{(ratio)}$ and $\beta_{(side)}$ between blocks A and B but the difference was not significant (both ps > .706).

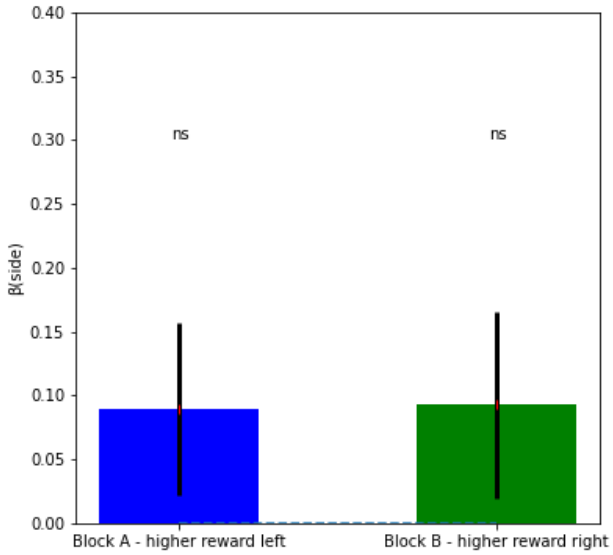

Fig. 2. Mean *β(side)* across participants for each block. Side bias not significant in either block. Error bars denote SEM.

Fitting psychometric curves confirmed our finding that there were no significant differences in *β(ratio)* between the two blocks A vs. B (see Fig. 3).

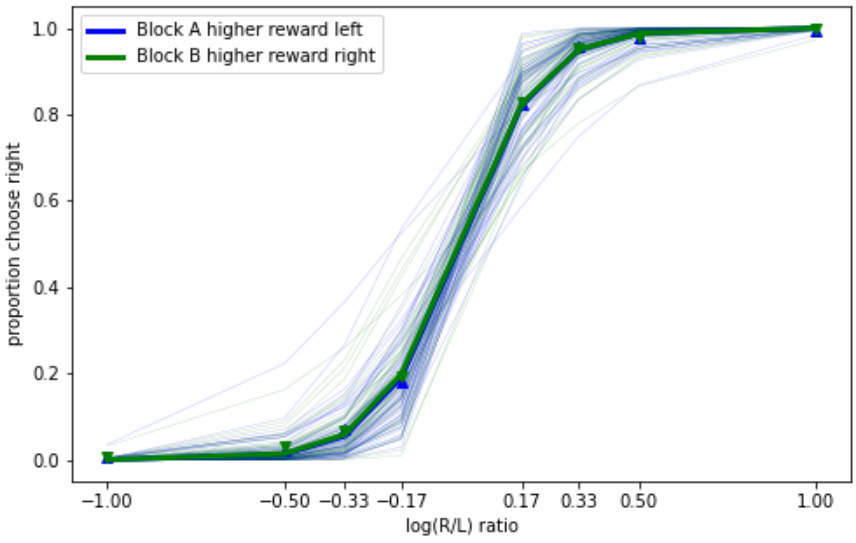

Fig. 3. Psychometric curves for both blocks. The more transparent curves are for individual participants' choose right proportions in each block. The two thicker curves and the markers show choose right proportions averaged across participants for both blocks as a function of log(R/L) ratio.

Response times averaged across participants did not differ significantly between the two blocks for any of the log(R/L) ratios (all *ps* > .065; see Fig. 4). We expected to have a significant effect of ratio on the probability of choosing right. But the absence of a side bias despite incentive motivation was surprising. So, to delve deeper, we fit DDMs.

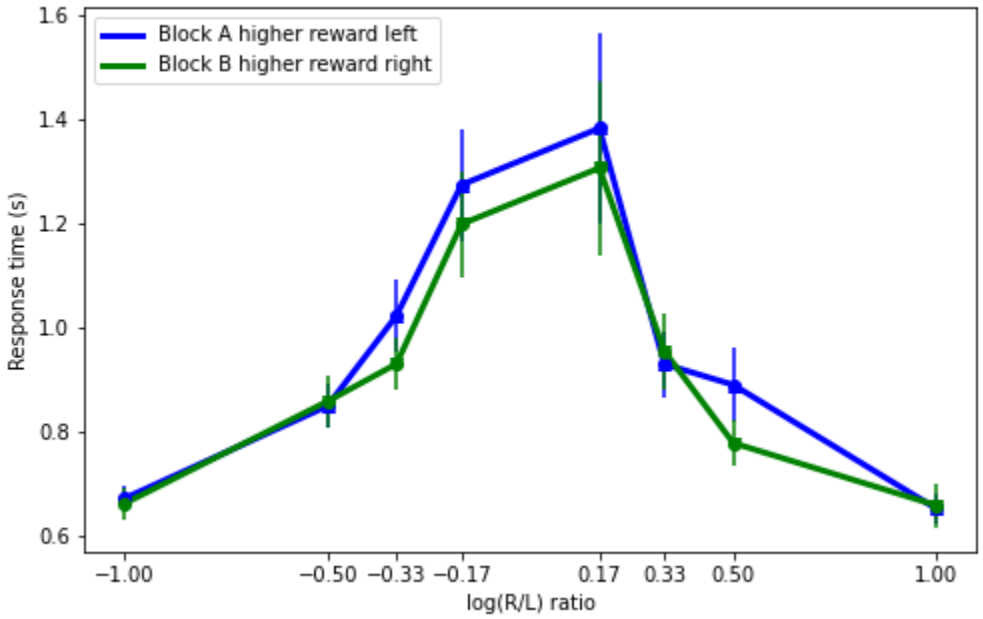

Fig. 4. Response times. Response times for different ratios for both blocks are shown. Error bars represent SEM.

Calculation bias and response bias are dissociable for this numerosity judgment task

Next, we fit DDMs on behavioral data for both blocks A and B. Consider the trial presented in Fig. 1. The same trial was presented in block A and block B, but in block A, choosing the correct pattern, i.e., the left pattern, would have earned the participant 60 points, whereas in block B, choosing the same correct left pattern would have earned the participant 30 points. Therefore, in both blocks, participants should choose the same dot pattern but, perhaps, their response times would be different. Moreover, in trials with higher difficulty level, we might see the incentive motivation of a block act against the more numerous pattern.

Using this framework, we modeled each numerosity judgment as a noisy evidence-accumulation process that drifts between two boundaries (Fig. 5A). Reaching the upper boundary (denoted by 1) meant the culmination of the evidence-accumulation resulting in choice of the right dot pattern. On the other hand, reaching the lower boundary (denoted by 0) meant the end of the evidence-accumulation and choice of the left dot pattern. One would expect participants to choose, in both blocks A and B, the right dot pattern for log(R/L) ratios above 0 and the left dot pattern for log(R/L) ratios below 0.

Drift rate (v), or the speed of evidence accumulation, was determined linearly by a weighting coefficient, w multiplied by the log(R/L) ratio: v ~ w*log(R/L). Weighting coefficient in block A (w(A)) was assumed to be independent of the weighting coefficient in block B (w(B)). Calculation bias was calculated through the contrast index: [w(B) - w(A)] / [w(B) + w(A)]. A starting point at 0.5, i.e., right in the middle would mean no starting bias towards either dot pattern. Since block A (B) offered higher reward on the left (right), the starting point in block A (B), i.e., z(A) (z(B)), would be shifted toward the lower (upper) boundary of the left (right) pattern. So, response bias was calculated through the contrast index as well: [z(B) - z(A)] / [z(B) + z(A)].

In DDMs for both blocks A and B, we also included an intercept variable in the drift rate formula to account for variation in the behavioral data that rose not due to the log(R/L) ratio. Also, we added a, which denotes the boundary separation and t, which denotes the non-decision time.

Since blocks A and B differentially rewarded participants with higher reward for correct answers on the right (left) hand side in block B (A), we expected that participants would have a starting point closer to the upper (lower) boundary in block B (A), resulting in a positive response bias, irrespective of log(R/L) ratio. This is illustrated by the contrast of the starting points between blocks A and B in Fig. 5A and 5B. Variations in drift rate might be more nuanced and depend on the sign and magnitude of log(R/L) ratio. For positive ratios of easy difficulty (Fig. 5A), participants would select the more numerous right dot pattern in both blocks, albeit at a slower pace in block A since block A gets associated with higher reward to the left. This is reflected in Fig. 5A by the higher slope, i.e., higher drift rate for block B vs. block A, which in turn means w(B) > w(A) for such trials. Similarly for negative ratios of easy difficulty, we might expect w(A) > w(B). Interesting are the ratios of higher difficulty level. For example, consider positive ratios of hard difficulty (Fig. 5B), it would be hard to make a judgment on which pattern is more numerous (hence, the low magnitude drift rates) even though the right one is, but participants would be nudged to select the correct right pattern in block B since the right pattern is associated with higher reward, whereas in block A, the incentive motivation might overpower the approximate number system, leading to an error in judgment and selection of the left pattern. Nevertheless, in such difficult trials, we still expect to see w(B) > w(A) for positive ratios and w(A) > w(B) for negative ratios.

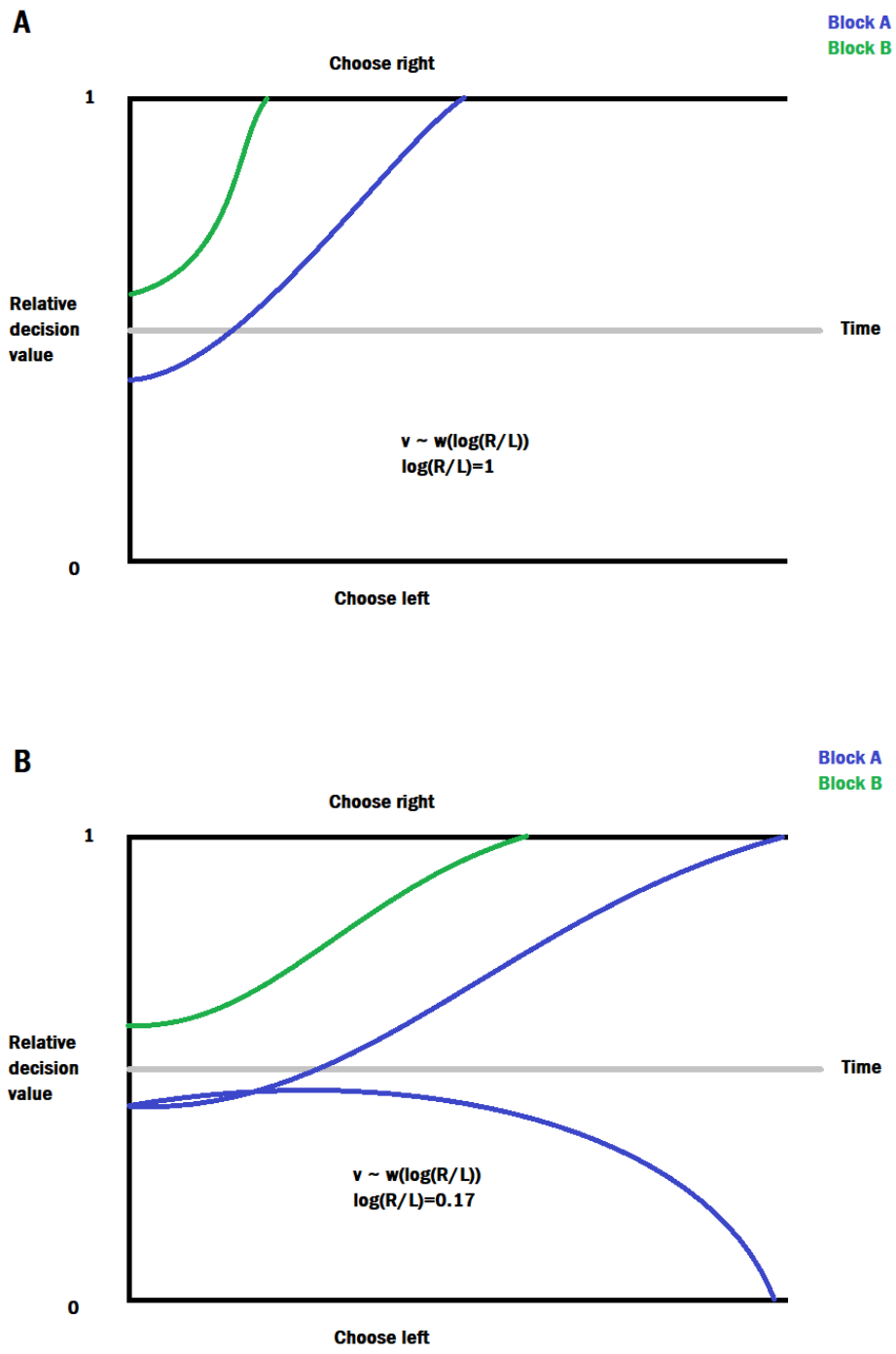

Fig. 5. DDM applied to numerosity judgment task. (A) Blue and green curves indicate the evidence accumulation process (with smoothed noise) in an easy trial with high magnitude log(R/L) ratio for blocks A and B, respectively. **(B)** Blue and green curves indicate the evidence accumulation process (with smoothed noise) in a hard trial with low magnitude log(R/L) ratio for blocks A and B, respectively.

Calculation bias and response bias were calculated using the formulae defined. First, we compare starting points and weighting coefficients between the blocks A and B.

We ran a one-sample t-tests to determine if the starting points z(A) and z(B) were significantly different from 0.5 (no a priori starting bias) across participants. Mean z(A) was not significantly different from 0.5, but mean z(B) was significantly different from 0.5 (Block A: *z(A) M* = 0.500, *SEM* = 0.005, t_{52} = 2.091, *p* = .999; Block B: *z(B) M* = 0.515, *SEM* = 0.006, t_{52} = 2.573, *p* = .013). Moreover, z(B) was significantly different from z(A) across participants (t_{52} = 3.263, *p* = .002; see Fig. 6).

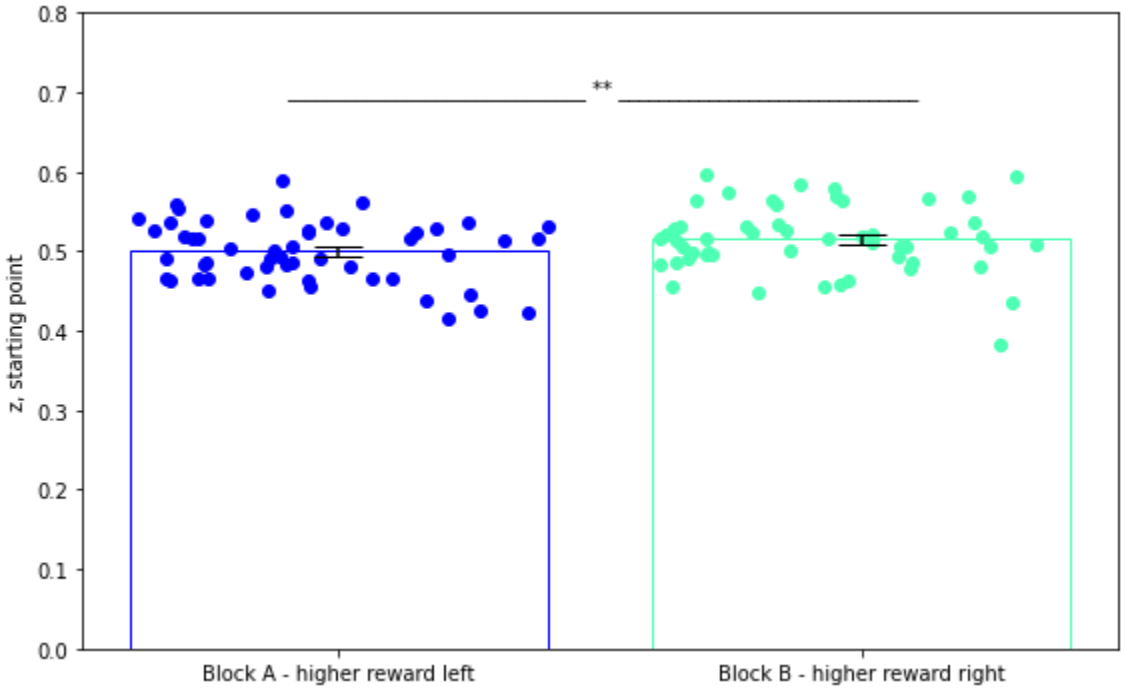

Fig. 6. Starting point, z. Estimated starting points in both blocks. Dots represent individual subjects and bars represent means across subjects. SEM are shown.

We also found that w(B) was significantly different from w(A) across participants (t_{52} = 2.999, *p* = .004; see Fig. 7).

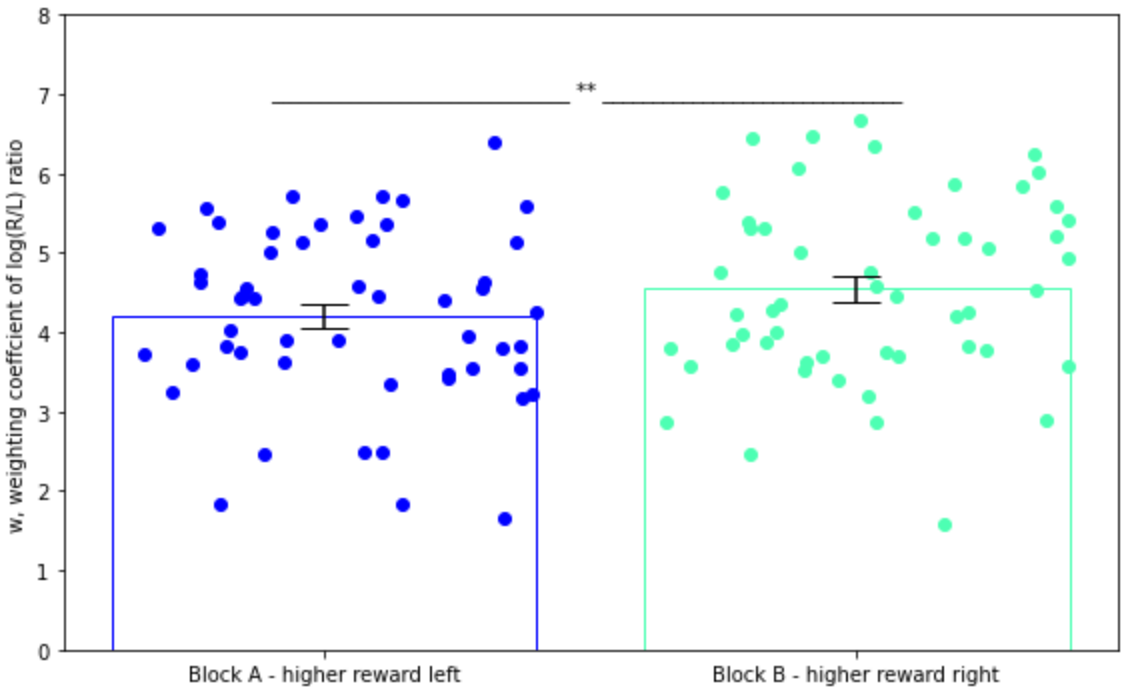

Fig. 7. Weighting coefficient, w. Estimated weighting coefficients of log(R/L) ratio in both blocks. Dots represent individual subjects and bars represent means across subjects. SEM are shown.

Mean calculation bias across participants was significantly different than zero (calculation bias M = 0.040, SEM = 0.015, t_{52} = 2.681, p = .010), and so was response bias (response bias M = 0.015, SEM = 0.005, t_{52} = 3.199, p = .002). This suggests higher weighting coefficient and higher starting point in block B vs. block A. Across participants, calculation bias ranged from −0.285 to 0.382, and response bias ranged from −0.041 to 0.120 (Fig. 8). These two biases were not correlated with each other (r = -0.175, p = 0.211), which means both are necessary to account for variations in the behavioral data.

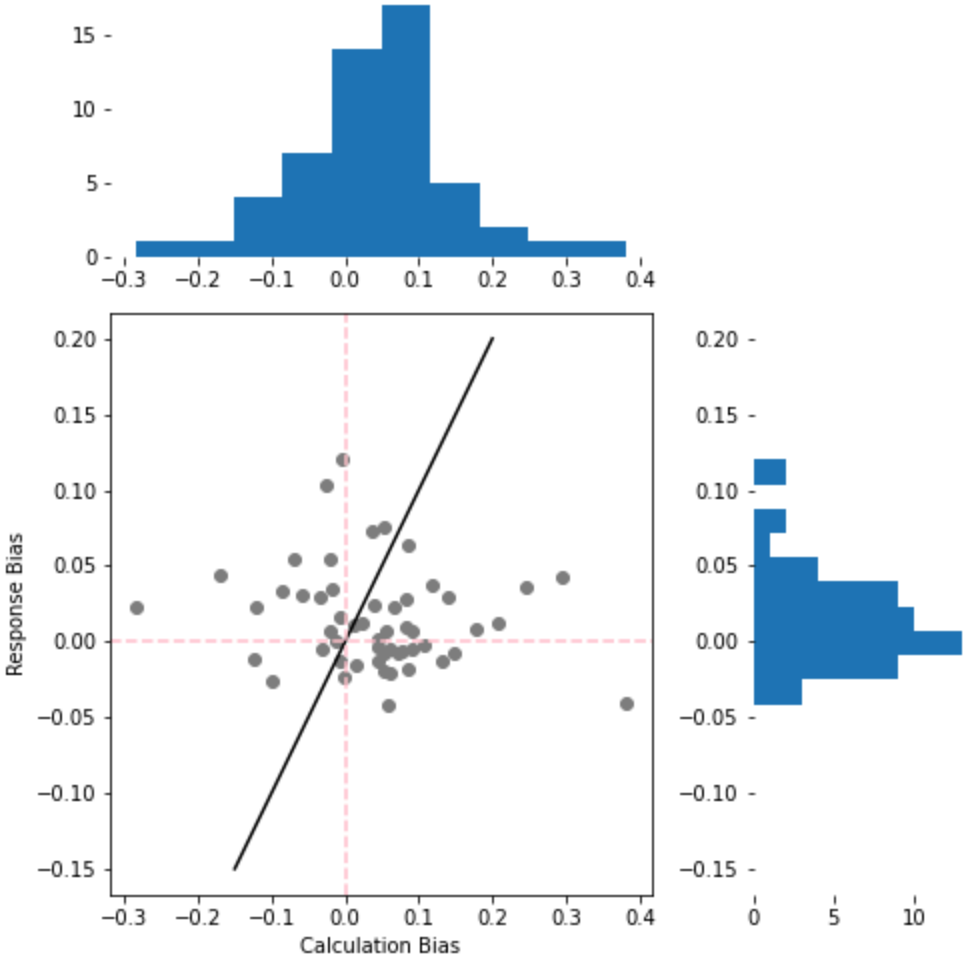

Fig. 8. Calculation and response bias. Each dot is a participant. Identity line is shown to bifurcate participants who were influenced more by one bias than the other. Histograms represent distributions of respective biases.

Calculation bias manifests in gaze allocation

We examined the gaze bias between the left and right dot patterns. This required calculating right gaze proportion as the time spent fixating on the right dot pattern relative to the total time spent fixating on both left and right dot patterns. We then averaged the right gaze proportion for each participant for each block. Across participants, as expected, right gaze proportion was significantly higher in block B (M = 0.530, SEM = 0.010) than in block A (M = 0.517, SEM = 0.010, t_{44} = 2.947, p = .005) but surprisingly right gaze proportion was more than 0.5 in block A as well even though block A is associates higher reward with the left dot pattern.

Next, we examined the relationship between gaze allocation and the two biases: calculation bias and response bias. First, we calculated gaze bias as the difference in right gaze proportion between block B vs. block A for each participant, then we regressed the calculated gaze bias on calculation bias and response bias as follows: gaze bias ~ response bias + calculation bias. Out of the 45 participants for whom we had eye-tracking data available, we removed four

participants from this analysis. These participants were outliers for response or calculation bias (more than two standard deviations away from the respective means). Calculation bias marginally predicted gaze bias (β = -0.103, *SEM* = 0.052, t_{38} = -1.967, p = .057), but response bias did not (β = 0.252, *SEM* = 0.168, t_{38} = 1.496, p = .143; see Fig. 9). It is surprising that on average, we found higher weighting coefficients in block B than in block A, and higher right gaze duration in block B than in block A, but a negative relationship between gaze bias and calculation bias. This means that participants with lower (higher) contrast between blocks B vs. A in terms of the weighting coefficient had a higher (lower) contrast in blocks B vs. A in terms of the right gaze duration. Nevertheless, these findings show that calculation bias, and not response bias, is related to gaze bias.

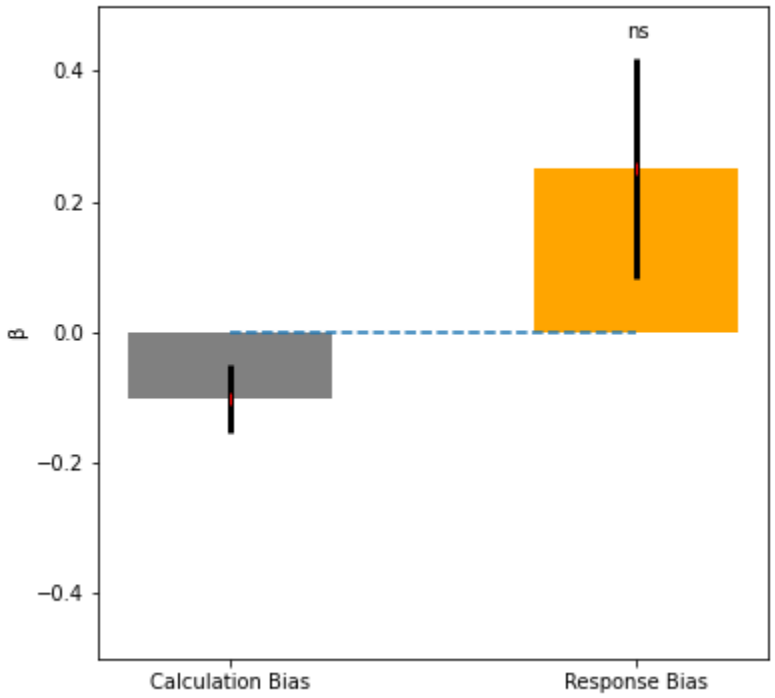

Fig. 9. Gaze bias predicted by calculation bias and response bias. Calculation bias, but not response bias, marginally predicts gaze bias (for the right dot pattern in blocks B vs. A). Error bars denote SEM.

We digged deeper into the relationship between gaze bias and calculation bias using an aDDM framework. We expected that the weighting coefficient of log(R/L) ratio would be higher when the gaze was on the right dot pattern and lower when the gaze was on the left dot pattern. Testing this required fitting aDDMs that allowed the weighting coefficient of log(R/L) ratio to vary with right gaze proportion vs. left gaze proportion (see Fig. 10): v ~ w(right gaze proportion)*log(R/L) + w(left gaze proportion)*log(R/L).

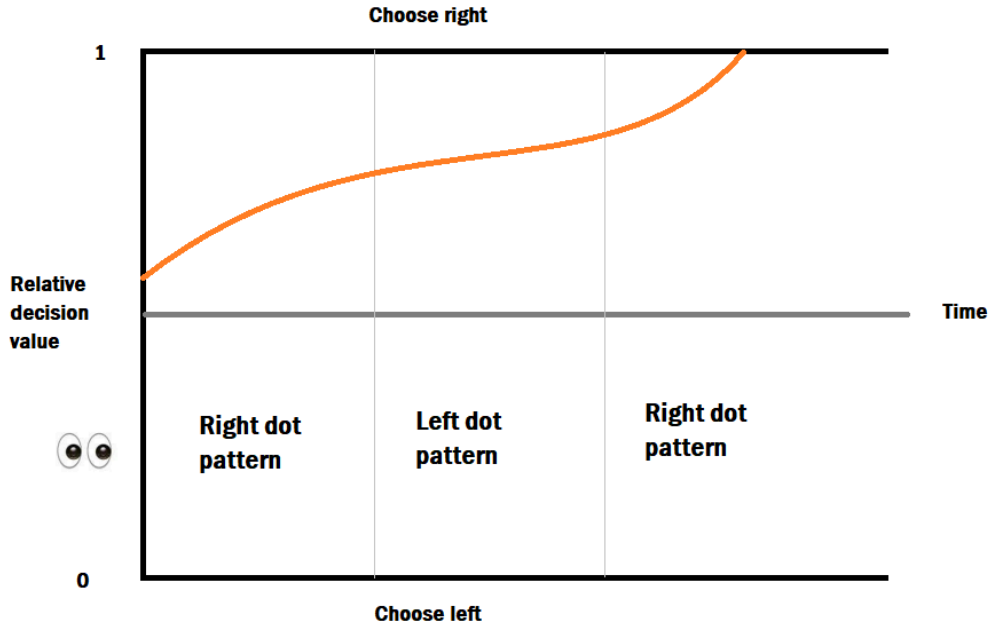

Fig. 10. aDDM modeling. Drift rate of evidence accumulation varies with gaze allocated to the right dot pattern vs. the left dot pattern.

As expected, within block B, subjects assigned more weight to log(R/L) ratio when subjects gazed at the right dot pattern (M = 5.015, SEM = 0.200) compared to when they gazed at the left dot pattern (M = 4.435, SEM = 0.190). Surprisingly, this pattern held in block A, subjects assigned more weight to log(R/L) ratio when they gazed at the right dot pattern (M = 4.881, SEM = 0.233) compared to when they gazed at the left dot pattern (M = 3.920, SEM = 0.239; Fig. 11), even though we expected the opposite since block A associates more reward with the left dot pattern.

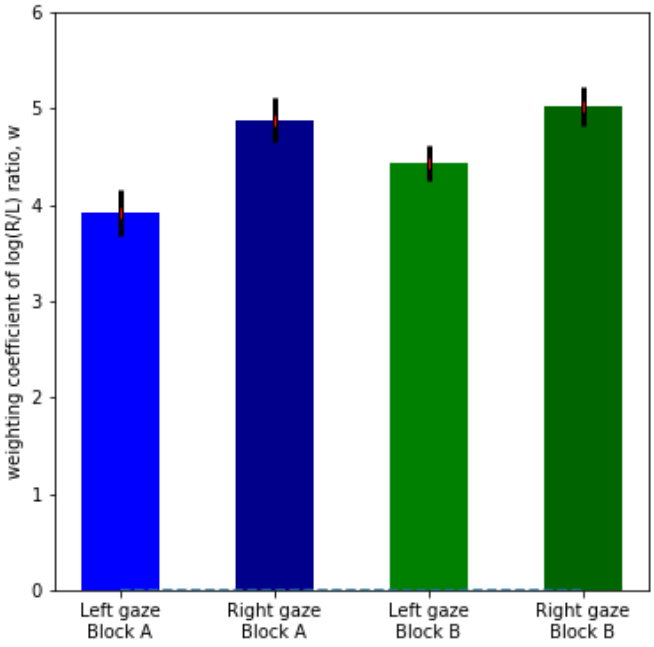

Fig. 11. Weighting coefficients vary with gaze. Weights assigned to log(R/L) ratio in blocks A vs. B when the left vs. the right dot pattern is fixated.

Response bias does not manifest in pupil dilation

Participants, in block A, showed no difference in pupil size when choosing the right pattern (M = -0.001, SEM = 0.001) vs. the left (M = -0.003, SEM = 0.001, t_{44} = -1.038, p = .305). In block B as well, there was no difference in pupil size when choosing the right pattern (M = 0.0002, SEM = 0.001) vs. the left (M = -0.002, SEM = 0.001, t_{44} = -1.215, p = .231). Therefore, in both blocks, choosing right over left did not require more mental effort, even though we expected that participants would exert more mental effort in choosing right (left) over left (right) in block A (B) given the higher reward associated with the left (right) pattern in that block.

These results could have depended on the trial difficulty level, so to test the impact of log(R/L) ratio on pupil dilation, we ran participant-wise regressions with chose right and log(R/L) ratio as predictors and pupil size as the predicted variable. We ran a one-sample t-test to determine whether *β(chose right)* and *β(ratio)* from each block were significantly different from zero across participants. *β(chose right)* and *β(ratio)* for block A were not significantly different from zero and the same held true for block B (Block A: *β(chose right)* M = -0.001, SEM = 0.004, t_{44} = -0.241, p = .810; Block B: *β(chose right)* M = -0.001, SEM = 0.003, t_{44} = -0.220, p = .827; Block A: *β(ratio)* M = 0.004, SEM = 0.003, t_{44} = 1.079, p = .287; Block B: *β(ratio)* M = 0.004, SEM = 0.003, t_{44} = 1.457, p = .152). Adding response time as a third predictor in the model did not show significant influence of chose right, log(R/L) ratio, or response time on pupil size in either block (all ps > .106)

We also tested whether pupil size reflected calculation bias or response bias. Pupillary reactivity bias was calculated for all subjects as pupil size differential in block B - pupil size differential in block A. Pupil size differential for both blocks was calculated as pupil size when choosing right pattern - pupil size when choosing left pattern. We regressed pupillary reactivity bias on calculation bias and response bias. Individual differences in pupillary reactivity bias were not predicted by response biases (Fig. 12, β = -0.086, *SEM* = 0.072, t_{42} = -1.196, p = .238) or calculation biases (β = 0.008, *SEM* = 0.023, t_{42} = 0.326, p = .746). These findings are surprising as they do not show the link between response bias and pupillary reactivity.

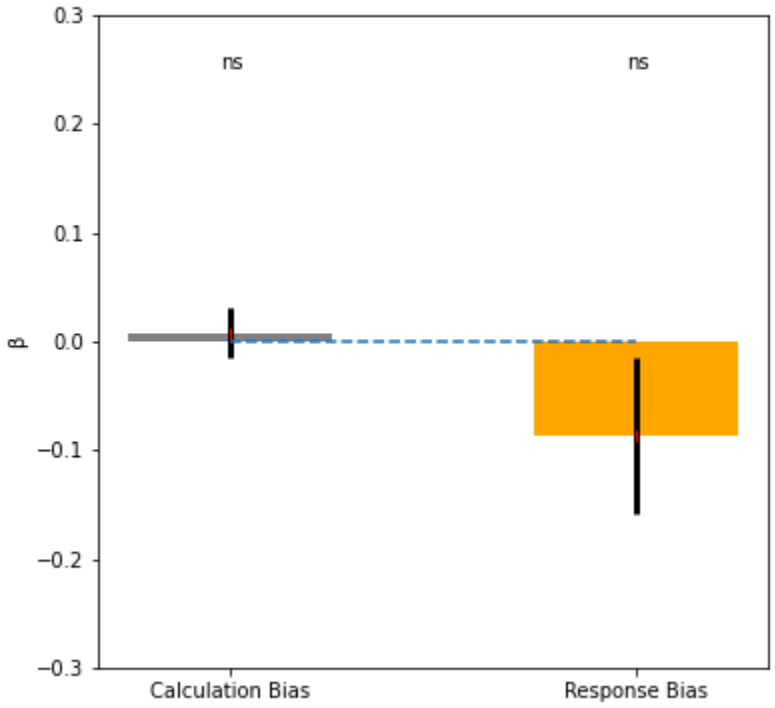

Fig. 12. Pupillary reactivity bias predicted by calculation bias and response bias. Neither calculation bias nor response bias predict pupillary reactivity bias. Error bars denote SEM.

Relationship between pupillary reactivity bias and gaze bias was also examined. Pupillary reactivity bias was not correlated with gaze bias (Fig. 13, r = 0.229, p = 0.129).

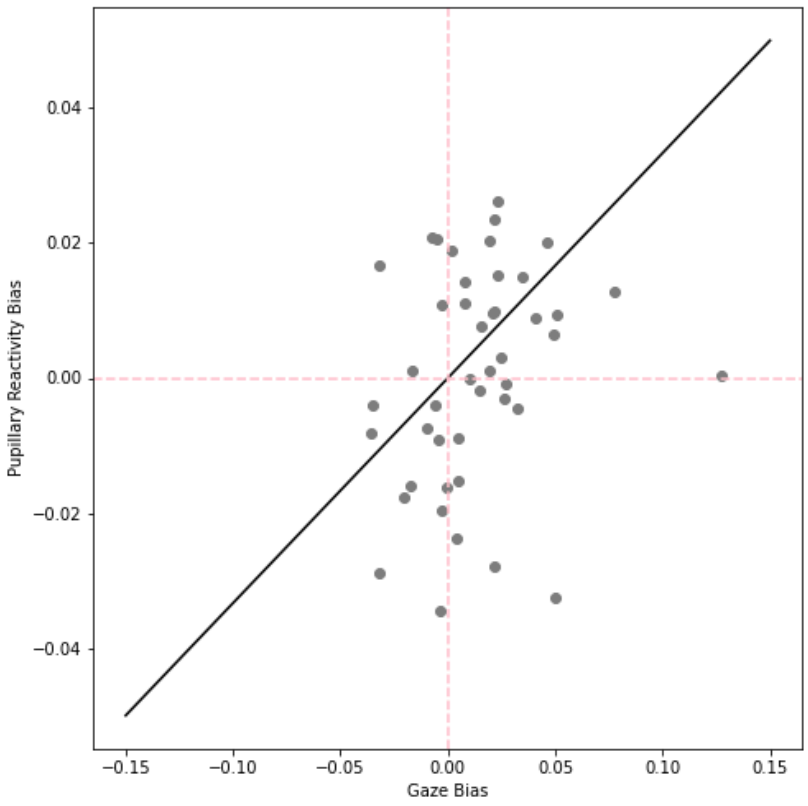

Fig. 13. Pupillary reactivity bias vs. gaze bias. Scatter plot of pupillary reactivity bias vs. gaze bias. Each dot denotes a participant. Identity line is shown to bifurcate participants who were influenced more by one bias than the other.

DISCUSSION

In the current work, we study how ANS and incentive motivation interact and how numerosity judgments are influenced by monetary motivations to make decisions. As expected, behavioral data analysis revealed that participants make accurate numerosity judgements and choose the right dot pattern over the left dot pattern as the number of dots in the right pattern outnumber the ones in the left dot pattern. But surprisingly, we did not observe a side bias, even though we expected participants to develop a bias towards the side that is more incentivized.

Nevertheless, fitting the combined choice and response time data through DDMs revealed the presence of significant calculation and response biases across participants. Participants had a starting point biased towards the right dot pattern when the right pattern was associated with high reward. This led to manifestation of response bias even though we did not observe a starting point biased towards the left dot pattern when the left pattern was associated with high reward. Presence of calculation bias was supported by the gaze data we collected. Irrespective of reward associations, participants always looked more at the right dot pattern than the left dot pattern, but more so when the right pattern was associated with higher reward. This differential gaze allocation behavior was related to calculation bias, but not to response bias, thereby

confirming the role of gaze allocation in calculation bias. aDDMs confirmed the gaze trends: right gazes were weighed more than left gazes in drift rate calculation.

We expected higher pupil dilation when choosing the pattern not associated with higher reward since that would require more mental effort but we did not observe any such trend. Response time, choice, or trial difficulty level did not predict pupil size. Moreover, pupillary reactivity bias was not related to response bias, which was surprising.

Our findings suggest that participants' numerosity judgements are not only influenced by the numerical ratio but also by the inherent biases in the evidence accumulation-based decision-making process. Incorporation of eye-tracking data provided a deeper understanding of how visual attention influenced decision-making. Relationship between gaze bias and calculation bias highlight interplay between visual processing and cognitive biases in numerical tasks. Effects seen due to differential reward associations suggest that participants adapt their numerosity judgments based on the reward context, and that too in different ways: some experience a perceptual change while others witness a shift in decision threshold. Confluence of numerosity judgements and decision-making has implications in educational settings. Insights from this study can inform development of interventions aimed at improving numerical cognition.

Future studies should aim at addressing the lack of pupil dilation effects, by perhaps introducing a starking contrast in reward allocations for the left and the right side. Moreover, further exploration into why the left pattern wasn't favored even with the presence of higher reward is warranted. Our findings open avenues to explore various other additional factors that can influence decision making in numerosity judgment tasks.

REFERENCES

Bendiksby, M. S., & Platt, M. L. (2006). Neural correlates of reward and attention in macaque area LIP. *Neuropsychologia*, *44*(12), 2411-2420.

DeWind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in human neuroscience*, *6*, 68.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247-265.

Dix, A., & Li, S. C. (2020). Incentive motivation improves numerosity discrimination: Insights from pupillometry combined with drift-diffusion modelling. *Scientific Reports*, *10*(1), 2608.

Dramkin, D., Bonn, C. D., Baer, C., & Odic, D. (2022). The malleable impact of non-numeric features in visual number perception. *Acta Psychologica*, *230*, 103737.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Medical Decision Making*, *27*(5), 672-680.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annu. Rev. Neurosci.*, *30*, 535-574.

Gottlieb, J., Hayhoe, M., Hikosaka, O., & Rangel, A. (2014). Attention, reward, and information seeking. *Journal of Neuroscience*, *34*(46), 15497-15504.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852-13857.
Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, *3*, 23998.
Kret, M. E., & Sjak-Shie, E. E. (2019). Preprocessing pupil size data: Guidelines and code. *Behavior research methods*, *51*, 1336-1342.
Lipkus, I. M., Samsa, G., & Rimer, B. K. (2001). General performance on a numeracy scale among highly educated samples. *Medical decision making*, *21*(1), 37-44.
Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255.
Park, J., & Starns, J. J. (2015). The approximate number system acuity redefined: A diffusion model approach. *Frontiers in psychology*, *6*, 173671.
Platt, M. L., & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741), 233-238.
Platt, M. L., & Plassmann, H. (2014). Multistage valuation signals and common neural currencies. *Neuroeconomics*, 237-258.
Platt, M. L. (2002). Neural correlates of decisions. *Current opinion in neurobiology*, *12*(2), 141-148.
Qin, N., Gu, R., Xue, J., Chen, C., & Zhang, M. (2021). Reward-driven attention alters perceived salience. *Journal of vision*, *21*(1), 7-7.
Ratcliff, R., & McKoon, G. (2018). Modeling numerosity representation with an integrated diffusion model. *Psychological review*, *125*(2), 183.
Ratcliff, R., Thompson, C. A., & McKoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition*, *137*, 115-136.
Sheng, F., Ramakrishnan, A., Seok, D., Zhao, W. J., Thelaus, S., Cen, P., & Platt, M. L. (2020). Decomposing loss aversion from gaze allocation and pupil dilation. *Proceedings of the National Academy of Sciences*, *117*(21), 11356-11363.
Sheng, F., Wang, R., Liang, Z., Wang, X., & Platt, M. L. (2023). The art of the deal: Deciphering the endowment effect from traders' eyes. *Science Advances*, *9*(34), eadf2115.
Wang, J. J., Libertus, M. E., & Feigenson, L. (2018). Hysteresis-induced changes in preverbal infants' approximate number precision. *Cognitive development*, *47*, 107-116.