ORIGINAL PAPER

A Tale of Two Literatures: A Fidelity-Based Integration Account of Central Tendency Bias and Serial Dependency

Ke Tong[1] · Chad Dubé[2]

Abstract

The ability to accurately estimate and reproduce the magnitude of stimulus features is critical for many daily tasks. However, experimental psychology has repeatedly confronted researchers with unexplained estimation biases stemming from preceding stimulus features. Serial dependency and central tendency bias are two particularly representative and ubiquitous examples. The core commonalities across these two response patterns raise a question: Are these seemingly different constructs re-describing a single phenomenon? The current paper tests this possibility by proposing a fidelity-based integration model (FIM) and testing it with three single item estimation experiments focusing on the visual features of line length and spatial frequency. The critical assumption of FIM is a fidelity-based sampling process that integrates information from both the target and recent non-target items. Our results suggest that central tendency bias and serial dependency reflect a single underlying process and that the distinct response patterns constituting these two phenomena are a by-product of the analytical methods traditionally favored within each domain. FIM also suggests that this implicit ensemble bias might be the basis for observed accuracy differences across implicit and explicit ensemble coding tasks.

The ability to accurately estimate and reproduce the magnitude of stimulus features is critical for many daily tasks. For example, when driving or parking, we need to continually estimate the distances between the cars and other objects and adjust these distances to desired levels. Failure to do so may lead to severe consequences. While healthy populations can generally do well on tasks like this, people with schizophrenia or autism have shown deficits in magnitude estimation in various feature domains (Notredame et al., 2014; Walter et al., 2009). This suggests that an understanding of feature magnitude estimation may provide critical insights into the functioning of normal and abnormal brains.

One of the most commonly used tasks to study our ability to estimate stimulus magnitude is the single item estimation task.[1] On any given trial, participants are asked to estimate the magnitude of a feature of a single stimulus (*target*)

by adjusting an adjustable stimulus (*probe*). This process is then repeated over multiple trials. Nominally, the main mechanism an experimenter hopes to recruit is some form of individual coding and retrieval of the stored item code. Intriguingly, our estimates of feature magnitudes seem to be systematically biased toward task-irrelevant preceding stimuli (e.g., Fischer & Whitney, 2014; Huang & Sekuler, 2010; Huttenlocher et al., 2000; Jazayeri & Shadlen, 2010; Petzschner et al., 2015; Rahnev & Denison, 2018).

The observed biases in single item estimation tasks are interpreted in different and sometimes overlapping contexts, such as *central tendency bias* (Hollingworth, 1910), *categorical effect* (Huttenlocher et al., 2000), *prototypical attraction* (Huang & Sekuler, 2010), *serial dependency* (Fischer & Whitney, 2014), *Bayesian updating* (Kalm & Norris, 2018; Petzschner et al., 2015), or *ensemble coding* (Crawford et al., 2019). Critically, many of the theoretical accounts that go along with the phenomena listed above contradict each other

✉ Ke Tong
ke.tong@ntu.edu.sg

1 Nanyang Technological University, 61 Nanyang Drive, ABN-01b-10, Singapore 637335, Singapore

2 University of South Florida, Tampa, FL, USA

---

[1] Also known as single item reproduction, magnitude estimation, direct match, delayed match, continuous/analog match. Also see other names summarized in Huang and Sekuler (2010). We use the name "single item estimation" to emphasize the critical task requirement that only one stimulus (*target*) is of direct relevance to the task on any given trial.
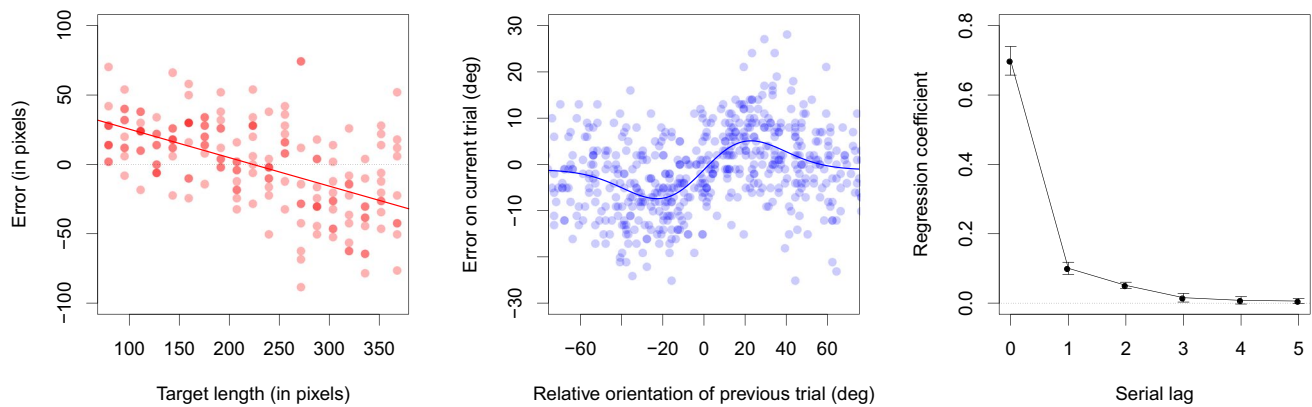
Fig. 1 The three commonly used analytical approaches for single item estimation tasks. Left panel: A linear regression line using stimulus values to predict estimation errors. Negative regression slopes suggest a central tendency bias, in that participants overestimate below-average stimuli and underestimate above-average stimuli (Huttenlocher et al., 2000). Middle panel: Estimation error plotted against the relative difference of stimulus value from the previous trial. Participants' responses on the current trials are biased toward the stimuli on the previous trials. For circular measures, the bias seems to peak and diminish within a certain range, showing a Derivative of Gaussian pattern between estimation error and the relative value of current and previous values (Fischer & Whitney, 2014). The peak amplitude of the curve is commonly used as a measure of the strength of the bias. Right panel: The estimate of the target regressed upon the stimulus values of the target and five preceding items. In addition to the target, the most recent few items show consistently non-zero regression coefficients, suggesting serial dependency on previous non-target items. The bias strength estimated from the analysis in the middle panel also shows decaying but significant non-zero values for the few recent items (Fischer & Whitney, 2014). The left and middle panels were reproduced using single-subject data from Duffy et al. (2010) and Fischer and Whitney (2014), respectively, with data kindly provided by the authors. The right panel was reproduced using Experiment 3 data from the current study, which is a replication of the memory condition of Experiment 1 in Huang and Sekuler (2010)

in the sources of such biasing influences. The central tendency, categorical, and prototype accounts attribute the bias to various forms of central tendency representations, while the serial dependency or Bayesian updating accounts interpret the bias as a result of integrating perceptual or memory traces from recent items. Not surprisingly, there exist dual-attractor models supporting both sources (e.g., Huang & Sekuler, 2010).

Since all the related studies used the same single item estimation paradigm, their data shared a similar structure. This allowed us to cross-analyze different datasets with a set of standardized analysis protocols. Interestingly, in the existing studies, different analysis protocols were used almost exclusively in their research contexts. For example, the analyses typically used to demonstrate central tendency effects (left panel in Fig. 1) are rarely used in the sequential dependency studies. Why not apply all the primary analytical tools to all the available data sets? Is it possible that the same dataset could reveal evidence for multiple theoretical accounts that are seemingly contradictory? Is it possible that these different theoretical accounts are describing particular ways of looking at the same data? If so, is a single unifying account likely?

The current paper aims to answer these questions. We first review the major analytical tools used to analyze data from the single item estimation task. Then, we evaluate existing models instantiating the different theoretical accounts. We show that the existing models share more similarities than

differences and that the core assumptions of most models are related to the change in exemplar fidelity over the trials and the implicit integration of exemplar information. Based on these ideas, we proposed a fidelity-based integration model (FIM) that provides a single unifying framework for understanding and explaining implicit inter-trial biases. A simulation compares the FIM and existing models' abilities to predict findings in the literature, and we provide additional tests of the models through three behavioral experiments.

Implicit Inter-trial Bias

To promote generalizability across literature, we use *implicit inter-trial bias* as the umbrella term for all the observed biases in the single item estimation tasks. We choose this term because every theoretical account seems to agree on the following three points: (1) such effects are not required by the task; thus they are *implicit*, (2) such effects stem from the multiple-trial format, hence *inter-trial* (further discussion is needed to determine if the bias is from prior stimuli, responses, or interactions), and (3) such effects draw responses away from the target; thus these effects are a type of *bias*.

Implicit inter-trial bias was found across a variety of stimulus feature domains, such as line length (Duffy et al., 2010), spatial frequency (Huang & Sekuler, 2010), orientation (Fischer & Whitney, 2014), spatial location

(Bliss et al., 2017; Manassi et al., 2018), numerosity (Fornaciai & Park, 2018), and facial expression (Liberman et al., 2018), suggesting it may reflect a domain-independent cognitive mechanism (Petzschner et al., 2015; Sanborn & Chater, 2016).

The empirical findings in these studies have been granted distinct terminologies and theoretical interpretations. Studies of *serial dependency* have tended to conclude that the observed implicit inter-trial bias is due to a handful of the most recent items, sometimes referred to as *response assimilation*.[2] Studies of *central tendency bias* have interpreted the corresponding inter-trial biases as indicative of a categorical influence from an arithmetic average, or *prototype*, computed over all the previously viewed stimuli. We wonder whether these two classes of effects might actually reflect a single, shared mechanism.

To understand why, there is an important issue to keep in mind: The two conclusions about the influence of prior items have each been reached via distinct analytical approaches that have each been characteristic within but not between the two research domains. The two methods in question are detailed in Fig. 1.

In analyses of *central tendency bias*, the estimation error (estimate—stimulus) is typically plotted against the stimulus values (Allred et al., 2016; Duffy et al., 2010; Huttenlocher et al., 2000). The expected data pattern for "central tendency bias" is described via a linear regression line through the data points with a negative slope. The negative slope indicates that observers generally overestimate stimuli whose feature values fall below the mean and underestimate stimuli whose values fall above the mean. This analysis is a general indicator of bias, but it does not address the source of the central tendency bias.

In analyses of *serial dependency*, item estimates are assumed to reflect a pull toward the most recent one or two items. The participants' estimates are thus predicted via short-range mechanisms tied to the specific items, such as trial-wise criterion shifting (M. Treisman & Williams, 1984) or representational leakage (Annis & Malmberg, 2013). The expected data pattern is that the current stimulus enjoys the majority of influence over responding, but the most recent few (one or two) items also have a non-zero influence, an influence which drops off rapidly with serial position. Studies exploring serial dependency often focus on analyses of subjects' estimates of a given stimulus feature as a function

of the individual features of previous trials, usually extending only a few trials into the past.

However, recent investigations found a pattern of recency-weighting in sequential averaging tasks (Hubert-Wallander & Boynton, 2015; Tong et al., 2019), suggesting that both perspectives are partly correct and partly incorrect. That is, participants appear to rely on a "prototype" or moving average; however, the computation is not an arithmetic average, but a weighted average in which the weights reflect the fidelity of items in memory (recency-weighting). If this hypothesis is correct, then a clear prediction arises: Both patterns of data, central tendency bias and serial dependency, should be demonstrable in the same data set depending only on the kind of analysis that is chosen.

To this end, in the current study, we use analytical methods from both sides of the conceptual divide to test the hypothesis that the corresponding predicted patterns could be revealed from the same data, depending only on the favored analytic method. Also, we wonder whether it is possible to relate implicit inter-trial bias to the more general literature of ensemble coding (Alvarez, 2011; Whitney & Yamanashi Leib, 2018). Although some types of ensemble codes are thought to be formed implicitly (Dubé & Sekuler, 2015), most ensemble coding studies employed explicit ensemble tasks, in which subjects were explicitly instructed to summarize multiple stimuli (e.g., mean estimation). What, if any, is the relationship between implicit and explicit ensemble coding?

Existing Models

Mechanistic models of the implicit inter-trial bias can be categorized into central tendency models and serial dependency models. Central tendency models share the assumption that implicit inter-trial bias is due to an implicitly formed category of all preceding items. The category center or "prototype," calculated as the moving average of the preceding stimulus values, pulls or attracts the memorized current stimulus toward the implicit category. The strength of the influence of this "category adjustment" process is determined by the relative variances of the current stimulus' memory representation and that of the memory representation of the category (Huttenlocher et al., 2000).

The serial dependency models generally assume that inter-trial bias is a result of activity or information from leftover memory traces of the most recent trials. The logic of the corresponding analytic strategy in the corresponding empirical work is that the contributions of these trial traces could be evaluated by their regression coefficients in multiple regression. The result from such analyses has shown that while the current stimulus takes the majority of the weight, the most recent one or two items also take

---

[2] The opposite pattern, *response contrast*, occurs over longer lags in studies of absolute identification but is typically smaller or absent in memory data, and is typically only observed in identification tasks when feedback is provided (Benjamin et al., 2009; Malmberg & Annis, 2012; Treisman & Williams, 1984). Hence, we focus on response assimilation in the current work, though we return to this issue in the General Discussion.

on non-zero weights in generating the response (Huang & Sekuler, 2010). From this point of view, the bias is likely to be linked to memory, but not perception (Bliss et al., 2017, but see Manassi et al., 2018), and the contributions of prior trial information are assumed to decay as they become older in memory (Kalm & Norris, 2018). In other words, the fall-off in weights implies that weights are chosen based on the resolution or presence of the corresponding items' memory representations (depending on which model of VSTM capacity one assumes; Ma et al., 2014; Zhang & Luck, 2008).

In what follows, we review the existing models of the single item estimation task. We started with a "null model" (Ideal observer model), followed by a central tendency model (Category adjustment model) and serial dependency models (Recency-weighted models, Bayesian updating models), and end with a dual-attraction model.

Ideal Observer Model

An ideal observer in the single item estimation task would base his or her responses solely on the current trials. As a result, no implicit inter-trial bias should be observed. In Eq. 1, R is response, S is stimulus, σ is the noise term, and the subscript i is trial number.

$$R_i \sim \text{Gaussian}(S_i, \sigma) \tag{1}$$

Although almost all data from the single item estimation studies contradict the assumptions of an unbiased, ideal observer model, this model itself has interesting implications. Consider that the "ideal" behavior can be caused by different underlying mechanisms. For example, one possibility is to have a perfect anti-interference device. Alternatively, a person with a minimal working memory capacity (e.g., one hard slot) would also behave like an ideal observer in this task. In the following sections, we show that major models are flexible enough to account for ideal observer-like behaviors.

Category Adjustment Model

In the category adjustment model (CAM, Huttenlocher et al., 2000), the implicit inter-trial bias is a "pull" from the category center. In Eq. 2, ρ is the category center, λ and $(1-\lambda)$ are the respective weights for the current stimulus and the category center. The weightings of the category center and the target stimulus (λ and $1-\lambda$ in Eq. 2) are modulated by the ratio of the standard deviations of the category representation and the target representation. The function $\lambda = g(\sigma_M/\sigma_\rho)$ was not specified in the original paper, but it was assumed to be "a monotonic decreasing function whose range is from 0 to 1" (Huttenlocher et al., 2000). The original paper used the term *memory inexactness* to refer to the representational

variability in the target or category center, qualifying the implicit inter-trial bias as a memory-based phenomenon.

$$R_i \sim \text{Gaussian}(\lambda \cdot S_i + (1 - \lambda) \cdot \rho, \sigma) \tag{2}$$

$$\lambda = g(\sigma_M/\sigma_\rho) \tag{3}$$

$$\rho = \frac{1}{i-1} \sum_{j=1}^{i-1} S_j \tag{4}$$

The original CAM paper stated that the category center is constructed by a Bayesian updating process (Huttenlocher et al., 2000), but apparently, CAM was also used in non-Bayesian ways (Crawford, 2019; Crawford et al., 2019; Duffy & Smith, 2018). In Eq. 4, the category center is an equally weighted running mean of all the preceding items. But regardless of the construction of the category center, CAM and related models assume the biasing influence is coming from a "category center," which summarizes all or a subset of the preceding trials.

CAM is widely used in a wide range of research areas (see a brief review in Duffy & Smith, 2018). However, the practical application of this model could be problematic. An earlier application to single item estimation suggested that the bias is from the category center, but not the recent preceding items (Duffy et al., 2010). Still, recently the authors showed disagreement on their early assessment (Crawford, 2019; Duffy & Smith, 2018). From a modeling perspective, none of their analysis in the debate applied the original CAM model. Instead, they used simplified versions of CAM without the critical function governing the relative weights and the Bayesian construction of the category center. The analysis in Crawford (2019), Duffy et al. (2010), Duffy and Smith (2018) used multiple regression with different sets of predictors. Their analysis essentially discarded the core mechanism of CAM, which is the critical $\lambda = g(\sigma_M/\sigma_\rho)$ function that implies a fidelity-based integration process.

Recency-Weighted Models

In the serial dependency account, the implicit inter-trial bias is from the residual information from the most recent items. The contributions of the recent trials are assumed to decay quickly as they become older. The response can be modeled as a weighted average of the current stimulus and a few previously shown stimuli. The key model specification is then how the weights decay over serial positions.

$$R_i \sim \text{Gaussian}(S_i \cdot w, \sigma) \tag{5}$$

$$S_i = S_{i-4} : S_i \tag{6}$$

$$w = r^{5:1} / \sum r^{5:1} \tag{7}$$

Equations 5 to 7 show one formalization using an exponential decay function, assuming the influence comes exclusively from the most recent five items. In Eq. 7, r is a decay rate over serial positions. r is greater than zero and not greater than one. A small r value means a steeper decay over serial positions, and the extreme case of a r value infinitely close to zero would lead to the ideal observer model. The idea of the exponential decay of weights of prior items was also implemented in a Bayesian updating framework referred to as the mixture model (Kalm & Norris, 2018).

The recency-weighted models were also used to model summary tasks like sequential averaging (Tong et al., 2019). Unlike single item estimation tasks, sequential averaging tasks explicitly ask participants to summarize, i.e., integrate information from prior items. Comparing modeling results from weighted-average models between single item estimation tasks and mean estimation tasks, we only find differences in terms of the degree of information utilization (through best-fit parameters, i.e., r in the exponential case in Eq. 7), but not mechanistic differences.

The weighted average model implies that the implicit integration and the explicit integration of information from prior trials are only different in terms of the degree of information utilization, but not mechanistically different. This hypothesis needs to be tested, since it has been long proposed that human observers apply qualitatively different processing modes or mechanisms across these different tasks (see a recent opinion on this issue: Baek & Chong, 2020).

Bayesian Updating Models

Sequential influences from prior states naturally fit in the Bayesian updating framework. The original CAM is assumed to be a Bayesian model, but the Bayesian updating occurs in the construction of the category center (Huttenlocher et al., 2000), making it closer to a central tendency bias view. The Bayesian updating models in this section are the recency-weighted models constructed in a Bayesian updating way, implying a serial dependency view.

$$P(\pi|s) \propto P(s|\pi) \cdot P(\pi) \tag{8}$$

$$\mu_{\pi|s} = w_s \cdot \mu_s + w_\pi \cdot \mu_\pi \tag{9}$$

$$w_s = 1 - w_\pi = \frac{1/\sigma_s^2}{1/\sigma_s^2 + 1/\sigma_\pi^2} \tag{10}$$

What is the proper prior on each new trial? Petzschner and Glasauer (2011) simply used the previous state. The information from the previous trial is integrated with information from the current trial, with weights determined by the relative variance in the representations, shown in Eqs. 8–10 (see a review in Petzschner et al., 2015). Kalm and Norris (2018) suggested that using only the previous state as a prior is not adequate, at least in the circular visual feature of orientation (Fischer & Whitney, 2014). Instead, a mixture of multiple recent states is suggested (Kalm & Norris, 2018). Critically, the function governing the mixing coefficients over recent preceding states is also an exponential decay function.

Setting the differences aside, these Bayesian updating models all incorporated the idea of fidelity-based integration, i.e., that the representation with higher fidelity or precision should contribute more in the item integration. However, these Bayesian models are still likely to be oversimplified since they do not specify how participants compute those explicit probabilities. Sanborn and Chater (2016) posited that humans might not calculate explicit probabilities but rather approximate them with posterior sampling. This view characterizes the brain as a rough, error-prone Bayesian sampler (Sanborn & Chater, 2016), rather than a perfect analytical Bayesian machine (e.g., Jazayeri & Shadlen, 2010; Petzschner et al., 2015).

Dual-Attractor Model

Huang and Sekuler (2010) used the analytical tools from both the central tendency view and the sequential dependency view. By separating the trials with congruent influences (in which the differences in feature values between the target to the mean and between the target and the previous item have the same direction) and incongruent influences (different directions), their analysis isolated the influences from the two sources. More critically, they found that the two sources of influence differed in their sensitivity to modulators like selective attention and inter-item similarity (Huang & Sekuler, 2010). Based on their results, they proposed a dual-attractor model showing that both a prototype and recent preceding items yielded biasing influences on a current trial's estimate, acknowledging both the central tendency view and the serial dependency view as explanations of the implicit inter-trial biases they observed.

However, the dual-attractor model in Huang and Sekuler (2010) is more of a conceptual model than a computational model. Thus, in its present form, it is of limited usefulness in producing simulations to make model predictions. Also, strictly speaking, their experiments using the post-cue conditions are not single item estimation tasks per se, because the post-cue condition ensured that more than one stimulus was task-relevant at the sensory processing stage. We develop this topic further in the General Discussion.

Fidelity-Based Integration Model (FIM)

From the model review, we found that different models shared many important similarities. The core assumptions of most models are related to the exemplar fidelity change over the trials and the implicit integration of exemplar information. The differences lie in the details of how information is integrated.

Here, we detail a fidelity-based integration model (FIM). FIM assumes that the observed implicit inter-trial bias is a result of an implicit information integration over the preceding items. The fidelity of item representations changes over trials, and the contribution of item information is dependent on the relative representational fidelity of the item. A fidelity-based sampling process is used to achieve the weighting on different sources factored into information integration. FIM is also a model framework that is conceptually compatible with many existing models.

FIM Assumptions

Memory representations consist of distributions of response potentials over a feature value axis. Representations from previous trials, in general, have much lower fidelity than the target, but nevertheless, some of the preceding items are still precise enough to influence a participant's response.

When doing the single item estimation task, participants generate their responses by sampling from the response representation. Ideally, this response representation should be solely the target representation; however, an implicitly formed ensemble representation, composed of samples from the representations of prior items, also has a role in this response representation. The relative contributions from the target and the ensemble are based on their representational fidelity: the number of samples drawn from a given representation is assumed to be a monotonically increasing function of the representation's fidelity. In this way, the probability density distribution of the resulting pooled distribution is inferred from fidelity-based samples from the target and the stimulus ensemble.

In the single item estimation task, the current trial stimulus, the target, has the highest representational fidelity, which is consistent with the task requirement. However, the prior trials still have some low fidelity representations, which could be integrated to generate the final response.

FIM's Account for Implicit Inter-trial Bias

The fidelity-based integration process results in responses that incorporate influences from noisy, recency-weighted average of prior trial response values. Because the weighted average will typically be closer to the arithmetic means than any single random value, the model will tend to produce a negative slope in the central tendency regression analysis, thereby manifesting a central tendency bias.

The same process is also responsible for the sequential dependency effect. In Fig. 2, note that the bumps in the probability density from the prior items are concentrated at the relatively high fidelity items ("1" and "2"). This means, in the FIM framework, with the assumptions of fidelity change
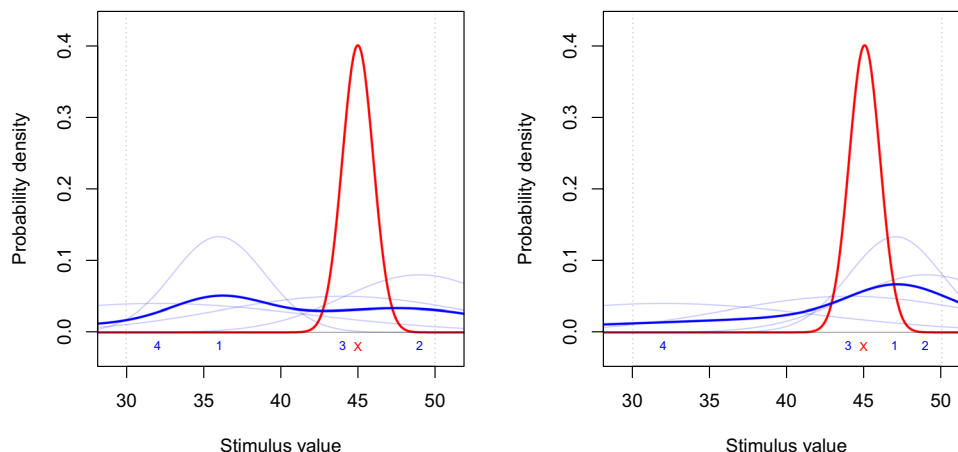

Fig. 2 The red curve is the probability density distribution of the target. The light blue curves in the backgrounds are probability density distributions of the preceding items. The blue curve is the Gaussian mixture of the preceding items. The implicit inter-trial bias is decomposed into the influence from recent preceding items ("1," "2," "3," and "4" on the axis. "1" means one item before target) on the response on the target (the red x on the x-axis). A given target value on the right side of the axis is more likely to have its 3–4 prior stimulus values falling to the left, which is the same direction toward the stimulus center (left panel) than the opposite direction (right panel), so estimates of the targets are overall more likely to be biased toward the center. Due to the fidelity change over the trials, the bumps in the probability density from the prior items are concentrated at the relatively high fidelity items ("1" and "2"). This means, in the FIM framework, with the assumptions of fidelity change and fidelity-based integration, central tendency bias, and serial dependency are the observed outcome of the same information integration process
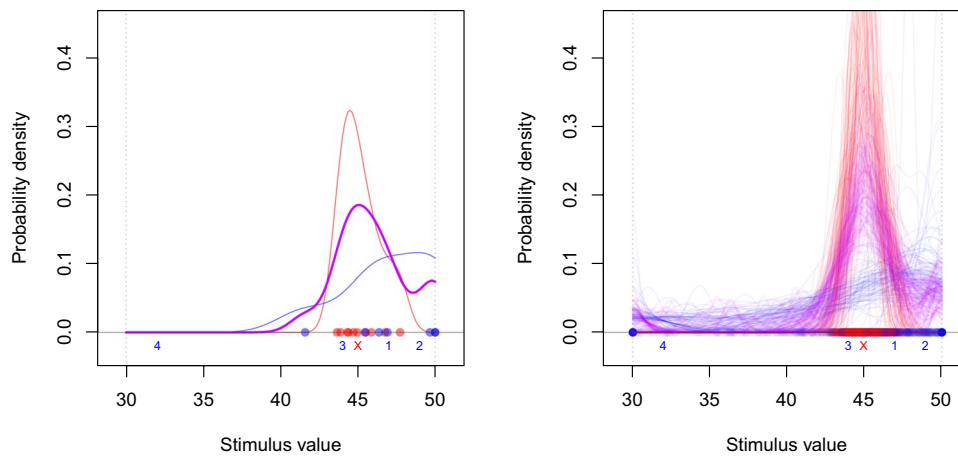
Fig. 3 FIM assumes that observers use finite samples and fidelity-based sampling without calculating the exact probabilities. They use density estimates based on the limited number of fidelity-based samples from the target and preceding items. The resulting response distributions for one trial (left panel) resemble the right panel of Fig. 2, but with random errors. The average of the long-run simula-tions (right panel) resembles the exact probabilities shown in the right panel of Fig. 2. For simplicity, Gaussian kernels with "nrd0" smoothing bandwidth were used in the density estimation in the simulations. The implementation and interpretation of these parameters in human sensory systems are open to investigation

and fidelity-based integration, central tendency bias, and sequential dependency are the observed outcome of the same information integration process.

Instead of calculating the exact probabilities shown in Fig. 2, FIM assumes that observers estimate representations from a finite number of samples, shown in Fig. 3. This sampler view from Sanborn and Chater (2016) makes FIM predictions to be error-prone on single trials, but given the long run, the average will approximate the exact calculation (Fig. 3). This assumption represents the conventional notion of a limited cognitive resource with a sampling account (Schneegans et al., 2020).

The assumption of fidelity change is deeply rooted in the memory literature. Here, the notion of a fidelity-based weighted sum in memory runs counter to the simple equally-weighted construction of the "prototype" or "category center" in early accounts of single item estimation (Duffy & Smith, 2018), as well as in influential memory models (Nosofsky & Zaki, 2002; Smith & Minda, 2000). Note that, if the "prototype" participants use is not an arithmetic aver-age, but a recency-weighted average or equivalent (as in FIM), this would challenge the conclusions of prior work that has rejected prototype models based on goodness-of-fit. We also addressed this point in explicit sequential mean estimation tasks (Tong et al., 2019), and now the same idea is applied to the single item estimation task.

Fidelity-based integration also adopts some of the ideas of the existing models we discussed in the model review. Intuitively, it makes sense to rely more heavily upon more reliable sources of information, and on this point, our approach philosophically aligns with "rational" models of cognition (Anderson, 1990). However, is this method of information integration really useful to us? In what follows, we generate some model predictions and test the predictions with empirical data. In the General Discussion, we expand further upon the usefulness to the organism of implicit bias and fidelity-based integration.

FIM Specification

A formal statement of the FIM model is straightforward. Equations 11–12 state that a vector of length j contains sam-ples, E_{ij}, that are drawn from V_i, the stimulus value (if it is the current trial) or response value (if it is one of the previous trials) that was on the ith trial. The number of samples, j, is a function of representational fidelity, σ_i. The fidelity parameter σ_i is a scalar function of time (t), serial position (s), and a host of other variables, depending on the particular task and stimuli in use. The specific forms of the functions $f(\cdot)$ and $g(\cdot)$ are to be determined. Equation 15 shows that samples drawn from all stimuli are combined into a sample pool, E_R, which is then used to make inference about the density distribution of the response.[3] Equation 16 shows that the response, R_i, is a sample from the density distribution of E_R.

$$E_i = <E_{i1}, E_{i2}, \dots, E_{ij}> \tag{11}$$

$$E_{ij} \sim Gaussian(V_i, \sigma_i) \tag{12}$$

---

[3] We use a set notation here as a tensor representation is difficult to implement without zero paddings when the vector lengths are vari-able across items.

$$\sigma_i = f(t, s, \dots) \tag{13}$$

$$j = g(\sigma_i) \tag{14}$$

$$\boldsymbol{E_R} = \{\boldsymbol{E_1}, \boldsymbol{E_2}, \dots, \boldsymbol{E_i}\} \tag{15}$$

$$R_i \sim \text{density}(\boldsymbol{E_R}) \tag{16}$$

Figure 4 demonstrates one realization of the FIM framework addressing the single item estimation task. As is clear in the figure, both serial dependency and central tendency bias arise from the same mechanism in FIM. More specifically, the

assumption of fidelity-based sampling and lag-based changes in fidelity result in a response that "depends on" (is sampled from) an ensemble representation that is neither an arithmetic average of prior stimuli nor a shift in responding based on one or two items. Instead, when fidelity drops with lag, the expected value of the ensemble representation in question is formally equivalent to a noisy, recency-weighted average of prior stimulus values.

FIM as a Model Framework

As we have detailed, the ideas instantiated in FIM are not limited to any single model. We view the key FIM assumptions as providing a general framework for models that share
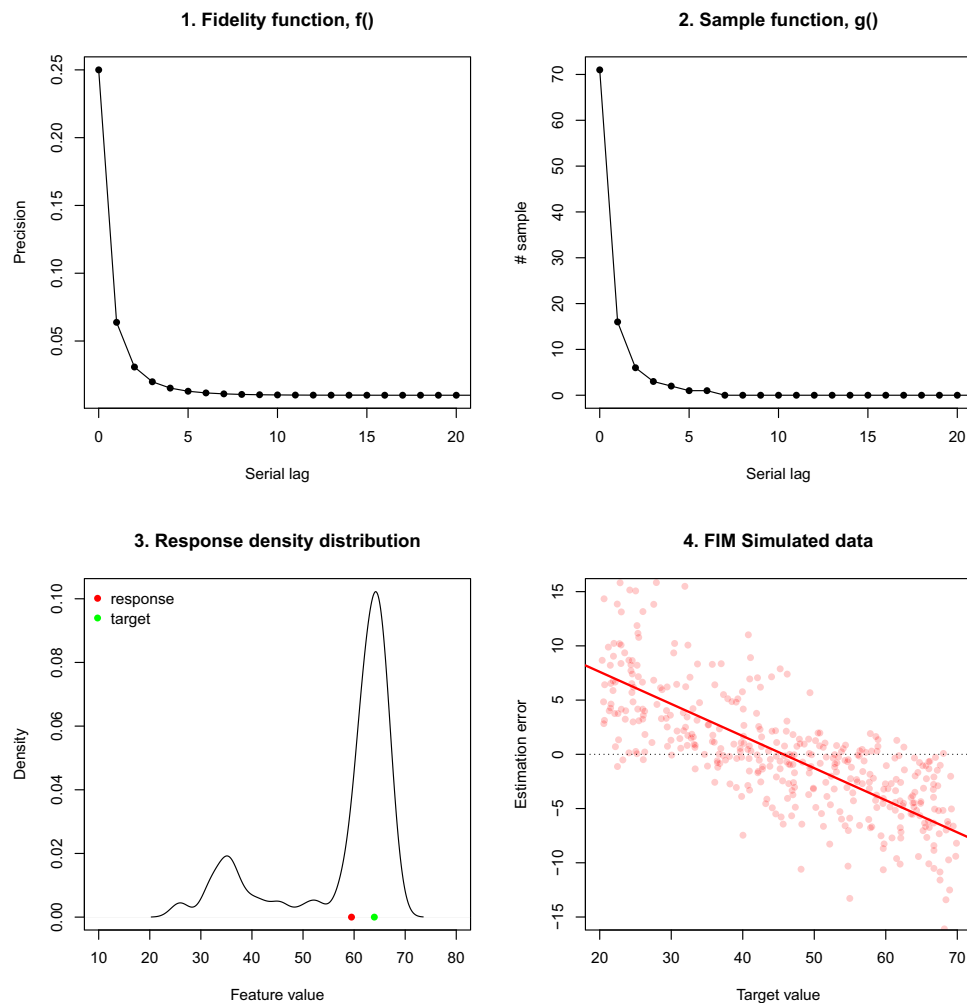

Fig. 4 A demo implementation of the FIM. In this demo, the fidelity function, $f()$, uses a logistic function to link representational SD and serial lag. To represent the concept of fidelity more intuitively, we transform SD into precision (inverse of variance) and show the fidelity (precision) curve in Panel 1. The sample function, $g()$, uses an identity function with linear scaling to assign samples according to representations' relative precision. A fixed number of 100 samples are used in this simulation. The number of samples are rounded to integers at each serial lag. Most serial positions will not be assigned

to any samples due to low fidelity, as shown in Panel 2. The response probability density distribution of one random trial is displayed in Panel 3, showing that samples from the current target stimulus and the preceding responses together influence participants' response selection. Panel 4 shows that the FIM-simulated data show the "central tendency bias," when FIM does not include any model component for an explicit representation of the global central tendency. Group-level simulated data from FIM and other models can be found in Fig. 7

these key assumptions but may differ in specific details or applications to specific tasks.

For instance, although the ideal observer model does not satisfy any FIM assumptions, the ideal-like estimation behavior is the FIM prediction when the target fidelity is much more precise than the preceding items. This prediction is intuitive: when the target representation is much better than those of leftover traces, why not just use the target representation? This reserves the possibility that in some cases (e.g., cases in which there are large changes in context or attention cues), implicit inter-trial bias should not be observed.

The Bayesian and recency-weighted models both show the idealized generative and observed sides of the FIM process. The explicit calculation of probability distributions in Petzschner and Glasauer (2011) and Kalm and Norris (2018) are simplified to finite samples in FIM (note that Kalm & Norris, 2018, also used finite samples in the actual calculation of their analysis). Figure 5 shows that such Bayesian sequential dependency models could recover the negative slope patterns (seemingly supporting the central tendency view) without any central tendency model component. These results confirmed our preference for sequential dependency models over central tendency models. The difference between FIM and sequential integration Bayesian models is the construction and integration of prior and likelihood. Bayesian models rely on full integration, assuming posterior to be Gaussian as well. The current implementation of FIM simply uses a Gaussian mixture of the density estimated from the finite number of samples. FIM shows that a sampling approach could simulate results that closely resemble the observed patterns with a finite number of samples (as low as 20 in total, as shown in Fig. 6). This shows that full mixture or high sample counts are unnecessary and even potentially redundant due to its potential high computation cost.

Compared to the Bayesian models and recency-weighted models, FIM provides more explanatory value by specifying the fidelity-based sampling process. FIM decomposes this process into two steps (Eqs. 13 and 14), making the fidelity of preceding representations the critical component in modulating the information integration process. The finite number of samples implements the limited processing capacity of our sensory systems. FIM also provided a wider generality, due to the flexible function construction. For instance, though the demo version of FIM (Fig. 4) uses the logistic function as $f()$, any other suitable functions can be tested in the FIM framework.

Model Predictions

FIM predicts many data patterns in experimental investigations of implicit inter-trial bias. We provide three testable predictions in this section. First, FIM models can simulate all the typical data patterns introduced earlier in Fig. 1, whereas non-FIM models cannot. The second prediction states that the relative fidelity difference between target and non-targets will modulate the strength of inter-trial bias. The third prediction states that in the analysis separating the two sources of bias (Huang & Sekuler, 2010), the influence from non-targets will generally be larger than the influence from the "prototype" in single item estimation.

Prediction 1

We simulated data from three models: FIM, non-FIM (the simplified CAM, which models responses as weighted averages of the targets and running means), and an ideal observer model as the null model. Then, we compared the predictions
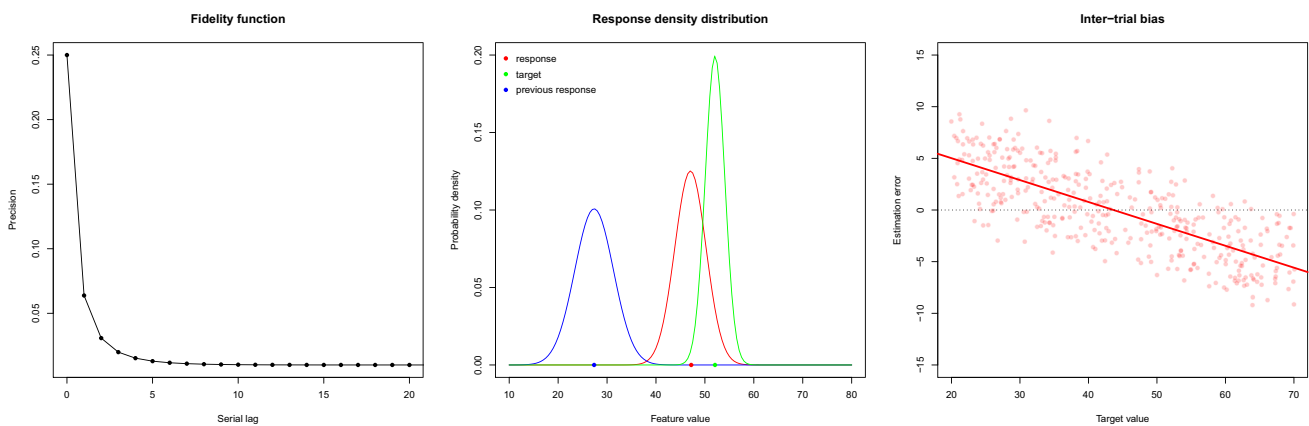

Fig. 5 Bayesian model using only one recent response as prior. Simulation show that such models could also recover the "central tendency" pattern (right panel) without any explicit central tendency model construction. This simulation uses the same fidelity function as FIM. Probability integration uses full integration shown in Eqs. 8–10
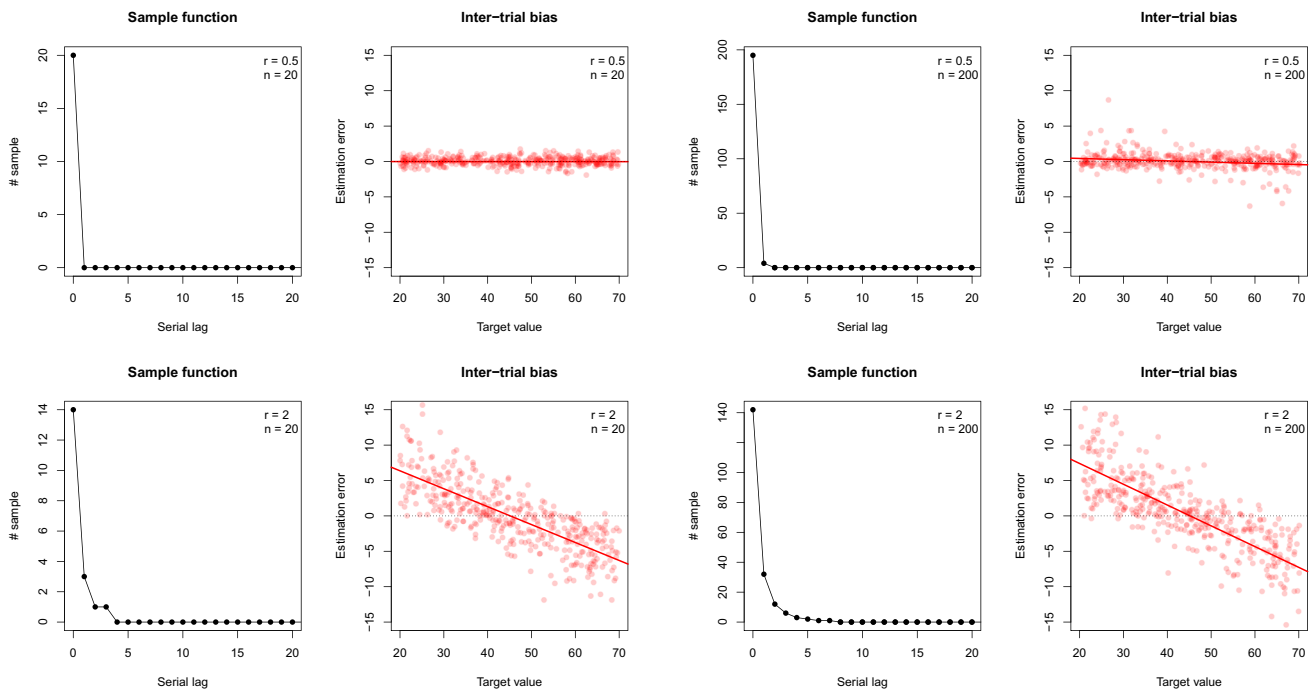
Fig. 6 In the FIM framework, the source of inter-trial bias is the relative precision of previous responses, governed by the fidelity function. The upper panels show that when the fidelity function has a scale parameter of 0.5, which translate to extreme relative precision of the target, no or very low inter-trial bias is simulated, when total sample count is 20 or 200. Lower panels show that then the scale parameter = 2, meaning that the recent few responses have relatively low but not extremely low fidelity, some samples will be assigned to them, thus causing the inter-trial bias. The total sample count does not influence the direction or the magnitude of the bias. The total sample count will only affect inter-trial bias when it is very low, so that no integer value of samples could be drawn from non-target items, as shown in the upper left panels

from these models in all four analyses introduced earlier. The model descriptions can be found in the *Existing models* section.

For each model, we simulated 20 virtual subjects with 400 trials per subject. The stimulus values were sampled from the same uniform distribution (upper = 50, lower = 30). For the ideal observer model, the population noise parameter, σ, was set to 3. Subject-level variation in the noise parameter was set to an SD of 0.3. For the simplified CAM, the population means of λ were set to 0.8. Noise term σ was set to the same level as the ideal observer model at 3. Subject-level λ values were sampled from Gaussian distributions centered at population means with SDs equal to 10% of the population means.

In FIM, the noise levels vary according to a logistic function with the location parameter of 1 and a scale parameter of 1 and linearly transformed to starting noise of 3 (for target) and a max noise of 10 (representing a guessing level), shown in the upper left panel of Fig. 4. The number of samples is the inverse of exemplar fidelity, rounded to the nearest integer, and linearly scales to have a total of 65 samples. The response is the mean of 10 samples from the response density distribution.

Simulated data from the three models (Top row: ideal observer; middle row: simplified CAM; bottom row: FIM) are shown in Fig. 7. As expected, data simulated from the Ideal observer model did not show any inter-trial bias in any analysis. Data simulated from CAM only recovered the first analysis. Although a slight effect of prior trials is evident in the data (blue lines), the regression-based analyses make clear that this stems from a small, uniform weighting over the prior items. This is an obvious reflection of the CAM's dependence on the arithmetic mean. FIM was able to recover all characteristic data patterns.

Prediction 2

Prediction 2 states that the relative fidelity difference between target and non-targets will modulate the strength of inter-trial bias. Figure 8 shows that when the target is relatively more precise (left panel), the information integrated from previous trials is lower compared with the condition when target has lower fidelity (right panel). This predicts that task configurations that result in lower target fidelity will increase inter-trial bias. This relative
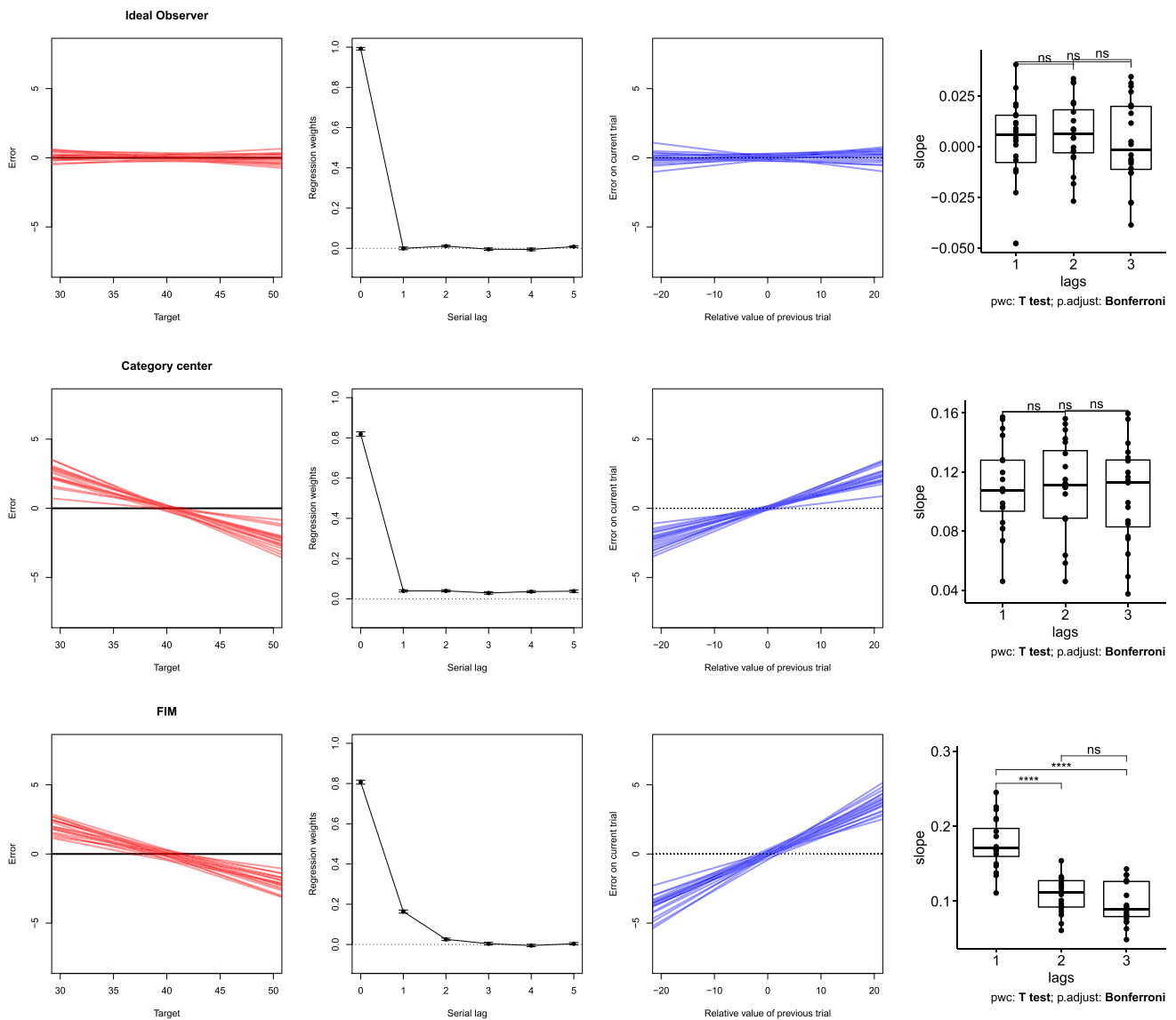
Fig. 7 Model predictions from the following models: ideal observer model (top row), category center model (middle row), and FIM (bottom row). As expected, data simulated with the Ideal observer model did not show any inter-trial bias in any analysis. Data simulated with the category center model showed central tendency bias in the first analysis and serial dependency in the third analysis, but no inter-trial bias in other analysis. FIM recovered all the typical data patterns

fidelity difference can be manipulated in many ways, for example, by varying selective attention (Dubé et al., 2014; Fritsche & Lange, 2019) or target-probe lags (Bliss et al., 2017; Corbin et al., 2017). One possible implication is that increasing the target-probe lag may increase implicit inter-trial bias. The FIM implementation is also easy: just add a time parameter in $f()$, as shown in Eq. 13.

On the opposite direction, this prediction also states that when the target fidelity is much higher than the preceding trials, ideal observer-like behavior is expected. Unlike conventional models like equal model or weighted average models that offer fixed weights for serial positions, FIM offers flexible predictions on information integration, depending on the relative fidelity.

Prediction 3

FIM predicts that the estimation error attributes mostly to recent non-target trials, rather than the central tendency representation. One particular analysis from Huang and Sekuler (2010) attempts to separate estimation error from prototypical sources and recent trial sources. The key assumption of this analysis is that the different sources of errors are additive and exclusively responsible for
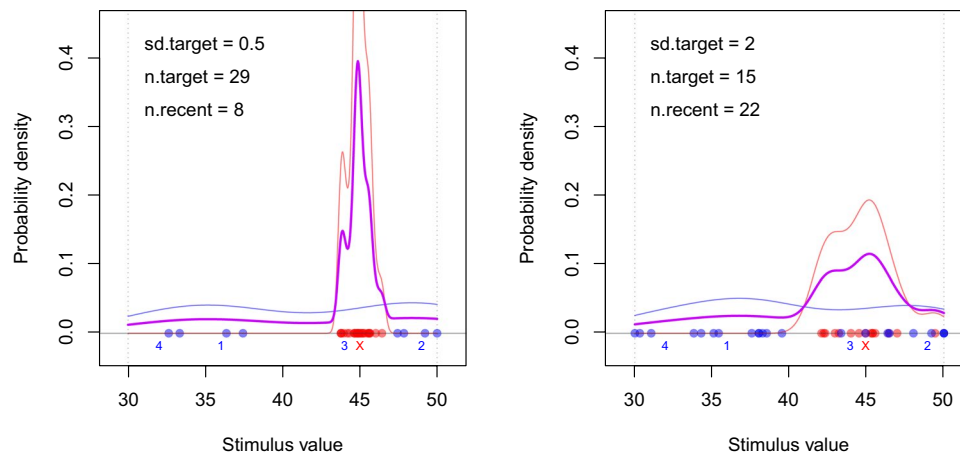
Fig. 8 FIM's assumptions of finite samples and fidelity-based sampling predict that the implicit bias is low when the target fidelity is relatively high compared to recent items (left panel), and vice versa (right panel). The relative fidelity between the target and the non-targets can be manipulated experimentally, e.g., by varying target-probe

lag time or the use of selective attention. Note that this figure is a one-time simulation. Slight variation may occur in different runs of the same simulation, but the basic pattern holds in simulations conducted over the long run

estimation errors. The recent stimulus and prototype are in the same direction to the target in the Same direction trials (Sdir), and in opposite directions in the Opposite direction trials (Odir). By adding the two sources and rearranging items in the equation, the two sources can be separated by the calculations shown in Eqs. 17 and 18 (see details in Huang & Sekuler, 2010). The calculation of the non-target part is limited to one trial before the target trial (lag-1 response). The calculation of the prototype is the moving (arithmetic) average of the stimuli preceding the target. Note that Experiments 2 and 3 in Huang and Sekuler (2010) used a cued reproduction task, and their "non-target" is the within-trial item that is not cued, which is different from the single item estimation paradigm in the current manuscript.

$$\text{Bias}_{\text{non-target}} = \left(\text{Error}_{\text{Sdir}} + \text{Error}_{\text{Odir}}\right)/2 \qquad (17)$$

$$\text{Bias}_{\text{prototype}} = \left(\text{Error}_{\text{Sdir}} - \text{Error}_{\text{Odir}}\right)/2 \qquad (18)$$

Figure 9 shows such analysis on FIM-simulated data. This reflects FIM's prediction that the majority of bias comes from inter-trial information integration, rather than from an explicit central tendency representation. Twenty "participants" were simulated using the same set of parameters, and the differences are sampling errors. The absolute bias values are in arbitrary scales. Note that this analysis is based on several strict assumptions, which might not be fully satisfied in actual data. This analysis assumes the total error to be exclusively from the two sources, and errors from the two sources are additive. However, we are interested to see whether we could observe similar patterns in actual single item estimation data.

Experiments

We tested the model predictions in three experiments. We investigated implicit inter-trial bias in single item estimation tasks with features of line length (Experiment 1 and 2) and spatial frequency (Experiment 3).
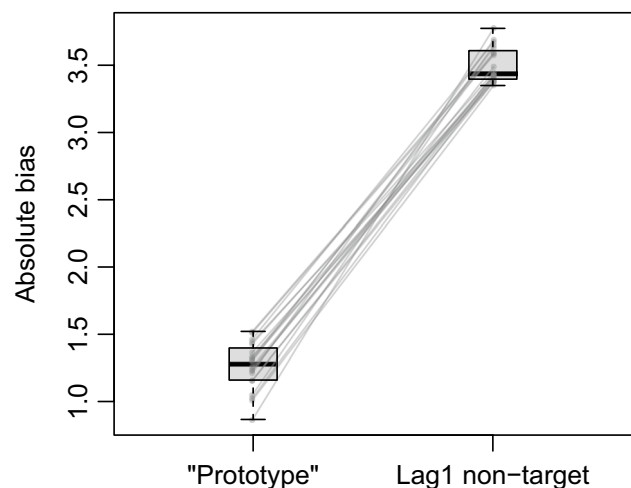

Fig. 9 A bias separating analysis on FIM-simulated data reveals that the estimation error in the FIM process is more from the non-target previous trials than the prototypical influences. In this analysis, "Prototype" is calculated as the moving average of all previous trials. "Lag-1 non-target" is the immediate preceding trial before the target trial. Prediction 3 suggests that similar patterns should be found in observed data

Experiment 1

Experiment 1 uses single item estimation task in the feature of line length. We found ideal-observer-like data patterns, which align with FIM's flexible prediction on target fidelity modulating implicit inter-trial bias. The results motivate us to further test the fidelity-modulating factor in Experiment 2.

Participants

Nineteen undergraduates participated in Experiment 1a. All procedures were approved by IRB at the University of South Florida (IRB Number: Pro00024055).

Stimuli and Procedure

On each trial, one vertical line (target) was displayed on the center of the screen. The target was presented for 0.5 s followed by a 1-s blank. After the blank, an adjustable line (probe) was displayed on the screen. The task was to reproduce the target by adjusting the probe using the mouse. When participants moved the mouse up or down, the probe increased or decreased in line length. When satisfied with their estimates, participants pressed the space bar on the keyboard to submit. No time limit was set for the adjustment phase. A trial ended with the participant's submission, followed by a 1-s blank screen before the next trial. Each participant completed 360 trials.

The target line length was randomly sampled from a uniform distribution between 200 and 400 pixels. The line width was set to 4 pixels. For ten participants, the starting values of the probe was set to 984 pixels, which is significantly longer than all possible target values. The remaining nine participants had a low probe starting value (4 pixels). Prior works showed that the initial values had no significant influence on participants' estimates (Corbin et al., 2017; Duffy et al., 2010; Huang & Sekuler, 2010).

Results

No implicit inter-trial bias of any sort was observed in Experiment 1 (all $p > 0.1$). Figure 10 showed analysis patterns referring to the analysis tools introduced in Fig. 1. The upper left panel showed the regression lines for all participants using target values to predict estimation error. The slopes are not significantly different from zero. The upper right panel showed that when using the stimulus values to predict the estimate of the target, the target took the majority of the total weight, leaving almost no influence on the preceding items. The bottom panels showed that the estimation errors were not biased toward the relative value of the previous trial, and their regression slopes, as a measure of the strength of the serial dependency, did not change over lags.

Results from Experiment 1 were contrary to prior reports (Allred et al., 2016; Duffy et al., 2010). We tested out alternative versions of the experiment in pilot data using a different probe placement and a normal distribution of target lengths. No implicit inter-trial bias was found in either of these two alternative conditions.

Experiment 2

Prior work suggests that central tendency bias might increase with longer target-probe lags (Corbin et al., 2017). Some work also suggests that serial dependency could increase with lag (Bliss et al., 2017). However, Huang and Sekuler (2010) manipulated the target-probe lag time and found no significant difference in implicit inter-trial bias.

These inconsistent results were from studies using different visual features, which precludes a strong conclusion regarding general mechanisms. For instance, it is possible that different features show different timescales or fundamentally different functions for the emergence of lag-dependent inter-trial bias. Indeed, given the wealth of literature demonstrating differential decay rates of different visual features in short and long-term memory (see, e.g., Carterette & Friedman, 1978), it seemed necessary to examine this question more closely with respect to the null results for line length in Experiment 1.

Therefore, in Experiment 2, we manipulated the target-probe lag time to test whether prolonged lag might result in inter-trial bias in the single item estimation task using line length.

Participants

Twenty-nine undergraduates participated in Experiment 2. All procedures were approved by IRB at the University of South Florida (IRB Number: Pro00024055).

Stimuli and Design

The target-probe lag was manipulated across five levels (0, 0.5, 1, 1.5, and 2 s) within subjects. Each participant completed 400 trials (80 trials per lag level). We randomly assigned the target-probe lag levels in the sequence of trials without any notice to the participants. All other configurations are identical to Experiment 1.

Results

Implicit inter-trial bias was found in the prolonged target-probe lag trials, as shown in Fig. 11. The regression slopes (predicting estimation errors with target values) showed a trend toward negative values as target-probe lag increased. The slopes were significantly different from zero in the 1.5-s trials,

Fig. 10 Observed data in Experiment 1, analyzed in four ways. The line length data did not show inter-trial bias in any analysis. The upper left panel shows error in responding as a function of the target feature value; negative slopes indicate central tendency bias, which is absent here. Each line is one participant. Other panels focus on serial dependency. Upper right shows the aggregate-level individual-trial contributions to responding, introduced in Fig. 1. The bottom panels show individual-participant error data against the relative stimulus value of the prior trial. Positive slopes indicate assimilation bias, which is absent in Experiment 1. The bottom right panel shows the regression slopes using lag-1, lag-2, and lag-3 stimulus at the non-target stimulus in the analysis. No serial dependency was found for any lag analyzed
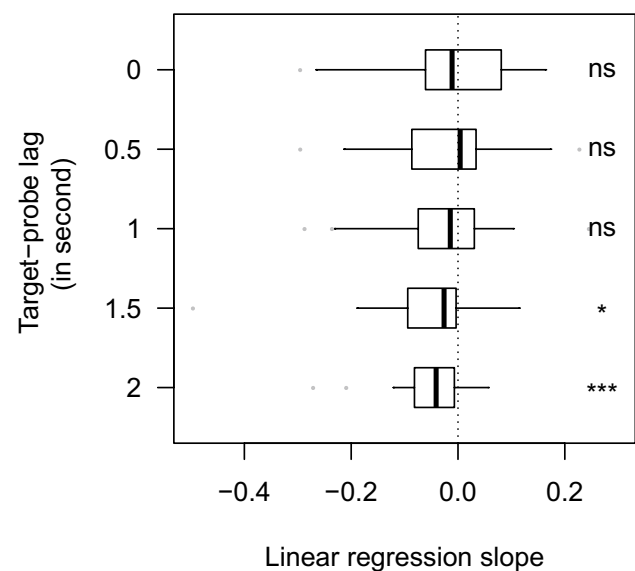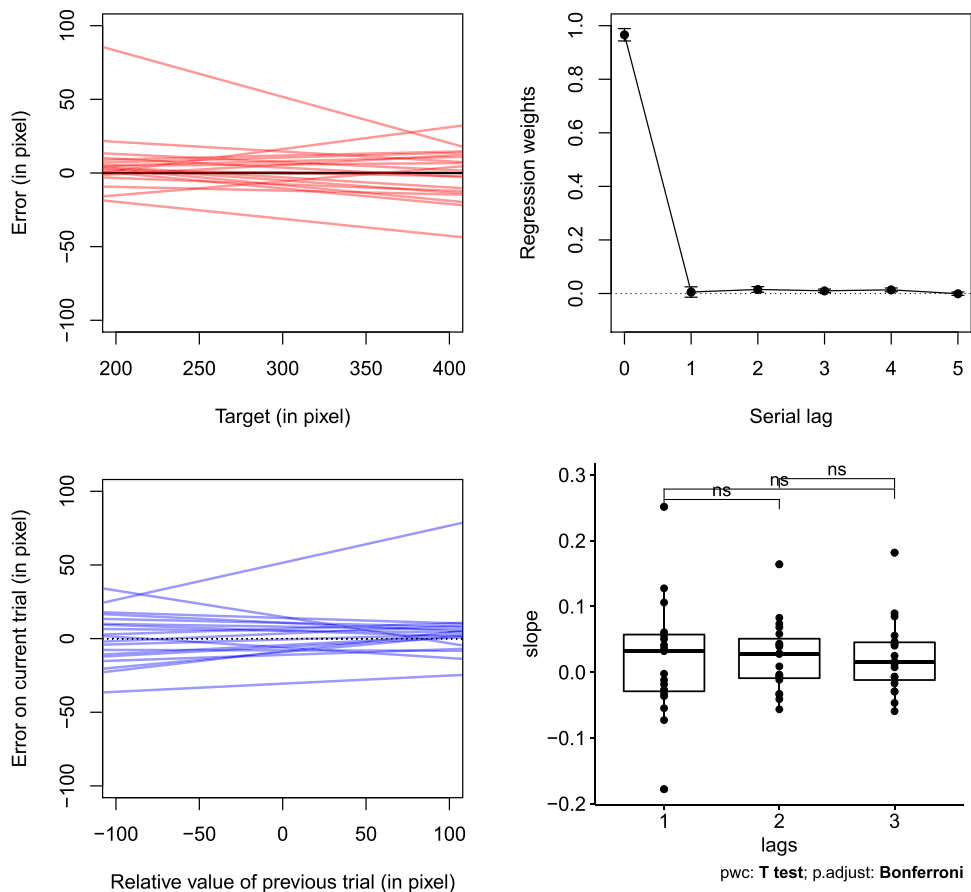

$t(28)=-2.41$, $p=0.02$, and the 2-s trials, $t(28)=-3.69$, $p<0.001$. Repeated-measure ANOVA was not significant at 0.05, $F(2.8, 78.48)=2.47$, $p=0.07$. A larger manipulation in the lag might show more significant results. The 1-s condition in Experiment 1b showed the same non-significant result as Experiment 1, which also used a 1-s target-probe lag. The trend suggests that longer target-probe lag might induce stronger implicit inter-trial bias, which is up to test in future studies.

What caused the target-probe lag's influence on the inter-trial bias? A prolonged lag may reduce the precision of the current trial's target representation, thereby leading to more reliance on representations from prior trials (Crawford et al., 2019), which is also predicted by the fidelity-based sampling assumption in FIM. Regarding the null effect of lag on implicit influence in Huang and Sekuler (2010), one possible explanation is that both their long (2400±100 ms) and short (1400±100 ms) lag conditions were long enough to elicit inter-trial bias. Also, it is conceivable that the difference between visual features (spatial frequency vs. line length) may also produce different functional forms relating lag and inter-trial bias. However, our current data do not allow us to map such functions out in sufficient detail to test this particular notion.

Fig. 11 Prolonged target-probe lag induced implicit inter-trial bias in Experiment 2. Negative regression slopes using target to predict estimation error indicate implicit inter-trial bias. The regression slopes were significantly different from zero in the longer target-probe delay conditions (1.5 s and 2 s) The 1-s condition replicates the results in Experiment 1. The dotted vertical line shows zero slope, which indicate no implicit inter-trial bias

Experiment 3

Experiment 3 aims to replicate a robust effect of implicit inter-trial bias and test the FIM's prediction 3 on source separating analysis. We choose the memory condition of Experiment 1 in Huang and Sekuler (2010), in which participants completed single item estimation task in reproducing the spatial frequency of Gabor patches. Their original analysis used the regression method shown in the right panel of Fig. 1. In Experiment 3, we aim to replicate their analysis, as well as all other analyses shown in Fig. 1. More interestingly, we are interested to see if the bias separating analysis pattern shown in FIM-simulated data (as shown in Fig. 9) can be found in empirical data.

Participants

Eleven individuals participated in Experiment 3. All procedures were approved by IRB at the University of South Florida (IRB Number: Pro00024055).

Stimuli and Procedure

On each trial, one Gabor patch (target) was displayed on the center of the screen. The target was presented for 0.5 s followed by a 1-s blank. After the blank, an adjustable Gabor (probe) was displayed on the screen. The task was to reproduce the spatial frequency of the target by adjusting the probe using the mouse. When participants moved the mouse up or down, the probe increased or decreased in spatial frequency. When satisfied with their estimates, participants pressed the space bar on the keyboard to submit. No time limit was set for the adjustment phase. A trial ended with the participant's submission, followed by a 1-s blank screen before the next trial. Each participant completed 168 trials. We generated the Gabor patches using Psychtoolbox-3 in Matlab. The target spatial frequency was randomly sampled from a uniform distribution between 0.5 and 5 cycle/deg. The adjustable range of the probe was 0.1 to 6 cycle/deg. Both the target and the probe Gabor patches were displayed in the center of the screen, subtending 5.15° in visual angle, on a black background. Many configurations from Huang and Sekuler (2010) were adopted (e.g., size of the Gabor patches, range of spatial frequency, number of trials per participant). Other configurations (e.g., display durations, ISI, and background color) were kept identical to Experiment 1 in the current manuscript.

Results

We conducted the same analysis as in Experiment 1 and recovered all major data patterns of the implicit inter-trial bias in Experiment 3, as shown in Fig. 12. For the central

tendency bias test, the slopes are significantly less than zero, $t(10) = -7.69$, $p < 0.0001$. The regression coefficients in the upper left panel are significantly greater than zero for the lag-0 to lag-2 (all $p < 0.001$) and not significant for lag-3 to lag-5. The slopes for the bottom panels are all significantly different from zero ($p < 0.001$), and a significant decay of bias strength was found as shown in the bottom right panel.

To separate the dual sources of ensemble bias, we modified the analysis method in Huang and Sekuler (2010) to suit our study design, introduced in the model prediction section (Prediction 3). Results in Fig. 13 confirmed the FIM prediction regarding the sources of ensemble bias. The magnitude of biasing influence is larger from the non-target than the "prototype", $t(10) = -3.28$, $p < 0.01$. As discussed earlier in Fig. 2, the underlying mechanism in FIM accounted for both central tendency bias and serial dependency. The differing patterns of "prototype effect" and "central tendency bias" were by-products of the analyses rather than distinct underlying mechanisms.
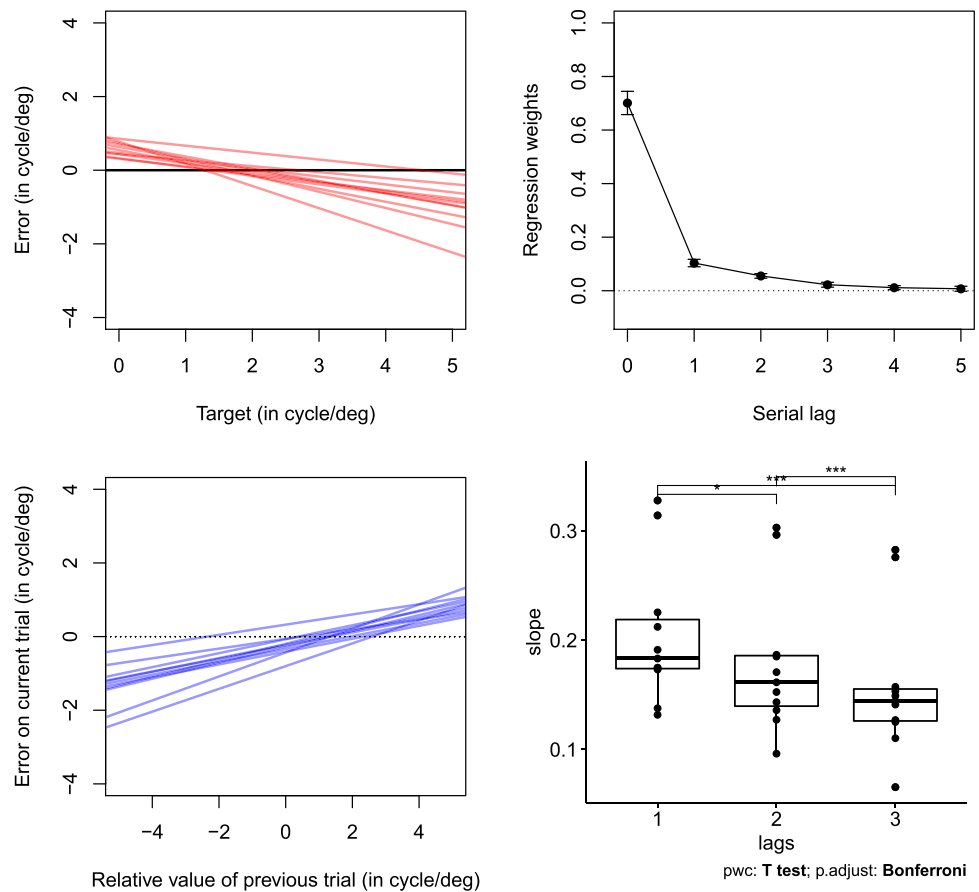
General Discussion

The current paper proposed a fidelity-based integration model (FIM) of implicit inter-trial bias. FIM simulations showed that the observed central tendency bias or serial dependency patterns might be different analytical descriptions of the same underlying information integration process. FIM's account is based on a few intuitive assumptions rooted in the perception and memory literature, such as a fidelity decay over trials, a finite number of cognitive resources (samples), and a fidelity-based sampling process that prioritizes sampling from high fidelity representations. FIM's model predictions were validated by the successful recovering of all the analysis patterns in a simulation and three empirical experiments using different types of visual stimulus features. In what follows, we discuss the potential of this work for understanding and relating shared response patterns in estimation and production tasks in visual perception, visual short-term memory, perceptual categorization, and long-term memory.

Inter-trial Bias in Perception, Categorization, Short-Term Memory, and Long-Term Memory

As we have noted elsewhere, implicit inter-trial biases are prevalent across tasks in many domains of cognitive psychology. This intriguing fact has naturally led many investigators to suspect an underlying commonality in the processes supporting those domains. This may seem natural today, but it is important to keep in mind the historical context within which many of the key tasks were developed. For

Fig. 12 Experiment 3 recovered all four analyses from Fig. 1. Experiment 3 is also a replication of the memory condition of Experiment 1 in Huang and Sekuler (2010)

Fig. 13 In Experiment 3, the magnitude of biasing influence is larger from the non-target than the "prototype," confirming FIM's account for ensemble bias. The two sources of bias were separated with the analysis in Eqs. 17 and 18. This analysis assumes that the two bias sources are additive. For details, see Huang and Sekuler (2010)

instance, in the 1950s a great deal of energy was directed toward the characterization of the perceptual system as an encoder/decoder or 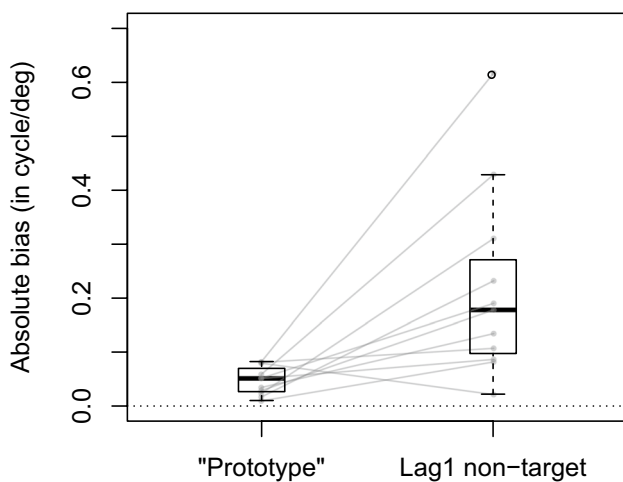processor in a digital communications system, following Shannon (1948) which happened to include a model of language production using successive approximations to English. Tasks such as absolute identification, in which the magnitude of a stimulus feature requires a previously acquired categorization response, and production tasks, which are the recall analogue of identification tasks, were used to quantify human processing capacity limits in bit units. However, in one highly influential paper by Miller (1956), a strict line was drawn between these "perceptual" tasks and their recognition and recall analogues in memory. Specifically, the author showed that while a bit limit seemed to hold for identification and production, no such limit was present in short-term memory tasks, which instead appeared to be limited by "chunks." Interestingly, this early work does not appear to question the premise that perception and short-term memory are distinct to begin with.

Despite the bit/chunk distinction, the sharp divide between perception and memory implicit in early work rapidly crumbled in the ensuing decades. For example, Ward and Lockhead (1970, 1971) uncovered inter-trial dependencies between responses and prior stimuli and responses in absolute identification tasks. Specifically, they found that responses are often pulled toward the prior response (or stimulus, under particular feedback

conditions). They referred to the "pull" as "assimilation." The authors also demonstrated a complementary pattern holding in relation to more distant trials, referred to as "contrast." These kinds of "sequential dependencies" were also observed in visual threshold estimation tasks (Howarth & Bulmer, 1956; Verplanck et al., 1952) and many perceptual signal detection tasks (Atkinson & Carterette, 1964; Carterette et al., 1966; Friedman et al., 1968; Parducci & Sandusky, 1965; Tanner et al., 1967, 1970; Treisman & Faulkner, 1984; Treisman & Williams, 1984; Zwislocki et al., 1958). These findings inspired renewed interest in relating perceptual and memory tasks via their patterns of dependencies, which has continued in more recent work demonstrating such short-term memory-based dependencies within long-term memory tasks such as judgment of frequency and old/new recognition tasks that have similar formats to visual and auditory signal-detection and absolute identification tasks (Annis & Malmberg, 2013; Malmberg & Annis, 2012).

Arguments have also been made in the opposite direction, from memory findings to predictions about perception. Specifically, powerful models of perceptual categorization tasks that closely resemble absolute identification have been successfully applied to short-term memory tasks as well (Nosofsky et al., 2011). More recent work has also suggested that predictions of the Generalized Context Model for short-term memory tasks could be improved if one allows for memory of statistical properties to inform judgments under conditions of low representational fidelity (Dubé, 2019).

At present, the most promising models of absolute identification frequently include memory processes and/or memory for statistical properties of stimuli, such as the mean, min, max, and range of feature values subjects must respond to (Duffy et al., 2010; Huttenlocher et al., 2000; Petrov & Anderson, 2005; Stewart et al., 2005). Clearly, the field has evolved a great deal since the time of Miller's paper, and today perceptual judgment tasks of the sort we have described are no longer viewed as passive or "memoryless." In this way, the implicit inter-trial biases of this literature have suggested that a common set of memory mechanisms underlies performance across all of these tasks. In FIM, we have attempted to formalize one such mechanism: the fidelity-driven integration of recent, low-dimensional memory traces on a common feature dimension. We envision this process as an operation upon the contents of short-term memory. For recent evidence bearing on this STM control hypothesis with respect to "ensemble coding" studies, see Zepp et al. (2021).

Next, we expand on the subject of sequential or "serial" dependency, by demonstrating how our results and modeling may help to answer longstanding questions about the inter-trial biases we see in multiple literatures.

Serial Dependency: Criterion Shifting or Representational Change?

In the serial dependency literature, prior approaches have often contrasted representational and decision-based models. For instance, Treisman and Williams (1984) accounted for response assimilation with a variable-criterion signal detection model. They assumed two processes, a short-term tracking process, and a longer stabilization process, that moves the response criterion to favor repetitions of the prior responses in the short term. The longer-term process prevents the criterion from wandering too far off into the tails of the evidence distributions, which would produce extreme and suboptimal response biases.

Representational models, on the other hand, generally assume that memory representations from prior stimuli influence the memory representations of more recent stimuli and/or the encoding of response probes. For instance, Annis and Malmberg (2013) proposed a REM-based model (Shiffrin & Steyvers, 1997) of sequential dependencies in memory and perceptual judgments that assumes information from feature vectors representing prior test probes carries over into the representations of current test probes. The authors interpreted this process as being dependent on the state of attention, with greater carryover occurring during states of low vigilance during the test and provided some empirical evidence in support of this assumption through manipulation of the dimensionality of the stimulus materials (words vs. pictures).

The FIM is closely aligned to the approach of Annis and Malmberg (2013), in that representations supporting responses to a probe are composites of current and prior test items. Like Annis and Malmberg (2013), we assume that most information will be sampled from the strongest representations. Although our model does not specify attentional probabilities, we note that decades of research has supported the idea that low attention produces low fidelity representations (see Ma et al., 2014, for a recent review from the standpoint of visual STM). In the Annis and Malmberg model, this means the current test probe produces less influence, and the prior items more impact on responding. In FIM, the same logic holds, though the emphasis is placed on the passage of time and intervening items without including an attentional mechanism. Future work should provide a more precise delineation of how FIM sits in relation to the memory and attention mechanisms detailed in the Atkinson-Shiffrin framework and successful mathematical instantiations of its subprocesses, such as REM and SAM (Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997).

On the other hand, we note our results and approach are inconsistent with criterion variability models such as that of Treisman and Williams (1984). Specifically, it is difficult

to see how the SDT model, which was designed to account for data from experiments using Likert or binary response procedures, can possibly account for data from the method of adjustment task. In short: How can a criterion shift produce a recall response that is distorted in the direction of a prior stimulus? The information carryover approach of representational models, while sharing the limitation of application only to absolute identification/Likert scale and binary response formats, can at least in principle support the kinds of tasks and data used here through their delineation of the feature space encoded and sampled from during retrieval in the SAM/REM framework.

A Useful Bias?

What, if any, are the benefits of ensemble bias to human cognition? The bias has been interpreted as a mechanism to maintain visual stability (Fischer & Whitney, 2014). The authors found that the magnitude of ensemble bias could be modulated by the similarity of the consecutive targets: the bias almost equals the relative difference within a certain range but starts to decrease when the difference is outside the range (middle panel in Fig. 1). In this way, visual stability is maintained by discarding small perturbations of the targets (bias within a small range), without losing the ability to detect real changes (less bias when the difference is outside the range).

This visual stability mechanism maximizes its benefit when the prior stimuli are informative in predicting the next stimulus in the sequence. This is often true when the stimulus sequence comprises representations of the same identity. For example, the sequence may be a stream of snapshots of the same moving object. However, in most single item estimation studies, the stimuli are randomly drawn from a uniform distribution, making the prior stimuli less informative about the next stimulus, despite some information about the general range.

This distinction relates to the choice of prior in Bayesian models: do we use information from all previous trials as prior or do we use information from a few previous trials as prior, for more general magnitude reproduction? This discussion might be not only related to data models (Petzschner & Glasauer, 2011), but also the item identity concept in the nature of the task. The item identity might be a modulating factor of the usefulness of this implicit influence. Instead of the same identity case, where perceptual stability is useful, there are cases in which the sequentially presented items are described as different identities. Why do we need to keep "stability" in such cases, where the stimuli are not supposed to be stable in the first place? Why do we bias our evaluation of a person's emotion toward the previous person?

Most existing ensemble coding studies do not clearly state the identities of the items. Could a prior description

modulate the sequential dependency and biasing effect? For example, the same set of stimuli can be described as either locations of the same target or the locations of different targets. Will this description difference lead to a measurable difference in the biasing effect? Further investigation is called for to understand the usefulness and detriment of this bias and the potential modulators for scenarios with different task goals, such as way-finding, tracking targets on radar images, or job performance rating in sequence.

Beyond Single Item Estimation

We initially proposed FIM as an integrated model for both implicit and explicit ensemble coding (Tong, 2020). In the current paper, the implicit part of the FIM is used to address single item estimation tasks.

In the full FIM framework, the single item estimation task is categorized as a *single-target isolate task*, in which participants are asked to process items in isolation, and every trial has only one target stimulus. The other two categories are *multi-target isolate tasks* (e.g., membership identification, Ariely, 2001) and *summary tasks* (e.g., mean estimation, Chong & Treisman, 2003). The three types of tasks are categorized by their task requirement, shown in Table 1.

One of the critical differences between summary and isolate tasks is the relative fidelity of the ensemble representation and the exemplar representations. Assuming participants deploy their attention among the items in the manner demanded by the task, one expects that in isolate tasks (like single item estimation), the ensemble representation is not as precise as exemplars. In contrast, in summary tasks, the ensemble representations should be more precise than exemplars.

Neither multi-target isolate nor single-target isolate tasks ask participants to summarize. However, the number of target stimuli differs in the two tasks. One example is the commonly used pre-cue/post-cue manipulation in perceptual or memory judgment tasks (Dubé & Sekuler, 2015; Dubé et al., 2014; Huang & Sekuler, 2010). When given the cue before the stimulus, participants know which item to attend to and can ignore the rest, thus rendering the task functionally equivalent to a single-target isolate task. But when given the cue after the stimulus presentation, participants need to process multiple items and keep them in the working memory store before the cue can guide them to further action. This makes a post-cue task functionally equivalent to a multi-target isolate task. The shared storage of the multiple items in multi-target isolate tasks may induce more information integration, as shown in membership identification tasks (Ariely, 2001; Haberman & Whitney, 2007). However, it

Table 1 FIM's categorization criteria for ensemble coding tasks

	Summary task	Multi-target isolate task	Single-target isolate task
Is summarizing multiple stimuli task-relevant?	Yes	No	No
Is processing multiple stimuli task-relevant?	Yes	Yes	No
Example tasks	*Mean estimation, mean comparison*	*Member identification, post-cue reproduction*	*Single item estimation, pre-cue reproduction*

is beneficial to isolate item information according to the task requirement.

How is this task-dependent characteristic achieved? FIM assumes that participants apply different statistical inferential goals in those two types of tasks. In isolate tasks, they make inferences about the distribution of the exemplars, whereas, in summary tasks, they make inferences about the distribution of the means of the exemplars. The central limit theorem tells us that the distribution of the exemplar means in the summary tasks will be more precise than the distribution of the exemplars.

This task-dependent distinction reconciles existing findings in implicit ensemble coding, found with isolate tasks such as membership identification and single item estimation. The FIM assumption of fidelity-based sampling predicts the recency pattern found in explicit ensemble tasks such as mean estimation. Critically, it points out that the implicitly formed ensemble representation is the basis of the more precise explicit summary ensemble representation. The implicit influence is also found in explicit summary tasks (Tong, 2020), suggesting a hierarchical structure of item representation and information integration, based on the fidelity-based sampling principles.

Open Questions

FIM's unifying account of central tendency bias and serial dependency implies that local similarity, rather than global similarity, may play a more important role in the biasing influences. Does this predict that for the same set of stimuli, different sequential order with different local similarities may show different levels of implicit ensemble bias?

What is the role of a potentially automatically perceived range (Ariely, 2001)? In Fig. 3, we corrected the samples outside the range to the nearest values within the range. If the range perception would relocate the probability density to areas within the stimulus range, even non-targets away from the center would be sampled more toward the center, which might be a secondary source of the observed ensemble bias.

The models introduced in the current paper do not distinguish the roles of prior stimulus values and prior responses. In the conventional single item estimation design, the high correlation of stimulus and response made it difficult if not impossible to separate the influences, but recent work showed ways to disentangle the influences from past stimuli and responses (Bosch et al., 2020). An alternative design, cued single item estimation, in which participants are asked to be ready to reproduce the target items but only do so when provided with a cue, may be useful to address this question.

Feedback has a notable effect on serial dependency and on mean estimation. In studies of sequential dependency using absolute identification, feedback appears to produce long-range response contrast as well as a greater influence of the prior stimulus (as opposed to the prior response) on short-term assimilation. In one study of mean estimation, feedback reduced recency weighting (Juni et al., 2010). Does this suggest that feedback in single item estimation would also reduce reliance on prior trials, even if the representational fidelity of the current target is low? How does feedback modulate attention's role in this process?

The current models do not provide any detailed, biologically plausible mechanism of encoding and retrieval of target and non-target information. FIM only assumes the representations to be samples, but how are those samples drawn from the sensory stream? How are they stored and retrieved? This raises the more general question of how the FIM mechanisms fit into the larger theoretical literature on human memory processing systems (Atkinson & Shiffrin, 1968). We aim to address this question, as well as the others we have posted here, in future work.

Code Availability All code is available upon request.

Declarations

Ethics Approval All procedures were approved by IRB at the University of South Florida. IRB Number: Pro00024055.

Consent to Participate All participants gave their permission to participate in the current study with written consent.

Consent for Publication The current study does not include any identifiable details about the participants.

Conflict of Interest The authors declare no competing interests.

References

Allred, S. R., Crawford, L. E., Duffy, S., & Smith, J. (2016). Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychonomic Bulletin & Review, 23*(6), 1825–1831.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*(3), 122–131.

Anderson, J. R. (1990). *The adaptive character of thought* (pp. xii–276). Lawrence Erlbaum Associates Inc.

Annis, J., & Malmberg, K. J. (2013). A model of positive sequential dependencies in judgments of frequency. *Journal of Mathematical Psychology, 57*(5), 225–236.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*(2), 157–162.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *The psychology of learning and motivation: II* (pp. xi, 249–xi, 249), Academic Press.

Atkinson, R. C., & Carterette, E. C. (1964). The effect of information feedback upon psychophysical judgments. *Psychonomic Science, 1*(4), 83–84.

Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics, 82*(1), 63–79.

Benjamin, A. S., Diaz, M., & Wee, S. (2009). Signal detection with criterion noise: Applications to recognition memory. *Psychological Review, 116*(1), 84–115.

Bliss, D. P., Sun, J. J., & D'Esposito, M. (2017). Serial dependence is absent at the time of perception but increases in visual working memory. *Scientific Reports, 7*(1), 14739.

Bosch, E., Fritsche, M., Ehinger, B. V., & de Lange, F. P. (2020). Opposite effects of choice history and evidence history resolve a paradox of sequential choice bias. *Journal of Vision, 20*(12), 9.

Carterette, E. C., Friedman, M. P., & Wyman, M. J. (1966). Feedback and psychophysical variables in signal detection. *The Journal of the Acoustical Society of America, 39*(6), 1051–1055.

Carterette, E. C., & Friedman, M. P. (1978). Perceptual ecology. New York: Academic Press

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*(4), 393–404.

Corbin, J. C., Crawford, L. E., & Vavra, D. T. (2017). Misremembering emotion: Inductive category effects for complex emotional stimuli. *Memory & Cognition, 45*(5), 691–698.

Crawford, L. E. (2019). Reply to Duffy and Smith's (2018) reexamination. *Psychonomic Bulletin & Review, 26*(2), 693–698.

Crawford, L. E., Corbin, J. C., & Landy, D. (2019). Prior experience informs ensemble encoding. *Psychonomic Bulletin & Review, 26*(3), 993–1000.

Dubé, C. (2019). Central tendency representation and exemplar matching in visual short-term memory. *Memory & Cognition, 47*(4), 589–602.

Dubé, C., & Sekuler, R. (2015). Obligatory and adaptive averaging in visual short-term memory. *Journal of Vision, 15*(4), 13–13.

Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research, 96*, 8–16.

Duffy, S., & Smith, J. (2018). Category effects on stimulus estimation: Shifting and skewed frequency distributions—A reexamination. *Psychonomic Bulletin & Review, 25*(5), 1740–1750.

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review, 17*(2), 224–230.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience, 17*(5), 738.

Fornaciai, M., & Park, J. (2018). Attractive Serial Dependence in the Absence of an Explicit Task. *Psychological Science, 29*(3), 437–446.

Friedman, M. P., Carterette, E. C., Nakatani, L., & Ahumada, A. (1968). Comparisons of some learning models for response bias in signal detection. *Perception & Psychophysics, 3*(1), 5–11.

Fritsche, M., & Lange, F. P. (2019). The role of feature-based attention in visual serial dependence. *Journal of Vision, 19*(13), 21–21.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology, 17*(17), R751–R753.

Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods, 7*(17), 461–469.

Howarth, C. I., & Bulmer, M. G. (1956). Non-random sequences in visual threshold experiments. *Quarterly Journal of Experimental Psychology, 8*(4), 163–171.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: Two classes of attractors at work. *Journal of Vision, 10*(2), 1–27.

Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of Vision, 15*(4), 5–5.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129*(2), 220–241.

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience, 13*(8), 1020–1026.

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2010). Integration of visual information across time. *Journal of Vision, 10*(7), 1213–1213.

Kalm, K., & Norris, D. (2018). Visual recency bias is explained by a mixture model of internal representations. *Journal of Vision, 18*(7), 1–1.

Liberman, A., Manassi, M., & Whitney, D. (2018). Serial dependence promotes the stability of perceived emotional expression depending on face similarity. *Attention, Perception, & Psychophysics, 80*(6), 1461–1473.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*(3), 347–356.

Malmberg, K. J., & Annis, J. (2012). On the relationship between memory and perception: Sequential dependencies in recognition memory testing. *Journal of Experimental Psychology: General, 141*(2), 233–259.

Manassi, M., Liberman, A., Kosovicheva, A., Zhang, K., & Whitney, D. (2018). Serial dependence in position occurs at the time of perception. *Psychonomic Bulletin & Review, 25*(6), 2245–2253.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81–97.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 28*(5), 924–940.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological Review, 118*(2), 280–315.

Notredame, C.-E., Pins, D., Deneve, S., & Jardri, R. (2014). What visual illusions teach us about schizophrenia. *Frontiers in Integrative Neuroscience, 8*, 63.

Parducci, A., & Sandusky, A. (1965). Distribution and sequence effects in judgment. *Journal of Experimental Psychology, 69*(5), 450–459.

Petrov, A. A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review, 112*(2), 383–416.

Petzschner, F. H., & Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *Journal of Neuroscience, 31*(47), 17220–17229.

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences, 19*(5), 285–293.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*(2), 93–134.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, 41.

Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences, 20*(12), 883–893.

Schneegans, S., Taylor, R., & Bays, P. M. (2020). Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences*, 117(34), 20959–20968.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145–166.

Smith, D. J., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 3–27.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review, 112*(4), 881–911.

Tanner, T. A., Haller, R. W., & Atkinson, R. C. (1967). Signal recognition as influenced by presentation schedules. *Perception & Psychophysics, 2*(8), 349–358.

Tanner, T. A., Jr., Rauk, J. A., & Atkinson, R. C. (1970). Signal recognition as influenced by information feedback. *Journal of Mathematical Psychology, 7*(2), 259–274.

Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? Averaging visual features in a sequence. *Attention, Perception, & Psychophysics, 81*(6), 1962–1978.

Tong, K. (2020). A Fidelity-based Integration Model for Explicit and Implicit Ensemble Coding (Doctoral dissertation), University of South Florida, Tampa, United States.

Treisman, M., & Faulkner, A. (1984). The effect of signal probability on the slope of the receiver operating characteristic given by the rating procedure. *British Journal of Mathematical and Statistical Psychology, 37*(2), 199–215.

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review, 91*(1), 68–111.

Verplanck, W. S., Collier, G. H., & Cotton, J. W. (1952). Nonindependence of successive responses in measurements of the visual threshold. *Journal of Experimental Psychology, 44*(4), 273–282.

Walter, E., Dassonville, P., & Bochsler, T. M. (2009). A specific autistic trait that modulates visuospatial illusion susceptibility. *Journal of Autism and Developmental Disorders, 39*(2), 339–349.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology, 84*(1), 27–34.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics, 9*(1), 73–78.

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129.

Zepp, J., Dubé, C., & Melcher, D. (2021). A direct comparison of central tendency recall and temporal integration in the successive field iconic memory task. *Attention, Perception & Psychophysics, 83*(3), 1337–1356.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–235.

Zwislocki, J., Maire, F., Feldman, A. S., & Rubin, H. (1958). On the effect of practice and motivation on the threshold of audibility. *The Journal of the Acoustical Society of America, 30*(4), 254–262.