Memory of relative magnitude judgments informs absolute identification

Nisheeth Srivastava (nsrivast@cse.iitk.ac.in)
Department of CSE, IIT Kanpur
Kalyanpur, UP 208016 India

Abstract

The question of whether people store absolute magnitude information or relative local comparisons of magnitudes has remained unanswered despite persistent efforts over the last three decades to resolve it. Absolute identification is one of the most rigorous experimental benchmarks for evaluating theories of magnitude representation. We characterize difficulties with both absolute and relative accounts of magnitude representation and propose an alternative account that potentially resolves these difficulties. We postulate that people store neither long-term internal referents for stimuli, not binary comparisons of size between successive stimuli. Rather, they obtain probabilistic judgments of size differences between successive stimuli and encode these for future use, within the course of identification trials. We set up a Bayesian ideal observer model for the identification task using this representation of magnitude and propose a memory-sampling based approximation for solving it. Simulations suggest that the model adequately captures human behavior patterns in absolute identification.

Keywords: absolute identification; relative judgment; mental representations; memory; Bayesian learning

A theoretical debate

One of the most compelling mysteries of human behavior is the nature of the information we store that enables us to decide what to do rapidly and effectively in everyday life. Theoretical opinion currently lies on a spectrum conceptually defined by two strongly divergent positions: one camp, generally better represented among economists and neuroscientists, assumes that people have direct psychophysical access to the magnitude of entities in the world (how big was this stimulus on the scale I'm interested in?) (Brown, Marley, Donkin, & Heathcote, 2008); the other, mostly containing psychologists, claims that people store only the relative results of comparative evaluations (which stimulus was bigger?) (Stewart, Brown, & Chater, 2005).

A classic problem for the absolute magnitude camp is that of absolute identification. Across a range of sensory modalities line lengths, sound frequency, sound loudness, etc. observers are quicker and more accurate in identifying stimuli at the extremes of the presented stimuli set than for stimuli in the middle (Lacouture & Marley, 2004). In addition to identification, it is also possible to ask participants to categorize perceptual stimuli into one of two groups, in which case a similar pattern of results is seen to hold. The same 'bow-tie' seen in identification experiments is seen in perceptual categorization experiments, with extreme stimuli within the stimuli set categorized more accurately and rapidly (Lacouture & Marley, 2004; Ratcliff & Rouder, 1998). Why should the range of stimuli presented in a set affect observers' responses to individual stimuli, if each stimulus has its own independent internal magnitude representation?

A classic challenge posed to the relative magnitude camp is the distance effect seen in closely related experiments. Distance effects are seen when people are asked to identify which of two presented stimuli is larger (Ratcliff & Rouder, 1998). Participants in Ratcliff & Rouder's brightness discrimination experiment were more accurate and quicker to respond when pairs of presented stimuli were far apart in actual brightness. The distance effect in perceptual choice finds an exact counterpart in the distance effect observed in economic experiments, where participants are seen to be more inconsistent and late in responding when the value of competing options is close (Dickhaut, Smith, Xin, & Rustichini, 2013). If people aren't storing absolute magnitude information, why do they find stimuli farther apart easier to categorize and differentiate?

Explanations for subsets of these phenomena have been previously proposed, and are briefly described below. However, the universality of these effects in perception and cognition demands a universal explanation, one equally applicable to simple perceptual identification tasks and to cognitive preference judgment tasks. Sequential sampling models, best represented by Roger Ratcliff's seminal drift diffusion model (Ratcliff & Rouder, 1998), are in fact universally used to fit data from all of these tasks, and commonly provide very good empirical fits. Drift diffusion models (DDM) assume that evidence favoring multiple alternatives rises in separate evidence accumulators, and the first accumulator to reach an implicit decision threshold is emitted as the overt choice. Even the simplest DDM style accounts of the choice process, though, use $O(N)$ parameters to fit choice and response time data in N-alternative experiments (Brown & Heathcote, 2008). While empirically valuable, such modelling is descriptive; it does not provide insight into why the error and RT distributions with respect to stimulus order take the typical 'bowtie' shapes they do (Brown et al., 2008). All DDMs can say in such cases is that evidence for extreme stimuli accumulates faster than for intermediate stimuli. They cannot explain why this happens.

Challenges to current accounts

Sophisticated models of absolute identification place the source of this pattern in the process by which observers map their internal representations of perceived stimuli magnitude onto discrete number labels. For instance, in (Lacouture & Marley, 1995), Lacouture & Marley showed that treating the magnitude-label mapping problem as an *encoder problem*, to be solved by a feed-forward network, yields mappings for response strengths quadratic in the stimulus order, immediately yielding the bowtie effect when coupled with a DDM (Brown

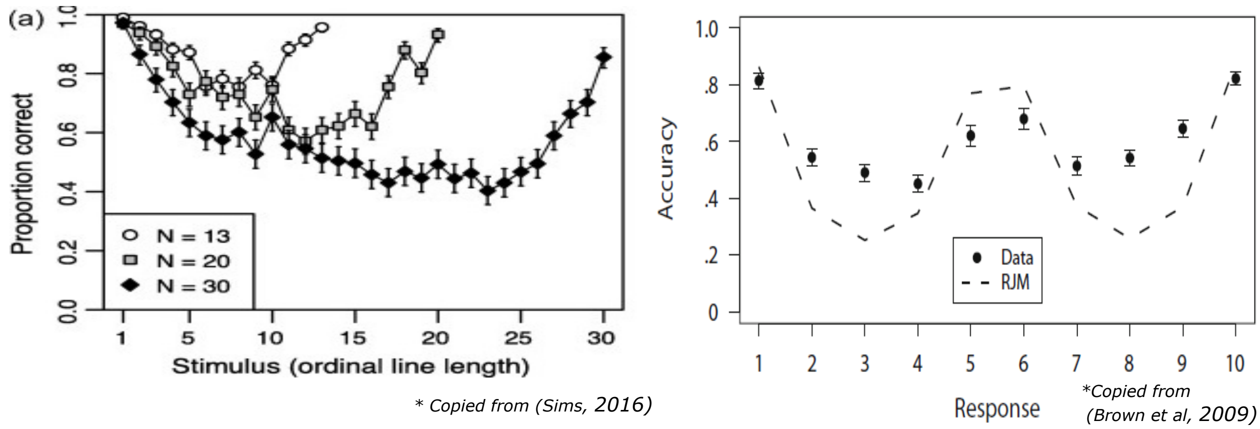

Figure 1: Problems for current accounts of absolute identification. *(Left)* Accounts that assume that observers maintain long-term distributions of stimuli magnitude find it hard to explain results where the same stimulus is responded to differently by the same subject depending on its order within different stimuli sets. *(Right)* Accounts that assume that observers maintain only a memory of 'which was bigger' judgments find it hard to explain effects of unequal spacing in magnitudes within stimuli sets.

et al., 2008). The psychophysical rationale for such a mapping more precise for extreme stimuli is illustrated in Figure 1. Assuming that long-term absolute internal representations of stimuli magnitude are noisy, the efficient encoding hypothesis holds that when confronted with a specific stimuli set, humans will respond to the specific task challenge of mapping stimuli to labels by comparing the presented stimulus to all available internal referents. The strength of the evidence for the mapping is information-theoretically stronger for stimuli corresponding to fewer overlapping internal referents, thus privileging points closer to the extremes, since they will have less interference from stimuli representations from one side of the scale.

A prominent empirical challenge to such accounts comes from the finding that stimuli of the same length are responded to differently when they are members of stimuli sets of different lengths, even within the same subject (Sims, 2016). If long-term stimulus magnitude representations exist, then they should be indifferent to the impact of adding more stimuli to an existing stimuli set, and the pattern of response should not change for the side of the stimulus order where new stimuli are not added. However, empirical evidence (see Figure 1 left) shows that it does. The solution to this problem presented in (Sims, 2016) is to adjust the noise levels in the internal stimuli representations 'adaptively' as a function of the set of stimuli to be represented. Such solutions, while mathematically feasible, call into question exactly how long-term the internal representations are, if they are to be so responsive to extraneous context.

Adopting a representation of stimuli that stores only local comparisons, Stewart et al have argued that observers, once given feedback about the previous trial, and comparing the current stimulus to the immediately previous one, can restrict the range of possible responses by using the previous stimulus as an upper or lower bound for the new one (Stewart et al., 2005). This range restriction naturally proves to be more informative for stimuli closer to the edge of the stimulus set range, making responses to these stimuli more accurate. Thus, a convex relationship between response strength and stimulus order, specific to the presented stimuli set, is obtained.

Prominent challenges to such relative comparison-based accounts include the fact that they do not provide easy explanations for differences in response patterns induced as a function of unequal distances between stimuli in absolute identification tasks. Including a large gap in the middle of an otherwise linear in log space stimuli range, Brown et al show that people find stimuli surrounding this gap easier to identify, whereas relative judgment models find it hard to even fit such data without detracting from predictive performance for the other stimuli (Brown, Marley, Dodds, & Heathcote, 2009). The core problem is that the model in question, Stewart et al.'s relative judgment model (RJM) uses a hard threshold in inter-stimulus distance to determine if a stimulus is larger or smaller than its predecessor, and fits this threshold as a parameter. Changes in spacing end up compromising the quality of the model fit.

It is intriguing to note that what is hard to explain using one family of models is easy using the other. Relative judgment models will have no difficulty explaining the effect of multiple stimuli sets on the response pattern, since there are no long-term response strength mappings to expect consistent responses from. Absolute magnitude models will find it straightforward to explain heightened accuracy across large gaps - assuming the same variance for each internal representation, shifts in the mean by adding a gap increases the discriminability of neighboring stimuli, increasing the response strength for the corresponding stimuli.

Finally, both classes of models find it hard to explain practice effects in absolute identification - the fact that participants in these experiments actually get better at the task given practice (Rouder, Morey, Cowan, & Pealtz, 2004). Since neither class of model posits any form of learning mechanism for observers, they fail to explain the actual learning curves seen in real experimental subjects (Dodds, Donkin, Brown, & Heathcote, 2011).

Judgments are formed from memories

The striking complementarity of the strengths and weaknesses of absolute and relative models of absolute identification suggest an opportunity to formulate an intermediate account that bridges this theoretical divide. We make an effort to do so in this paper.

In contrast with previous work, we make three novel assumptions about the process by which observers perform absolute identification and related tasks. First, we assume that the mental representation actually used by people in such tasks is a judgment of *relative* magnitude made using comparison to the immediately preceding stimulus during the experiment. Second, we assume that observers learn the stimulus-label mapping via a process well-described as an approximately Bayesian learning algorithm that explicitly samples memory engrams corresponding to the internal representations of stimulus magnitude learned during earlier trials of the experiment. Finally, we assume that this memory sampling self-terminates according to an information-gain criterion during each trial, and that the learned distribution of stimuli ranks at the time of termination is what the observer uses to emit an overt label response.

The relative magnitude representation. Rather than assume the existence of a stable internal magnitude scale, or a binary 'bigger than the previous' representation, we propose that observers actually store, upon each stimulus presentation, an instantaneous probabilistic judgment of how much bigger (or smaller) the incoming stimulus is with respect to its predecessor. In the same way that a beam balance need not know the actual weights it is loaded with to tell which is bigger, this internal representation need not require the existence of an internal magnitude scale in order to be coherently calculated for local pairwise comparisons. In the same way that the deflection of the beam balance scale, or even the velocity of its change, can tell us how unequal the two weights are, this probabilistic representation of the pairwise difference between successive stimuli will contain more information than a simple binary judgment. For any pair of successive observations $\{x_{t-1}, x_t\}$, we denote this probabilistic container of relative magnitude $p(r|x, o = \{x_{t-1}, x_t\})$, where r takes on the interpretation of magnitude.

Unlike absolute magnitude accounts, which assume the existence of long-term absolute internal referents for stimuli of specific magnitudes, relative magnitude judgment requires only the existence of a short-term (1-back) absolute internal referent. On the typical time-scales of absolute identification experiments, such a referent could stem either from perceptual hysteresis, or from short term memory. Once the comparison of the new stimulus with the referent is made, the referent dissipates on the time-scale of either mechanism, without affecting the information content of the relative magnitude judgment, which is stored in long-term memory. The nature of this mental representation matches exactly the intriguing phenomenology that triggered the original absolute identification experiments - people are much better at making immediate local comparisons (discrimination) but poorer at long-term identification. If the shorter timescale discrimination is conducted via perceptual processes, while identification requires more elaborate memory-accumulation, this difference becomes easy to explain.

Bayesian stimulus-label mapping. Given this assumption about the nature of the long-term internal referent, an observer's goal in absolute identification is to extract a relative magnitude judgment across stimuli in the stimulus set given access to a history of pairwise relative magnitude observations, and to do so using their own history of stimulus exposure within the task. We model the stimulus-label mapping process in the absolute identification task as Bayesian marginalization over relative magnitude judgments seen in pairwise comparisons (Srivastava, Vul, & Schrater, 2014). The mathematical machinery of sequential Bayesian updating allows us to formalize this learning process sequentially on a trial-by-trial, instead of treating the stimulus-label mapping and experimental responding as separate events as is classically done. A graphical view of this process is illustrated in Figure 2.

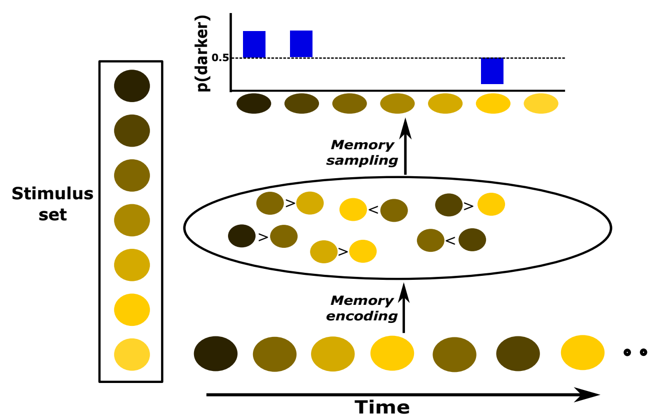

Figure 2: An approximately Bayesian learning model of absolute identification. Observers encode a history of relative magnitude comparisons into memory, and sample from this set of memories to sequentially learn the relative rank order of individual stimuli upon repeated presentation.

The relative magnitude of each stimulus, as we describe above, takes on a probabilistic interpretation formally expressed as $p(r|x, o)$, where r is the relative magnitude judgment, x is the currently visible stimulus, and $o = \{x_{t-1}, x_t\}$ is the relevant comparison *observation*. The ideal Bayesian observer learns $p(r|x,o)$ by combining comparison information from all previously observed comparisons. Thus, this quantity is obtained by marginalizing over the set of previously seen unique observations in memory $\mathcal{C} = \mathcal{P}(\mathcal{X}), s.t. \forall c \in \mathcal{C}, |c| = 2$ which we denote the memorized *comparisons*. Then,

$$D(x) = p(r|x,o) = \frac{\sum_c^{\mathcal{C}} p(r|x,c)p(x|c)p(c|o)}{\sum_c^{\mathcal{C}} p(x|c)p(c|o)}, \tag{1}$$

where it is understood that the *comparison* probability $p(c|o) = p(c|\{o_1, o_2, \cdots, o_{t-1}\})$ is a distribution on the set of all comparisons available from observation history. Here, $p(r|x,c)$ encodes the probability that the item x was found to be larger in the comparison c, $p(x|c)$ encodes the probability that the item x was present in the context c and $p(c)$ encodes the frequency with which the observer encounters these comparisons during the experiment. This frequency is updated via recursive Bayesian estimation,

$$p(c^{(t)}|o^{(1:t)}) = \frac{p(o^{(t)}|c)p(c|o^{(1:t-1)})}{\sum_c^{\mathcal{C}} p(o^{(t)}|c)p(c|o^{(1:t-1)})}. \tag{2}$$

This completes the computational description of the task an ideal Bayesian observer would perform in service of absolute identification, given access to local relative magnitude judgments. The practical approximation arises when we explicitly model the act of accessing previous relative magnitude judgments as memory sampling.

Self-terminating memory sampling. Evidence accumulation influences the shape of the distribution $p(c|o)$ via memory sampling. We model the process of memory recall as the activation of a subset $\mathcal{Q}$ of decision-relevant memory engrams. Using this notation, a general memory accumulation model could be expressed as,

$$p(c) = \sum_{q \in \mathcal{Q}} p(c|q)p(q), \tag{3}$$

where $c \in \mathcal{C}$ are stimuli comparisons available in memory and $q \in \mathcal{Q}$ are memory engrams corresponding to past relative magnitude judgments. Here, the probability distribution $p(q)$ - which we call the *memory prior* - encodes the likelihood of recalling the memory of experience q, while the distribution $p(c|q)$ encodes the knowledge of having seen c and its corresponding relative magnitude judgment stored in the memory engram q. For simplicity, we assume a trivial bijective mapping between c and q - each memory engram is assumed to be associated with a unique stimuli pair.

This memory-sampling variant of $p(c|o)$ plugs directly as the prior in the Bayesian comparison probability update for $p(c|o)$ in Equation 2, which then itself plugs into the two computations in Equations 1 and 2 that define the ideal observer model. This replacement is facilitated by one additional assumption: that the comparison-specific memories recalled are episodic, and therefore convey all comparison-relevant information once the comparison episode itself has been activated in memory[1].

Finally, we formalize our information-theoretic criterion for terminating memory sampling and emitting an identification response. We assume that observers continue to sample memory engrams until the rate at which these provide new information subsides below a threshold. Since the number of unique engrams is limited, the total information available in memory is finite, and any sampling strategy is bound to asymptote. Additional information gained by adding an additional engram q_n to the existing set can be expressed as,

$$IG(q_n) = \sum_i p(c_i|q_{1:n-1}) \log \frac{p(c_i|q_{1:n-1})}{p(c_i|q_{1:n})}, \tag{4}$$

so our sampling termination rule is,

$$\arg\min_n IG(q_n) < T, \tag{5}$$

where T is the termination threshold, constrained by imprecision induced by the influence of processing noise ε such that $T > \varepsilon > 0$, and also potentially informed by exogenous influences.

The observer's choice is determined from the relative magnitude judgment across all x available at the time memory sampling is terminated. We count instances where the observer's decision variable predicts the correct rank of the stimulus introduced on individual trials as accurate responses. Samples to termination are directly interpreted as response times.

Simulation Results

A complete evaluation of the capabilities of this model requires considerably more space than is available in the present format. We restrict ourselves to demonstrating that the model produces reasonable patterns of behavior by replicating well-established empirical benchmarks for absolute identification models. First, we demonstrate the ability of our model to qualitatively replicate the absolute identification results of (Lacouture & Marley, 2004), which are a common benchmark for evidence accumulation models (Brown & Heathcote, 2008). Then, we show how it can replicate a harder pattern of behavior - the crossover effect in RT (Luce, Nosofsky, Green, & Smith, 1982; Brown & Heathcote, 2008).

Reproducing the bowtie effect

Modelling our *in silico* experiment design after the design reported in (Lacouture & Marley, 2004), we showed 20 instances of the model 40 copies each of $N = 7$ stimuli, and asking them to assign number labels $1 \cdots n$ to them. On each trial,

[1]This assumption simplifies our analysis by ignoring the memory dependence of our other intermediate probability terms. While it is likely that such dependence exists, its effects will work in the same direction as the basic results of our approach, since it would further impoverish the preference representation we are already imposing sampling constraints on.

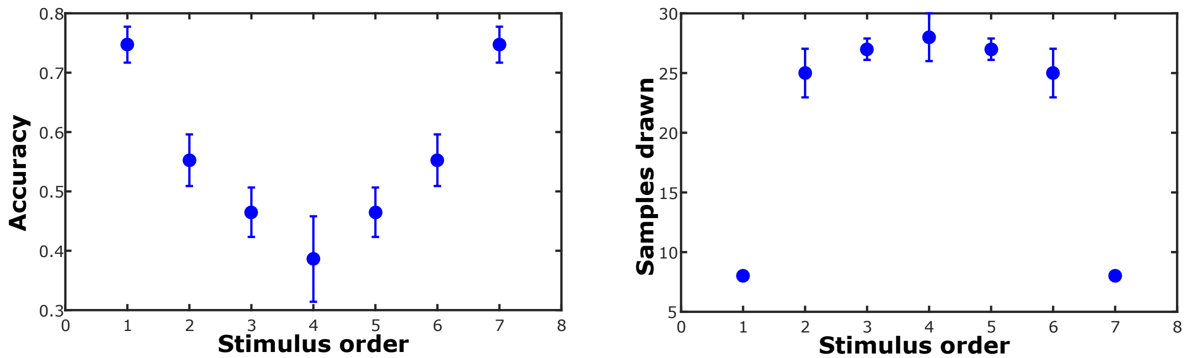

Figure 3: Combining the U-shaped pattern of informativeness with an information-based threshold yields patterns of (left) accuracy and (right) response time as a function of stimulus order. Simulated experiments were conducted using 40 trials per stimulus per session; results are shown averaged across 20 simulation sessions emulating 20 identical observers. Error bars represent ± 1 s.e.m. across these simulated observers.

agents updated their estimates for $p(r|x_t, o = \{x_t, x_{t-1}\})$ following the model described above. Since we assumed equal spacing on a log scale for stimuli as in the original experiment, we kept the relative magnitude judgments as 1 for simplicity, and used a threshold value $T = 0.001$. While our results are presented using a smaller stimulus set than the original experiment, the qualitative trends match (see Figure 3). Accuracy exhibits a convex relationship with stimulus order, and the response time distribution is concave, matching the profiles observed by (Lacouture & Marley, 2004) (also see Figure 10 in (Brown & Heathcote, 2008)).

The model's capacity to identify absolute stimuli arises from differences in the informativeness of memory samples corresponding to various stimuli. Because the evidence from comparisons involving extreme stimuli consistently points the same way, the marginal information gain from sampling saturates rapidly and the model terminates memory retrieval sooner, leading to faster RTs and accurate responses. On the other hand, for stimuli closer to the middle, samples will be split between comparisons where the stimulus is larger and ones where it is smaller, resulting in greater decision variable volatility, and hence, more sampling. This interaction then manifests summarily as slower RTs and noisier responses. In contrast with conventional DDM models, that require $O(N^2)$ parameters to fit decision processes with N alternatives, our model requires just one to reproduce the central finding of the multi-alternative perceptual decision-making literature - the psychophysical bowtie effect.

Reproducing the crossover effect

The crossover effect describes a more complicated pattern of behavior typically seen in perceptual choice experiments. When choice is easy and speed is emphasized, incorrect responses are quicker than correct responses; when choices are harder and accuracy is emphasized, the opposite is true (Brown & Heathcote, 2008). This pattern of RT behavior

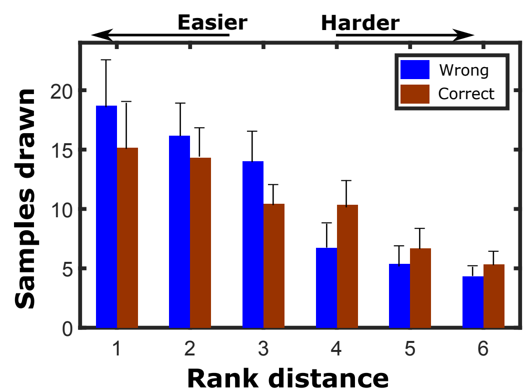

Figure 4: Replication of the crossover effect in perceptual choice.The x-axis plots the rank distance between compared stimuli on a given trial, and the y-axis plots the average number of samples drawn before responding during 20 model runs. Error bars represent ± 1 s.e.m. across these model runs.

has proved very challenging for several models of choice RT to fit, and is a challenging benchmark for models in this field.

Perceptual choice fits into our framework without affecting the formalism in the slightest. The only difference is that the observations o now represent two stimuli seen together instead of sequentially. All the other interpretations remain identical to those in the identification setting. We conducted *in silico* experiments using the same simulation setup as above. As Figure 4 illustrates, our model displays a crossover effect even ignoring the effect of the speed-accuracy tradeoff.

Further, our model offers a straightforward parameter-free explanation for the crossover effect. Simple choices correspond to situations where most samples in memory point in the same direction for a particular stimulus. In such cases, the only way the model could fail to produce the correct response is if the sampling was terminated prematurely. Thus, incorrect responses for simple choices have to be fast. Given sufficient time for integration, it would be impossible for the model to be incorrect. Hard choices correspond to situations where both options have memory samples supporting their case for being bigger. In such cases, the model is biased towards terminating when the marginal information gain is low. Thus, the model will fail to terminate when memory sampling fails to resolve to a modal response, which is more likely when the sampling has failed to discover the true mode of the relative magnitude judgment distribution, resulting in bigger RTs for errors.

Discussion

The results described above suggest that, at a minimum, the model presented in this paper functions as a reasonable model for absolute identification. The fact that our model reproduces the standard bowtie effect using a single parameter suggests that it is capturing a core intuition that existing DDM models struggle to incorporate using many more parameters - errors in the accumulation process occur due to sampling errors in memory retrieval, posterior distributions for difficult trials have greater entropy and so take more samples to resolve. Further work is obviously needed to verify that the theoretical novelty of the *relative* magnitude representation actually provides better fits to data for problem cases in existing models, such as the ones we highlighted above.

A priori, the representational flexibility provided by this representation, in conjunction with the fact that the representation itself is learned from trial-by-trial stimulus observations, is expected to reproduce both the shift in response patterns as a function of stimulus set shown by (Rouder et al., 2004) and the heightened response to unequal spacing documented by (Brown et al., 2009). The latter, in particular, can be obtained using a simple modification of the present experimental setup - changing the value of the local relative magnitude judgment to reflect the impact of unequal spacing, instead of keeping it uniform in the presence of equal spacing as in our demonstrations. Mathematically, changing this term simply constitutes an over-weighting of the relevant memory engrams in the accumulation process, which should easily yield lower error rates around these gaps. Crucially, our account predicts that this effect should also propagate to the RT distribution - an easily testable prediction.

Sequentially modelling the mapping process, in conjunction with the use of an information-based stopping criterion, sheds new light on the relationship between the psychophysical bowtie effect seen in identification experiments (Lacouture & Marley, 2004) and the economic distance effect (Dickhaut et al., 2013) observed in multiple behavioral economics experiments where it is found that extreme choice valence (distance in utility) appears to be correlated with lower error rate, response times and interestingly, levels of neuronal activation as measured by fMRI (Dickhaut et al., 2013). According to our model, constructing a decision variable using conflicting evidence requires more samples to breach the information-based threshold, resulting in greater effort, correlated with higher RT and brain activation for both perceptual and economic choices with greater mutual confusability, as determined by their history of pairwise comparisons.

Acknowledgments

We acknowledge financial support from the Research I Foundation.

References

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.

Brown, S. D., Marley, A., Dodds, P., & Heathcote, A. (2009). Purely relative models cannot provide a general account of absolute identification. *Psychonomic Bulletin & Review*, *16*(3), 583–593.

Brown, S. D., Marley, A., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological review*, *115*(2), 396.

Dickhaut, J., Smith, V., Xin, B., & Rustichini, A. (2013). Human economic choice as costly information processing. *Journal of Economic Behavior & Organization*, *94*, 206–221.

Dodds, P., Donkin, C., Brown, S. D., & Heathcote, A. (2011). Increasing capacity: Practice effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 477.

Lacouture, Y., & Marley, A. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology*, *39*(4), 383–395.

Lacouture, Y., & Marley, A. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & Psychophysics*, *66*(7), 1206–1226.

Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Attention, Perception, & Psychophysics*, *32*(5), 397–408.

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.

Rouder, J. N., Morey, R. D., Cowan, N., & Pealtz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review*, *11*(5), 938–944.

Sims, C. R. (2016). Rate–distortion theory and human perception. *Cognition*, *152*, 181–198.

Srivastava, N., Vul, E., & Schrater, P. R. (2014). Magnitude-sensitive preference formation. In *Advances in neural information processing systems* (pp. 1080–1088).

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