Psychological Review

A Unified Theory of Discrete and Continuous Responding

Peter D. Kvam, A. A. J. Marley, and Andrew Heathcote

Online First Publication, July 21, 2022. http://dx.doi.org/10.1037/rev0000378

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Review

ISSN: 0033-295X

https://doi.org/10.1037/rev0000378

A Unified Theory of Discrete and Continuous Responding

Peter D. Kvam[1], A. A. J. Marley[2], and Andrew Heathcote[3, 4]

[1] Department of Psychology, University of Florida
[2] Department of Psychology, University of Victoria
[3] School of Psychology, University of Newcastle
[4] Department of Psychology, University of Amsterdam

Understanding the cognitive processes underlying choice requires theories that can disentangle the representation of stimuli from the processes that map these representations onto observed responses. We develop a dynamic theory of how stimuli are mapped onto discrete (choice) and onto continuous response scales. It proposes that the mapping from a stimulus to an internal representation and then to an evidence accumulation process is accomplished using multiple reference points or "anchors." Evidence is accumulated until a threshold amount for a particular response is obtained, with the relative balance of support for each anchor at that time determining the response. We tested this multiple anchored accumulation theory (MAAT) using the results of two experiments requiring discrete or continuous responses to line length and color stimuli. We manipulated the number of options for discrete responses, the number of different stimuli, and the similarity among them, and compared the outcomes to continuous response conditions. We show that MAAT accounts for several key phenomena: more accurate, faster, and more skewed distributions of responses near the ends of a response scale; lower accuracy and slower responses as the number of discrete choice options increases; and longer response times and lower accuracy when alternative responses are more similar to the target response. Our empirical and modeling results suggest that discrete and continuous response tasks can share a common evidence representation, and that the decision process is sensitive to the perceived similarity among the response options.

Keywords: continuous report, cognitive modeling, perception, absolute identification, multi-alternative choice

Supplemental materials: https://doi.org/10.1037/rev0000378.supp

Both discrete and continuous response scales are used in many tasks that measure and/or assess components of psychological functioning, ranging over assessments of personality, physical and mental health, beliefs, opinions, confidence, and performance related to constructs such as intelligence, executive control, and working memory. Likert (1932) scales have a long history of extensive use in such areas, but there are many known limitations of such discrete measures (Paulhus, 1991). The addition of a continuous measure like response time or confidence, or additional responses like best-worst rankings, provide richer insights into the cognitive processes underlying responses and allow a researcher to make inferences that are not possible with simple binary choice data (Hawkins et al., 2014; Louviere et al., 2015; Reynolds, Kvam, et al., 2020). Of course, the degree to which we can determine whether the true underlying constructs we seek to measure are discrete or continuous is a major challenge (Luce, 1997), and there are cases like item response theory where we would like to use a discrete scale because the different levels of a continuum cannot be fully distinguished (Hambleton & Swaminathan, 2013). Across all of these contexts, a particular challenge for models of psychological and cognitive processes is to provide a unified account of how continuous underlying psychological constructs (e.g., confidence, strength of elief, commitment) or percepts (pitch, timbre, length, color) are mapped onto continuous and discrete response scales (Luce, 1997; Townsend, 2008). To address this challenge, we develop a modeling approach for, and empirical evidence about,

Editor's Note. Han L. J. Van der Maas served as the action editor for this article.—ELG

Peter D. Kvam https://orcid.org/0000-0002-3195-8452

Part of the material contained in this article was presented at the 2018 Midwest Cognitive Science and 2017 Society for Mathematical Psychology conferences, and the data set from Study 2 was part of the first author's PhD dissertation. Study materials and code for this article can be found on the Open Science Framework at https://osf.io/6d29q. This study was not preregistered.

This research has been supported by Natural Science and Engineering Research Council Discovery Grant 8124-98 to the University of Victoria for A. A. J. Marley, a National Science Foundation Graduate Research Fellowship and international research opportunity Grant DGE-1424871 to Peter D. Kvam, and an ARC Professorial Fellowship to Andrew Heathcote (DP110100234).

The second author, A. A. J. Marley, passed away shortly before this article was submitted. We are extremely grateful for his guidance, his insights on the anchor-based approach to modeling that were fundamental to the creation of multiple anchored accumulation theory (MAAT), his comments on this work, and to him personally as a colleague and collaborator.

Correspondence concerning this article should be addressed to Peter D. Kvam, Department of Psychology, University of Florida, 945 Center Dr, P.O. Box 112250, Gainesville, FL 32607, United States. Email: pkvam@ufl.edu

tasks where participants choose among discrete options and provide values on a continuum of responses for those same options.

The key to developing a unified account is understanding how people map internal representations onto external responses, controlling for the systematic biases or distortions that occur during the mapping process (Zotov et al., 2010). Cognitive models have been developed for several types of graded response measures, capturing the responses and response times associated with confidence and probability estimates (Busemeyer, Kvam, et al., 2019; Pleskac & Busemeyer, 2010; Ratcliff & Starns, 2009, 2013; Smith & Van Zandt, 2000), preference ratings (Bhatia & Pleskac, 2019; Kvam et al., 2021), and opt-out (deferred or "don't know") responses (Bhatia & Mullett, 2016; Reynolds, Garton, et al., 2021). Incorporating response times is important to understanding how these types of responses are generated and to accurately interpreting them, as an overwhelming body of empirical evidence shows that the time at which a response is made has a nontrivial interaction with the type or magnitude of the response and how diagnostic it is (Baranski & Petrusic, 1998; Donkin, Brown, Heathcote, & Marley, 2009; Luce et al., 1982; Yu et al., 2015). Thus, a complete account of behavior on mapping tasks must predict the joint distributions of responses and the time it takes to make them.

In the past, models have mainly focused on discrete-valued (choice) responses, but continuous-valued responses can theoretically confer more information about internal representations. Recent developments of models for tasks involving continuous responses (Kvam, 2019a; Ratcliff, 2018; Smith, 2016) have enhanced our understanding of the cognitive processes underlying these tasks and have begun to be applied successfully in domains such as orientation estimation (Kvam, 2019b; Ratcliff, 2018), color identification (Ratcliff, 2018; Smith et al., 2020), numeracy (Ratcliff & McKoon, 2020), and pricing (Kvam & Busemeyer, 2020). Critically, relations between continuous and discrete response measures have not been thoroughly explored. In this article, we examine the relationship between different response scales, and evaluate how (if) the response scale influences the representation of decision evidence and decision strategies.

Although the model we develop has the potential to be applied to almost any type of response and response time data, we focus here on using it to tie together continuous and discrete response formats for two fundamental types of paradigms that require participants to take a perceived stimulus and map it onto a response, as in traditional absolute judgment tasks. Such paradigms have been modeled by extensions of choice RT models (Nosofsky, 1997) and used to address questions about information processing capacity (Miller, 1956), the relationship between the number of stimuli (and/or responses) and accuracy (Luce et al., 1982), and the relationship between response times and the response chosen (Kent & Lamberts, 2005; Lacouture & Marley, 1991, 1995, 2004). Absolute judgments, therefore, provide an ideal test bed for a general model of mapping tasks, revealing fundamental phenomena that will also impact more complex tasks and measures. A natural question is whether mappings onto continuous or discrete scales involve fundamentally different cognitive processes. Busemeyer et al. (1997) reviewed learning differences in discrete and continuous response tasks, including a number of diverging patterns for categorical versus continuous function learning, and developed a computational model of the disparate systems associated with discrete and continuous tasks. However, this model addressed the "front-end" differences in learning between the tasks, and not the "back-end" response processes involved in the two paradigms. This leaves unresolved the connection between continuous and discrete response scales, which we address here.

Extrapolating to the Continuum

We might expect that a continuum of responses—as used in function learning tasks, for instance (Koh & Meyer, 1991)—would arise as the limiting case as the number of categories in a discrete-category task grows very large and fine-grained. This is arguably the most coherent relationship between discrete and continuous response paradigms (see Kvam, 2019a; Schurgin et al., 2020). In this case, we could simply examine how the parameters of our chosen model shift as the number of response options increases and extrapolate to the asymptotic case where the number of response options approaches infinity (or at least, hits the number of pixels on the response scale or the limits of spatial discrimination for human decision-makers). However, the continuous case as an asymptotic limit can be somewhat problematic for models initially formulated for discrete responses. Using Hick's (logarithmic) law as an example (Hick, 1952), the lack of an asymptote as the number of responses increases forces the bizarre prediction that response times will become infinitely long in the continuous case. Naturally, this prediction does not pan out in empirical response time data, as violations of Hick's law have been observed when there are (substantially) more than eight alternatives (Longstreth, 1988; Seibel, 1963).

Recent cognitive models are more promising in terms of identifying candidate processes that may differ for discrete versus continuous response scales. For instance, Usher et al. (2002) suggested that shifts in mean response times as a function of the number of response alternatives could be primarily attributed to changes in the threshold for making a decision. Due to the space of response options becoming more "crowded" and baseline or guessing accuracy decreasing as more alternatives are added (see Schurgin et al., 2020; Van Maanen et al., 2012), models that predict responses as a function of racing accumulators must increase the amount of evidence gathered before making a decision in order to maintain a desired level of accuracy and speed (Hawkins et al., 2012a) or optimize time on the task (Hawkins et al., 2012b). In this case, we might expect a model of continuous response measures to simply implement a different threshold that sets the trade-off between speed and accuracy. To preview the results of our experiments, threshold shifts wind up being a plausible explanation for differences in performance between tasks with continuous versus discrete responses, making it straightforward to jointly model the two types of paradigms.

To test whether differences in performance among tasks with different numbers of responses can be attributed to threshold shifts, we develop a model that uses a common underlying evidence accumulation process to generate either discrete or continuous responses. In the first study, we examine how this approach can account for changes in accuracy and response time across an evenly spaced span of stimuli ("bow" effects) and across different numbers of response options ("set-size" effects) with unidimensional (line length) stimuli. Our model utilizes a pair of "anchors," one at each end of the range in which the stimuli lie, which correspond to exemplars of very short or very long stimuli based on the stimuli a participant had seen (Brown et al., 2008; Lacouture & Marley, 2004; Marley & Cook, 1984, 1986; Petrov & Anderson, 2005). Note that a

participant would learn in the practice trials exactly how long stimuli in the longest/shortest categories would be, so they were well-calibrated to this range. The location of the stimulus relative to these anchors drives an evidence accumulation process. As desired or elicited, either discrete or continuous responses are triggered by a stopping rule based on the total amount of evidence that has been gathered, with the chosen response determined by the balance of evidence with respect to each anchor.

In the second study, the stimuli vary in hue and the data are modeled with three anchors, one each for red, green, and blue stimuli. Previous studies have shown that choosing the location of the brightest (target) stimulus is slower and less accurate when one of the distractor stimuli is a "near competitor" (i.e., similar in brightness to the target; Teodorescu & Usher, 2013). We find an analogous near-competitor effect for hue, which follows from our model because target options with near competitors (i.e., similar response options) require more support to be chosen than do target options that are distinct from other options. This result violates a core assumption of several continuous models; namely, that stopping rules (i.e., the threshold amount of evidence required to trigger a response) should be consistent across options. Instead, we suggest that the similarity between response options should be a fundamental consideration in selecting a stopping rule for each option in a discrete or continuous set of alternatives.

New Predictions

The key component allowing our multiple anchored accumulation theory (MAAT) to accommodate different numbers of discrete responses is its ability to divide the stimulus representation into separate categories but map this representation onto a continuum. In some ways, this is similar to general recognition theory, where a continuous feature space is divided into categories corresponding to regions bounded by hyperplanes (Ashby, 2000; Smith, 2019). The distinguishing feature of MAAT is how support for different categories is gathered and weighed against one another—the support for each choice is the relevant component of a vector defining the corresponding option. The location of the evidence vector (and hence its component along each of the option vectors) changes dynamically over time as a person considers the stimulus. This allows for decision rules that include the relative degree of support for each choice option based on the similarity relations between them (e.g., it is easier to select between options that are distinct, as opposed to options that are very similar).

MAAT makes several predictions that are quite different to those of other approaches to modeling continuous and discrete responses (Ratcliff, 2018; Smith, 2016). The first is that the distribution of continuous responses should become more skewed as the target stimulus gets closer to upper or lower anchors (see also Kvam & Turner, 2021, for a discussion of how this can be described in terms of representational similarity). As a result, responses near the ends of a scale should be faster and more accurate. That is, there will be a "bow" effect in performance as a function of stimulus magnitude (Luce et al., 1982).

Second, the distribution of responses should be wider when they are further away from an anchor point. This is related to the bow effects and naturally arises when participants have a point of reference against which they can compare and contrast a presented stimulus. Responses near a point of reference tend to be closely grouped and relatively precise (Hollands & Dyre, 2000), while those in between points of references tend to be relatively uncertain. In our model, this results from how the rate of evidence accumulation is determined by comparing the stimulus to the available anchors (memory traces or exemplars). In Supplemental Material, we show formally that the entropy of the accumulation process in MAAT increases as the distance between a stimulus and the anchors becomes greater.

Third, MAAT predicts that it is more difficult to select a particular response when it has similar competitors compared to when its competitors are very dissimilar. In this case, participants may set a greater threshold for response options with similar competitors in an attempt to improve accuracy, although at the cost of longer response times (Usher et al., 2002). Importantly, prominent theories of continuous responding, including the circular diffusion model (Smith, 2016, 2019; Smith et al., 2020) and spatially continuous diffusion model (Ratcliff, 2018; Ratcliff & McKoon, 2020) have only a single threshold or an invariant function (e.g., in the color experiments of Ratcliff, 2018, there are higher thresholds for nonprimary/nonsecondary colors) specifying the thresholds for all response options. Consequently, a set of unequally spaced response options, say A, B, and C, with C being further away (i.e., A-B—-C) must have the same thresholds for each option. In contrast to these other theories, MAAT is able to explain not only reduced accuracy for responses with near competitors, but also the ability of participants to maintain or even increase accuracy by selectively slowing such responses.

In the following sections, we apply MAAT to data from both discrete and continuous mapping tasks. In the first experiment, we use line-length stimuli, demonstrating the model's ability to provide a simple and tractable account of perceptual judgments on a unidimensional scale with anchors at either end. The model of this first task is fit only to accuracy data (i.e, was a response within or outside the correct range?), and the distribution of continuous responses is used to provide as an "out-of-sample" (cross-validation) test. In the second experiment, we use a hue-based task requiring responses on a color circle. We instantiate a simple theory of vision to account for how people perceive the stimulus, with anchors for red, green, and blue, but otherwise employ the same core principles as used for the first task. To show the generality of the model and fitting approach, we fit the model to the precise distribution of responses (multinomial in the discrete condition, and continuous in the continuous condition). Together, the two experiments and accompanying models elucidate how and why behavior changes with different response-set sizes.

Model Overview

Although our proposal uses a novel accumulation structure that allows it to model behavior on both discrete and continuous tasks, it invokes well-developed cognitive mechanisms from previous models of absolute identification, including dynamic models like the selective attention, mapping, and ballistic accumulation (SAMBA) model (Brown et al., 2008) and the relative judgment model (Stewart et al., 2005). It then extends these mechanisms to account for data involving continuous responses. The main hurdle to overcome with respect to previous models it that they associate a unique accumulator (and corresponding accumulation rate and threshold) with each possible response, which is difficult to translate to

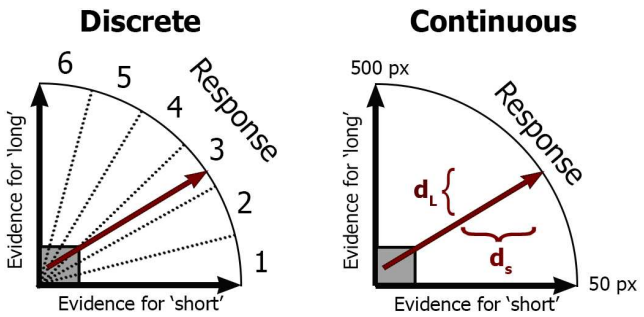
scenarios where the number of responses is very large. To yield an approach that simultaneously handles both continuous and discrete responses, we instead base our new model on the geometric (Kvam, 2019a) and multiple threshold race (Reynolds, Garton, et al., 2021; Reynolds, Kvam, et al., 2020) frameworks for modeling dynamic choice.

In unidimensional versions of these frameworks, evidence accumulation is represented in terms of two numbers: a balance of evidence among the options (How much does the evidence favor a response toward one end of the scale vs. the other end?) and a total amount of evidence (How strong are these beliefs/representations?). The information that a decision-maker accumulates changes on both dimensions over time, leading to different response options having different strengths over time, and increasing the overall amount of information that has been considered. An option is chosen when it has enough support, that is, when the match between a person's evidence state and the description of the choice alternatives align well and there is sufficient information to make a decision.

This type of evidence representation can be formally depicted in two dimensions. If "no information" corresponds to an evidence state at [0, 0], then the distance from the origin describes how much information has been collected, and the direction of the evidence state relative to the origin describes the option that is most favored at that moment. For example, in the first task that we study, the position of the state in the *x*-direction corresponds to the strength of evidence for responses at one end of the scale (i.e., a"short" responses) and the position of the state in the *y*-direction corresponds to the strength of evidence for responses at the opposing end of the scale (i.e., "long" responses) as shown in Figure 1. The response selected is determined by the *ratio* of these two dimensions (i.e., the angle), while response time is determined by the (squared) *sum* of the dimensions (i.e., the distance from the origin). In other words, a person enters a response when they have gathered enough information, but the response that they make is determined by the balance of evidence between support for short and long responses, reflecting Vickers's (Vickers, 2001; Vickers & Packer, 1982) seminal ideas relating to confidence judgments.

This framework can be contrasted with the standard assumption of many diffusion-based models, which track only the balance of evidence—information favoring one relative to another option. Theories producing graded estimates based only on the balance of evidence between two options can fail to produce magnitude effects (Miletić et al., 2021; Teodorescu et al., 2016) where increasing the magnitude of both stimuli/choice options (e.g., making both more coherent or easy to see) while maintaining the balance between them speeds up response times (but see Ratcliff et al., 2018, for an approach to this issue that makes variability in the balance of evidence proportional to its mean). As is true racing accumulator models in general (Heathcote & Matzke, in press), having more than one dimension to the evidence accumulation process allows MAAT to capture magnitude effect as well as typical difference effects, where adjusting the balance of evidence by manipulating the ratio of support for two options mainly affects the responses that are given rather than response times (Reynolds, Kvam, et al., 2020; Vickers, 2001).

Figure 1
Diagram of Discrete (Left) and Continuous (Right) Response Models



Note. Evidence accumulates over time (red arrow) from its starting point (gray box) according to the drift rates for each anchor, in this case v_S for short responses and v_L for long responses. A decision is made when sufficient support for one of the responses is gathered, corresponding to the location and time at which the accumulation process crosses the quarter-circular boundary. This particular model diagram corresponds to the one used in Experiment 1, whereas a full circle is used in Experiment 2. See the online article for the color version of this figure.

General Model Specification

We propose an "opponent processes" account of evidence accumulation during decision-making, where choices are driven by competing sources of information relative to *anchors* (Marley & Cook, 1984). In the first experiment, short and long anchors are represented as direction vectors at 0° and 90°, respectively. In the second experiment, red/green/blue are represented by directions vectors at 0°/120°/240°, respectively. Responses that are in between these anchors are represented by intermediate angles. For example, a medium line length in the first experiment might be at 45°, and a yellow stimulus in the second experiment might be located at 60°.

The state of the information that a decision-maker has at a particular point in time is described by a point, which can be defined by an x and y coordinate in two dimensions for the experiments and models presented in this article (although in principle they can be higher-dimensional; Kvam & Turner, 2021). The degree of support for a particular response is described by the match between the decision-maker's state and a vector v defining the response. Formally, it is the component of the state vector $\boldsymbol{s}$ along a vector $\boldsymbol{v}$ describing the choice option, $\text{comp}_{\boldsymbol{v}}(s)$. For example, suppose a decision-maker is choosing among different orientations in an orientation-detection task, and has two options: choice Option A is in direction $\boldsymbol{v}_A = [1, 0]$ (0°) and choice Option B is in direction $\boldsymbol{v}_B = [\sqrt{.7}, \sqrt{.3}]$. If the accumulated evidence corresponds to a state $s = [.5, .2]$, then the degree of support for Option A is equal to $\text{comp}_{vA}(s) = .5$, while the degree of support for Option B is equal to $\text{comp}_{vB}(s) \approx 0.53$. Therefore, the evidence state slightly favors Option B over Option A.

The state s changes over time according to how well the stimulus matches each of the anchors and the overall rate at which the decision maker gathers information. Its dynamics are determined by the degree of activation relative to each of the anchors, with such activation bounded between 0 and 1. The activation values vary randomly from trial to trial according to a β distribution, $\beta(cz, c(1-z))$, where z is the match between the stimulus and the anchor, and $c > 0$ is the precision of the representation. A value of $z = 1$ indicates a perfect match between the stimulus and the anchor and $z = 0$ indicates a perfect mismatch. The different ways in which the match is calculated for line

length stimuli and for hue stimuli are detailed below. Precision describes how consistently participants can discriminate between stimuli: larger values of c correspond to representations that are more precise (i.e., vary less from trial to trial) and hence support better discrimination. We call c a "precision coefficient" to reflect the idea that lower information variability is required to form more precise representations (Hick, 1952).

The model dynamics unfold as different anchors pull the state in different directions. The overall movement of the state is determined by the balance of activation across the anchors, taking into account the conflict in direction. Formally, the dynamics of the state are determined by a vector δ_{all} that is the weighted sum over anchors of unit length vectors $\boldsymbol{d}_i$ defining each anchor's direction. The weight for each anchor vector is determined by its degree of activation, given by the β distribution:

$$\delta_{all} = \delta \sum_i \boldsymbol{d}_i \cdot \text{Beta}\left(cz_i, c\left(1 - z_i\right)\right). \quad (1)$$

The parameter $\delta > 0$ scales the rate of evidence processing per unit time and can be viewed as a measure of overall information processing or "channel" capacity, with larger values corresponding to a wider channel and thus faster information accumulation (Eidels et al., 2010; Townsend & Wenger, 2004) and shorter response times.

The evidence state changes from its initial position at time 0, $\boldsymbol{s}(0)$, to its position at time t, as a function of δ_{all}. For simplicity, we assume a deterministic accumulation process, as in the linear ballistic accumulator model (LBA; Brown & Heathcote, 2008)[1].

$$\boldsymbol{s}(t) = \boldsymbol{s}(0) + t \cdot \delta_{all}. \quad (2)$$

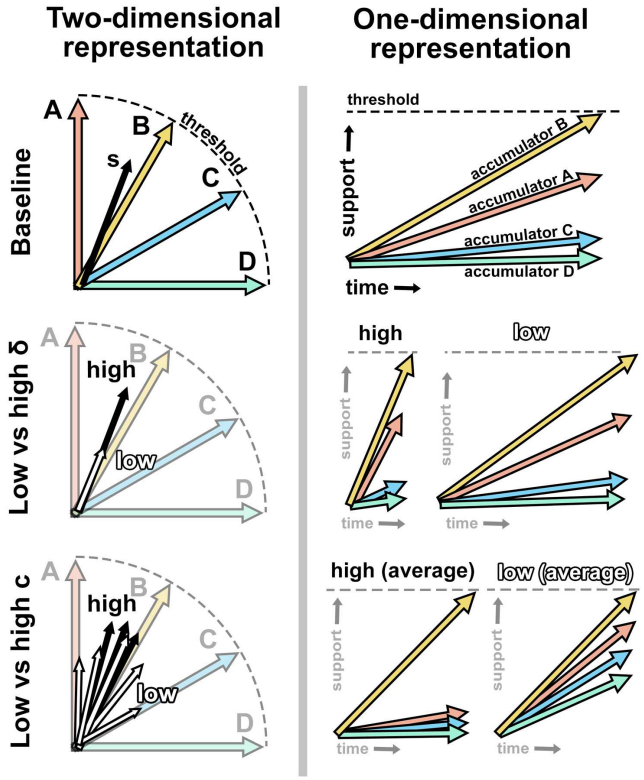
Evidence accumulation stops when one of the response options exceeds a threshold level of support, as determined by the match between the state and the direction describing that response. Formally, evidence accumulation halts and response (option) j is chosen at time t if it is the first option that satisfies the condition $\text{comp}_{\boldsymbol{v},j}(\boldsymbol{s}(t)) > \theta$. Thus, response times are determined by the shortest time t at which this condition is met, and choice is governed by which option j meets it first. The value of θ controls how strict the stopping rule is: as in the LBA lower values of θ result in faster response times but lower accuracy, while higher values result in slower response times but greater accuracy.

The models that we use for our two studies follow this general specification, although there are differences in certain details of the decision process, such as the anchors and the factors that determine the threshold(s) θ. Thus, the general model structure can be adapted to the details of specific task paradigms according to the demands that they place on participants.

An outline of the type of model that we used for the first task is shown in Figure 2. The two anchors—long and short (corresponding to A and D in the figure)—drift the evidence accumulation process until one of the responses along the continuum between them accumulates sufficient support to be selected. For a discrete set of responses, this model can also be instantiated as a competition between multiple correlated accumulators (Kvam, 2019a; Reynolds, Kvam, et al., 2020). This representation is shown in Figure 2 on the right. Readers familiar with competing accumulators can conceptualize drift as describing the average rate of evidence accumulation for all options, while precision is more similar to drift variability or the difference between the best and next-best accumulator(s) (Brown et al., 2009). However, we should caution these readers that as the number of options grows increasingly large, it becomes more difficult to instantiate the model as a competing-accumulator process and so we must defer to the model representation on the left side of Figure 2/right side of Figure 1.

Figure 2

Outline of the Structure of the Two-Dimensional Representation of the Model (Left) and Its Corresponding Racing-Accumulator Representation (Right)



Note. In the top left panel, the black arrow labeled s is the stimulus vector and the coloured arrows are vectors representing different possible choice options. The top panel on the right shows the increase in support for each choice in corresponding accumulators (indicated by having the same color). Note that support for options more closely aligned with s increases more quickly. The effect of manipulating drift (δ, white = low vs. black = high) is shown in the middle panels and the effect of manipulating the precision coefficient (c, white = less precise, black = more precise) is shown in the bottom panels. Drift affects all racing "accumulators" equally, while precision affects the average disparity between the best "accumulator" and its competitors across trials. See the online article for the color version of this figure.

[1] This assumption can be see as an approximation to an underlying system in which evidence is diffusive (i.e., fluctuates from moment to moment) but its effects are dominated by between-trial fluctuations in the average rate of accumulation. We acknowledge that the evidence state could instead be driven by the accumulation of stochastic samples of evidence, and so we still refer to this process as sequential sampling characterized by a "drift" or overall rate of accumulation.

Line Length Identification

The first study tested the connection between discrete and continuous responses in an experiment motivated by classical absolute judgment tasks (Miller, 1953). In absolute judgment, a participant assigns each of a finite number of uni-dimensional stimuli to a different experimenter-defined response; the responses usually have the same ordering as the stimuli. In our study, the stimuli are line lengths defined by the distance between two points on the screen. Participants were tasked with assigning the line length shown to either (a) one of a finite number of discrete categories or (b) a position on a continuous (radial) scale. As shown in Figure 3, in order to make the motor requirements of discrete and continuous responding as similar as possible, the response in each condition was to move a cursor to the selected response location on a semicircle. More details on the task are provided in the Methods section below; first, we introduce the structure of the model used to explain behavior on this task.

Line-Length Model

The model for the line length study is a two-dimensional evidence accumulation process that starts at randomly chosen coordinates in the first quadrant $\boldsymbol{s}(0) = [s_S, s_L]$, which define the initial degree of support relative to the short and long anchors, respectively, with $s_S > 0$ and $s_L > 0$. As a person samples information about the stimulus, this state moves in the x-direction toward "short" responses at a continuous rate d_S, and moves in the y-direction toward "long" responses at a continuous rate d_L. Accumulation terminates when evidence for any response between the shortest and longest stimulus exceeds θ. A diagram of the model is shown in Figure 1. For simplicity of computation, this response-selection process can be approximated as a circular boundary specified as $x^2 + y^2 = \theta^2$ (i.e., accumulation stops when it reaches radius θ from the origin). This results in miniscule differences relative to separate boundaries for each response option.

We assumed the position on the arc that a stimulus maps onto is linearly related to its length, although in general this mapping can be nonlinear (as in utility representations or nonlinear similarity between stimuli, see Kvam & Busemeyer, 2020; Kvam & Turner, 2021). Note that participants are *not* being asked to map each stimulus length onto an arc of equal length. Rather, they are mapping each stimulus length to a position on the semicircle from 180° to 0° with the position determined by the stimulus set.

The angle coordinate at the time when the state hits the response boundary determines the response. However, for the purposes of modeling, where starting points and drift rates are drawn with respect to Cartesian coordinates, we lay out the behavior of the model in terms of (x, y) coordinates. The Cartesian coordinates can then be mapped back into polar coordinates to compute continuous response distributions given the hitting point $[x_{\text{resp}}, y_{\text{resp}}]$ equates to angle $\phi = \tan^{-1}(y_{\text{resp}} / x_{\text{resp}})$. For the first model, the value of x_{resp} and y_{resp} will necessarily be positive, as the accumulation process starts in the first quadrant and accumulates in the positive x and y directions (i.e., no negative drift rates are allowed). For convenience in plotting fits of the theory, a response is calculated by first mapping its value onto [0, 1] by taking $\phi_{01} = \frac{2\phi}{\pi}$, which is then transformed into a bounded line-length response R on a length scale $[R_{\min}, R_{\max}]$ as follows:

$$R = \phi_{01} * (R_{\max} - R_{\min}) + R_{\min}. \tag{3}$$

In the present study, the minimum of the scale is $R_{\min} = 50$ pixels and the maximum of the scale is $R_{\max} = 500$ pixels.

For the discrete-response condition, the values for R are sorted into categories with boundaries $C = \{C_{1,2}, C_{2,3}, \ldots, C_{n-1,n}\}$, where n is the number of categories. We assume unbiased discrete responses by placing the category boundaries evenly from $R_{\min}$ to $R_{\max}$ at $C_{m,m+1} = \frac{m}{n}$ $(0 < m < n$; see the left panel of Figure 1).

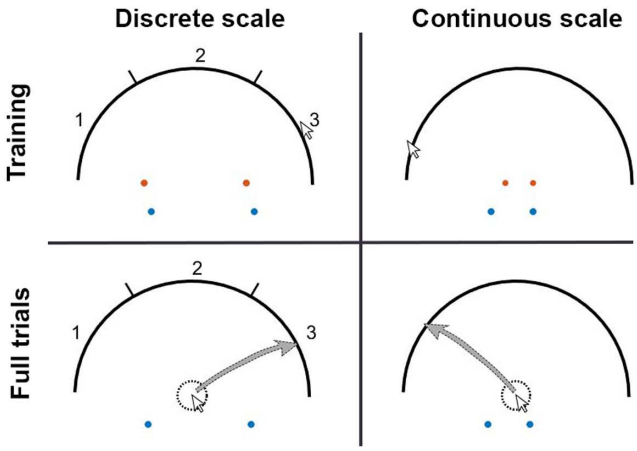
For both the discrete and continuous cases response time is the sum of the time it takes to travel from the starting point (s_S, s_L) to the point where the process hits the boundary θ, and a nondecision time τ, representing the sum of the times to encode the stimulus and produce a motor response.

We now present the detailed assumptions regarding the models starting points, drift rates, and thresholds.

Starting Points

On each trial, the activation relative to each anchor begins with some activation drawn from independent uniform random variables, $S_S \sim \text{Unif}(0, s)$ for the short anchor and $S_L \sim \text{Unif}(0, s)$ for the long anchor. Start-point variability, s, is a free parameter governing the degree of anchor activation before any evidence is collected, indexing activation either due to response biases or leftover activation from previous decisions (Heathcote et al., 2019). This is a fairly typical distribution of starting points utilized in both accumulator (Brown & Heathcote, 2008; Busemeyer, Gluth, et al., 2019) and diffusion (Ratcliff et al., 2016) models.

Figure 3
Diagram of Practice Trials (Top) and Experimental Trials (Bottom) for Discrete (Left) and Continuous (Right) Scales



Note. Blue dots indicate the stimulus, orange dots indicate the response corresponding to the mouse position in practice trials. For experimental trials (bottom), response times were recorded when the cursor moved outside the dotted circle and response location was determined by where the cursor crossed the scale semicircle. Mouse trajectories are shown in grey. For practice trials (top), responses were entered by clicking on the scale semicircle rather than simply crossing it in order to allow participants to match the response to the stimulus. See the online article for the color version of this figure.

Independent random starting points are a simplifying assumption that could be modified. Starting points could be arranged proportional to all of the possible responses as opposed to the anchors (creating a circular start point distribution in the continuous condition, for example). Starting points could also be used to specify systematic response biases, or be specified to reflect sequential effects, such as a dependence on the previous responses, as in models of absolute identification (Brown et al., 2008). For example, a bias favoring the previously chosen category on the immediately following trial might be implemented by fixing the ratio of short/long start points to that of the previously chosen category. This would result in an assimilation effect like that found in absolute judgment paradigms where response times and accuracy are improved when subsequent trials feature stimuli from the same category (Ward & Lockhead, 1970). However, it is also possible that these sequential responses are the result of fluctuations in selective attention (Matthews & Stewart, 2009)—in which case they might be better incorporated into drift rates and/or thresholds (Donkin, Brown, Heathcote, & Marley, 2009; Donkin et al., 2015; Treisman & Williams, 1984). The model we present here is theoretically neutral with respect to which mechanism is responsible for sequential effects, as any of these three mechanisms (start point, drift, or threshold) could potentially change across trials. For the sake of simplicity and computational tractability, and because these effects are not particularly pronounced in the data from the task described below, we do not include mechanisms to model sequential effects here. However, for those readers interested in these effects, we include a section in the Supplemental Material that gives an overview of the sequential effects in our data.

Drift Rates

The main index of the strength of a particular alternative in dynamic models is the drift rate. A typical model of discrete responses/decisions might assign separate drift rates to each of the response options and determine some covariance structure across the response options (the approach taken by Ratcliff, 2018). However, this becomes burdensome for the continuous case. A simpler way to represent the accumulation process is to reduce the large number of accumulators to a smaller number of dimensions and specify a drift rate for each dimension (Kvam & Turner, 2021). Each dimension can then be thought of as being represented by an accumulator. In the present case, there are only two dimensions—defined by the upper and lower anchors $R_{\max}$ and $R_{\min}$—and so we need only two drift rates.

To specify the drift rates for a stimulus relative to the lower and upper anchors (i.e., for the x and y directions), we build on the double-anchor model from the absolute judgment literature, which gives the strength of activation relative to each of the anchors based on the length of a given stimulus (Marley & Cook, 1984, 1986). In this theory, the strength of activation relative to an anchor for a given stimulus is based on the distance between the stimulus, L, and the anchor relative to the range between the two anchors, $R = R_{\max} - R_{\min}$. In our model, this activation is equivalent to the match, z, in Equation 1: for the long anchor $z_L = (L - R_{\min})/R$, and for the short anchor $z_S = (R_{\max} - L)/R$. Formally, the drift rates for short accumulator (δ_S) and long the accumulator (δ_L) are independent samples from β distributions that are both scaled by the channel capacity, δ, and with an overall level of trial-to-trial variability in rates determined by c:

$$v_L = \delta \cdot \text{Beta}\left(c \cdot z_L, c \cdot (1 - z_L) + b\right), \tag{4}$$

$$v_S = \delta \cdot \text{Beta}\left(c \cdot z_S + b, c \cdot (1 - z_S)\right). \tag{5}$$

The b parameter controls a bias set before the trial commences to encode the stimulus as either short or long. It increases the drift rate for one response (e.g., short) over the other (e.g., long), reflecting a tendency to encode the properties of a perceptual stimulus according to the participant's bias. Formally, the way it is included in the model results in adjusting the relative activation of the short and long anchors, reflecting participants' "stimulus bias" (White & Poldrack, 2014) to respond toward one end of the scale or the other. Larger values of b increase the drift rate of the short anchor relative to the long anchor, and so increase responses at the short end of the scale. Smaller values of b, conversely, increase the drift rate of the long anchor relative to the short anchor and results in more and faster responses at the long end of the scale. The former is a bias that was displayed by our participants, as we report below.

This is only one of several possible ways to instantiate a stimulus bias. For example, it might be alternatively achieved by shifting the values of stimulus length that goes into calculating z. We believe that these different implementations of bias would be difficult to tell apart in the present data, but might be discriminated by a manipulation of the stimulus range. We leave exploration of this issue to future research.

Response Time and Response Selection

The final ingredient is a rule that terminates the race between the accumulators associated with each anchor and selects a response. There have been a number of proposals for how response boundaries can be used to map a small number of accumulators onto a greater number of responses, which we examine in Supplemental Materials. Here, we focus on the circular boundary model shown in Figure 1, which corresponds to the asymptotic limit of both the multiple threshold race (Reynolds, Garton, et al., 2021; Reynolds, Kvam, et al., 2020) and the geometric framework for modeling decision-making (Kvam, 2019a; Kvam & Turner, 2021). A single boundary allows discrete and continuous cases to use a commensurate stopping rule, as with the circular diffusion model (Smith, 2019; Smith & Corbett, 2019). However, this assumption must be abandoned when we move to stimuli that are unequally spaced, as in our second experiment, because the similarity between options is a critical determining factor in how much evidence is needed to make a decision.

With all of these pieces specified, we can put together the full model, a process that halts whenever there is sufficient evidence for any alternative at distance θ from the origin—that is, when the length of the state (i.e., the norm of the state vector s) exceeds θ: $||s|| \geq \theta$.

This results in an extremely simple stopping rule, as all of the tractability of the circular boundary and linear accumulators translates into analytic equations for stopping locations and stopping times. For a state that moves around the (x, y) plane, we simply have to compute the location s_{final} at which it hits the circular boundary in

order to find the predicted response. The accumulation time—that is, the amount of time it takes to transit from the starting point to the hitting point—is given by taking the distance $D = ||s_{\text{final}} - s_{\text{initial}}||$ (where $s_{\text{initial}} = [S_S, S_L]$ is the initial state) and dividing by the accumulation rate. This gives the decision component of a response time, with overall response time, RT, given by the following:

$$\text{RT} = \frac{D}{\sqrt{v_S^2 + v_L^2}} + \tau. \tag{6}$$

Predictions

The model we have developed follows in the tradition of models of absolute identification (Luce et al., 1982; Marley & Cook, 1984, 1986), expanding their purview and predictions while accounting for fundamental phenomena. Among the most important effects in absolute judgment are "bow" effects, where discrete responses to stimuli near the ends of a scale tend to enjoy both faster and more accurate responses than middle categories (Luce et al., 1982). Responding also typically becomes slower and less accurate as the number of responses increases, although at the ends of the stimulus range this effect is more evident in response time. These "set-size" effects are typically thought to reflect a limit on performance in terms of the amount of information transmitted (Miller, 1956). The lengthening response times and shrinking accuracy toward the center of the scale have previously been attributed to perceptual and memory processes (Guest et al., 2018) or to perceptual processes alone (Lamberts, 2000)—we test these proposals by leaving the stimulus on-screen in our experiments, to focus on perceptual and response processes. Although in typical absolute identification paradigms the number of stimuli and responses are the same, Lacouture et al. (1998) found that almost all of the set-size effect, and much of the bow effect, are due to the number of responses rather than number of stimuli when the two factors are manipulated separately.

We now examine MAAT's predictions for bow and set-size effects on accuracy and response time for both discrete and continuous responses. We also examine predictions for response deviations, which provide an inherently graded metric that tracks the difference between where a participant should and does respond. Response deviation is, therefore, a measure that is richer than simple correct/incorrect—it tells us the error direction, and in the case of multiple responses, which incorrect response is given.

In order to understand MAAT's predictions for response time it is important to understand that although the distance to the boundary is the same for every response, the overall rate of accumulation, $\sqrt{v_S^2 + v_L^2}$, is not. This is because the effective threshold for one accumulator depends on the value of the other accumulator. For example, say the long accumulator has no evidence at time $t = 1$ then the short accumulator has to have a value of θ to trigger a response at the short end of the scale at that time. In contrast, if the long accumulator has a value of $\theta/\sqrt{2}$ at $t = 1$ then the short accumulator need only have the same value at that time to trigger a response in the middle of the scale. For the short response $\delta = \theta$ whereas for the middle response $\delta = \sqrt{2}\theta$, so that the overall rate required to trigger a response at the same time is higher for responses in the middle of the scale than responses at the ends of the scale. Given rates for each accumulator are sampled independently, this means the model predicts a "bow" effect, with the average time to respond being faster at the edges of the response scale than in the middle.

In addition to an inverted U (i.e., downward) bow effect in response times, the model also predicts an upward bow effect in accuracy, where responses in the the middle of the scale are less consistently correct than responses at the ends of the scale. This occurs because the trial-to-trial variance of the drift rates is greatest when stimuli are toward the middle of the scale. The variance of a β random variable, $\beta(a, b)$, is $(a \cdot b)/((a + b)^2 \cdot (a + b + 1))$. In our case, $a = c \cdot z$ and $b = c \cdot (1 - z)$ and so $a + b = c(z + 1 - z)$. Hence, the only part of the β variance that changes with the stimulus, L, is $a \cdot b = c^2 \cdot (L - R_{\min}) \cdot (R_{\max} - L)/R^2$, which has a maximum of $(c/2)^2$ when L is in the middle of the scale. As a result, we can expect the lowest accuracy in the middle of the scale, and expect it to monotonically increase as the stimulus moves toward either the long or short end of the scale. For both accuracy and response time, the bow effects are symmetric when there is no response bias. However, bias (e.g., positive values of MAAT's b parameter cause a bias toward short responses) will cause both faster and more accurate responding for the favored end, and hence an asymmetric bow.

Like almost any reasonable model, MAAT predicts that accuracy decreases as set size (N) increases. This prediction arises because the base rate of accuracy decreases as $1/N$. However, MAAT does not predict the large effects of set size on response time that are found in absolute identification experiments. This implies that either response threshold must increase or accumulation rates decrease as set size increases. Given Lacouture et al.'s (1998) finding that response rather than stimulus set size is the main driver of set size effects, an increase in the threshold appears most likely. If this is the case, MAAT predicts that there will be no differences in response time as a function of stimulus set size in the continuous condition.

MAAT also predicts an interaction that is typically observed between set size and bow effects in absolute identification: namely, faster responding for extreme categories as set size decreases. This pattern is not captured by Lacouture & Marley's (1995) mapping model of absolute identification but is by Brown et al. (2008) SAMBA model. The corresponding interaction in accuracy—less errors for extreme response categories—is also observed, although it is often smaller than the response time interaction (Stewart et al., 2005) and may even be largely absent (Lacouture et al., 1998). Both the accuracy and response time interactions can be modulated by response bias effects, reducing set size effects at the favored end of the scale.

Response deviations cannot be examined in traditional absolute identification tasks as different responses are assigned to different buttons rather than different positions on a continuum. In MAAT, the magnitude of a response is a linear function of the accumulation angle, ϕ (Equation 3). This angle equals the inverse-tangent transformation of the ratio of the β distributed long and short accumulator rates, v_L/v_S. The transform places bounds at 0° and 90°, causing the distribution of ϕ to be positively skewed when centred around small angles, symmetric when centred around 45°, and negatively skewed when centred around larger angles, at least when the β distribution variance is moderate. In both experiments, the latter condition held, as distributions of response magnitudes had a single peak around the true value. Hence, MAAT predicts that response magnitude distributions should be skewed toward the centre of the scale for small and large stimuli, and symmetrically distributed for middle stimuli.

We used mean absolute deviations in response magnitudes as a robust way to summarize bow and set-size effects on variability[2]. MAAT predicts bow effects on mean absolute deviations in the same upward direction as for response time (i.e., greater mean deviations toward the centre of the response scale) for the same reason it predicts skew effects—because of the bounded response range with anchors at either end. It does not, however, predict set-size effects unless precision (c) increases with set size, in which case variability will increase with set size. Such set-size effects in c may be plausible under an information-theoretic view (Hick, 1952) where precision decreases because limited information must be shared among more representations.

Response deviations also allow us to examine the degree to which any set-size effects on accuracy in the continuous condition are a result of dichotomizing responses as correct versus incorrect. Van Maanen et al. (2012) suggested that reduced accuracy is due to "crowding" caused by the increased similarity that necessarily occurs when set size is increased while holding the stimulus and response ranges constant, as was the case in our experiments. If mean absolute deviations are unaffected by set size, then any effects on accuracy in the continuous condition are solely due to crowding. However, if mean absolute deviations increase with set size, then the corresponding increase in overlap between adjacent response distributions will also play a role.

Study 1: Line Length

We initially report the results of our first experiment in a descriptive manner and then evaluate the ability of MAAT to fit the data. Our emphasis is on developing a relatively simple model that describes the main features of the data rather than a more complex model that captures smaller details. For example, better fits could be obtained in the discrete condition by uneven spacing of category boundaries and that may be appropriate in some circumstance. However, assuming equal spacing here makes comparisons of discrete and continuous responding more transparent.

Method

A total of six participants from the University of Tasmania and 31 participants from the University of Newcastle took part in the experiment. These participants were paid AU$10 for participating, plus an additional $5 if they accumulated sufficient points in the experiment. The points bonus was designed to be easy to achieve if the participants were putting serious effort into the experiment (i.e., not simply guessing), and so we included all participants who achieved this payoff in the analyses. A total of six participants were dropped from analyses for failing to achieve this criterion, resulting in 31 remaining participants whose data were analyzed. We describe below how they earned points on the task.

Each participant completed 120 trials of training and 504 trials of the main task, consisting of 12 blocks of 10 practice trials and 42 experimental trials (i.e., 52 total trials per block). These 12 blocks were divided evenly among each of the six conditions 2 (continuous vs. discrete) × 3 (3/6/9 stimuli) so that each participant saw each condition exactly twice. In the discrete-response conditions the number of responses was the same as the number of stimuli. In both conditions, stimuli were evenly spaced along the response scale (e.g., at 125, 275 and 425 pixels for 3 stimulus types). The range of the stimuli was 50–500 pixel, which was held constant across all conditions of the study. These blocks were randomly shuffled for each participant. This level of practice (results for which were not further analyzed) was employed to enable participants to adjust to the substantially different response requirements among conditions.

This study was not preregistered. Study materials including data, analysis code, and additional figures can be found on the Open Science Framework at https://osf.io/6d29q (Kvam & Heathcote, 2022).

Task

A diagram of the task participants were asked to perform is shown in Figure 3. The length of the stimulus was given by two horizontally aligned dots to avoid differences in overall screen brightness from full line segments that vary in length, as this would result in two-dimensional stimuli that vary in both length and brightness. Across trials, each stimulus category could result in one of three stimuli: a stimulus that was centered on the screen (so that its left and right points were equidistant from the center), a stimulus that was shifted slightly to the right, or a stimulus that was shifted slightly to the left. This was done so that participants could not rely on just one of the two points comprising the stimulus—if it were always centered, then participants could judge the distance of a single point from the center rather than judging the distance *between* the two points.

In the practice trials, participants could mouse over different response locations—numbers in the discrete condition or anywhere along the scale in the continuous condition—in order to match their response to the stimulus on the screen. This is shown in the top panels of Figure 3: the blue dots represent the stimulus and the orange dots represent the response corresponding to the location of the mouse. Participants would confirm their selection in practice trials by clicking the mouse on the chosen response location. The 10 practice trials preceding every full block gave them the opportunity to understand the scale they were about to use—whether that was continuous, or discrete with 3, 6, or 9 options.

In the experimental trials, the participant's task was to match the length of the stimulus with the location on the scale that they had learned was associated with the given stimulus. In the discrete condition, this was done by assigning it to one of the numbered categories (bottom left of Figure 3), similar to traditional absolute judgment tasks with the exception of the responses being locations on the screen rather than physical buttons. In the continuous condition, it was accomplished by assigning the stimulus to a position on the scale that corresponded linearly to the length of the stimulus (bottom right panel of Figure 3).

Participants received 10 points per trial for a correct response, made by moving their mouse across the semicircular response scale within 1° of the middle of the appropriate category (marked by a numeral) in the discrete condition, or within 1° of the location corresponding to the stimulus length in the continuous condition. In all conditions, they lost a point for every degree away from the correct location, meaning that they would receive points as long as

[2] We use mean absolute deviation as opposed to variance to index variability while providing robustness to outlying deviations that occur when the wrong response is occasionally chosen in the discrete condition, which can drastically inflate variance because the distance between responses is squared.

they deviated by less than 10°. This range was chosen to match the width of the categories in the most difficult condition (9 categories), where each category corresponded to a 20° arc on the scale. This ensured that the motor challenge of the task was constant across conditions, although it was still somewhat easier in the discrete condition because there was a number at the middle of the category that participants could use as a target.

Response times were recorded as soon as the mouse cursor moved outside the semi-circle (10 pixels away from the central fixation point), so that the time it took to reach the edge of the scale was not counted in the response times. This was done to minimize the impact of nonlinear trajectories that often accompany responses on the radial scale we used (see, e.g., the trajectories in Kvam & Busemeyer, 2020). While these trajectories can certainly be informative for understanding the evolution of the decision state over time (see Dotan et al., 2018; Friedman et al., 2013; Koop & Johnson, 2011; Lepora & Pezzulo, 2015), we focus here on accounting for the final responses themselves and the associated response times. Forcing participants to make ballistic movements, and cutting out the time it took them to reach the scale, provided better control over response times and reduced the covariance between response location and response time. Responses were recorded as the angle of the cursor relative to the center when it crossed the response scale. To encourage participants to move the mouse directly from the center to the location of their desired response, they received an error message anytime their cursor spent more than 300 ms between the starting circle and the semicircular boundary, and did not receive any points for these trials. Any trials on which this error message was triggered, as well as any on which participants responded too quickly (<0.25 s) or too slowly (>5 s), were removed prior to data analysis (4.01% of responses).

Descriptive Results

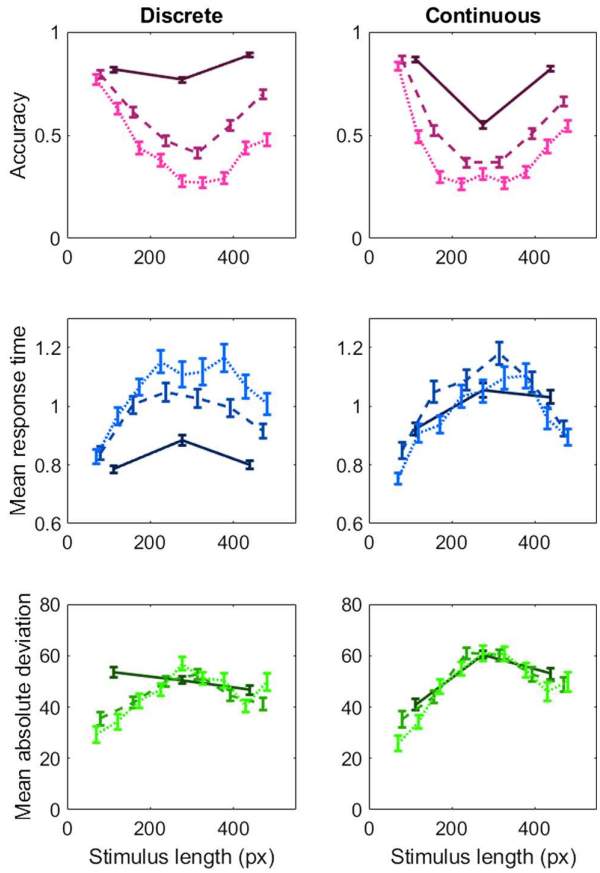
For both discrete and continuous responses, we scored accuracy in terms of the proportion of responses that fell within equal-width adjoining ranges around the stimulus values (e.g., 50–200, 200–350, and 350–500 pixels for three stimuli). We also analyzed response times and response deviations in both discrete and continuous conditions. Response times are the time to move the mouse from the center to the edge of the semi-circle. Response deviations are the raw difference between the response a participant gave and the response they were *supposed* to give, whether that was the center of a stimulus category (discrete condition) or the actual position of the stimulus length on the response scale (continuous condition).

Note that we report all results using Bayesian statistics, including the mean effect estimates (M) along with 95% highest-density intervals [HDIs] describing the interval containing the 95% most likely values of each estimated parameter.

As expected, there were bow effects for discrete responses, which were more accurate and faster toward the edges of the scale, and the same was true for the continuous conditions, as shown on the left and right of Figure 4. The discrete condition also produced the expected set-size effects, with overall accuracy and speed being less in conditions with more responses. However, for set-size 6 and 9, there is also a strong bias favoring the short end of the scale and and associated asymmetry in the bows. As a result, there was virtually no set-size effect in either speed or accuracy for the shortest response and an exaggerated set-size effect in both measures for longer

Figure 4

Mean Accuracy (Top Row, Proportion Correct), Response Times (Middle Row, Seconds), and Response Deviations (Bottom Row, Pixels) Across Conditions of the Task



Note. Error bars indicate ±1 unit of standard error with lines passing through the means. The *x*-axis indicates the length of the lines, divided into 3 (darkest/solid lines), 6 (medium/dashed lines), and 9 (lightest/dotted lines) line length conditions. See the online article for the color version of this figure.

responses. For the set-size 3 condition, in contrast, the response-time bow is quite symmetric and the accuracy bow, if anything, favors longer responses.

There was a similar pattern in accuracy for the continuous condition, except that it was a little higher for the shortest condition, the bows slightly deeper, and there was a small bias favoring shorter responses for the smallest set size. Participants were slightly less accurate in the continuous than the discrete condition, particularly in the middle stimulus categories, which is likely due to the presence of number labels on the scale in the discrete condition that were not present along the continuous scale, as shown in Figure 3. Mean response time in the continuous condition exhibited a bowed pattern that was biased toward short responses, with the asymmetry being present even for the smallest set size. However, in contrast to the discrete condition, there was no overall set-size effect on response time.

Asymmetric and short-biased bow effects also appeared in mean absolute response deviations, with the exception of the set-size 3 discrete condition where deviations were unusually high. Importantly, there was no set-size effect in the continuous condition for

either discrete or continuous responses, as illustrated by the overlapping lines in the bottom panels of Figure 4. This does not support the idea that the ability to represent stimuli precisely is subject to capacity constraints, and hence that a model in which precision (*c*) does not change with set size is likely to fit this data well. It also suggests that the accuracy effects in the continuous condition are due to crowding. In combination with the lack of set-size effects on response time, this suggests that the same response threshold was used for all continuous conditions. In contrast, the response time effects in the discrete condition suggest that participants used larger response thresholds for larger set sizes.

Modeling Results

The model had nine parameters per participant (see Table 1). In line with our aim to produce a simple unified model and the empirical results just reviewed, only threshold parameters differed between discrete and continuous conditions. Only one value of nondecision time, τ, was estimated on the assumption that stimulus encoding and motor production was the same for all conditions. Similarly, the same value of start-point variability, *s*, was estimated for short and long accumulators, so this parameter played no role in explaining response bias. In light of the lack of effect of set size on mean absolute deviations in the continuous condition, capacity, *c*, was assumed to be unaffected by set size. Two further parameters completed the specification of drift rates, δ, which controls their overall magnitude, and *b*, which controls stimulus bias. The remaining four parameters were all thresholds, one for each set size in the discrete condition (θ_3, θ_6, and θ_9) and a single threshold for all of the continuous conditions.

A participant had 84 trials in each condition, distributed uniformly across stimulus lengths. However, this does not result in a large number of trials in a category, especially when there were nine response categories and participants tended not to respond as often in Categories 3, 4, 6, and 7 as in the center and edge categories. To address associated issues with measurement error, we used hierarchical Bayesian estimation to fit the model, allowing the number of participants to compensate in part for fewer responses in each category within each participant by sharing parameter-relevant information across participants. Each parameter, with the exception of bias, was restricted to the positive reals, and we use relatively uninformative and independent hyperpriors on the group-level parameter estimates. All parameters were constrained by a group-level distribution. Specifically, for the drift, capacity, threshold, and nondecision time parameters, the group-level prior was a very wide γ distribution, $\gamma(.001, .001)$ (in a shape, rate parameterization). The start point variability parameter was set as a proportion of the θ_3 parameter (typically the lowest of the four thresholds) to avoid start points going above the threshold, using a group-level truncated normal distribution with hyperpriors μUniform(0, .99) and σGamma(.001, .001) (again in shape, rate parameterization).

We fit the model to accuracy and response time data in both discrete and continuous conditions. Response deviation data was used in a cross-validation exercise to test whether the model makes good predictions for the continuous conditions. Model fit was quantified using a multinomial likelihood for responses (number of responses in each category) and the probability density of the associated response times.

Hierarchical Bayesian estimation was carried out in Just Another Gibbs Sampler (JAGS; Plummer, 2003) with a MATLAB interface,

Table 1
Parameter Estimates From Study 1

S	*c*	δ	θ_0	θ_C	K_D	τ	θ_N	K_C
1	1.32 [1.18, 1.50]	1.61 [1.01, 2.61]	0.20 [0.00, 0.73]	1.06 [0.37, 1.59]	0.15 [0.08, 0.23]	0.53 [0.46, 0.60]	0.25 [0.00, 0.75]	0.06 [0.00, 0.16]
2	0.70 [0.52, 0.88]	5.42 [4.47, 6.46]	0.68 [0.00, 1.29]	5.92 [4.99, 6.89]	0.09 [0.00, 0.17]	0.19 [0.06, 0.27]	0.54 [0.00, 1.31]	1.96 [1.79, 2.18]
3	0.95 [0.68, 1.18]	5.07 [3.92, 6.41]	0.82 [0.00, 1.68]	9.74 [8.71, 10.91]	0.88 [0.80, 1.00]	0.17 [0.07, 0.28]	1.89 [0.78, 2.74]	3.28 [3.05, 3.50]
4	1.09 [0.93, 1.25]	2.89 [1.99, 3.67]	0.56 [0.00, 1.15]	12.44 [11.66, 13.28]	0.97 [0.91, 1.00]	0.63 [0.56, 0.72]	0.87 [0.06, 1.53]	4.86 [4.67, 5.03]
5	1.50 [1.32, 1.67]	6.68 [5.66, 7.42]	0.42 [0.00, 1.11]	23.66 [22.68, 24.89]	0.93 [0.85, 1.00]	0.24 [0.15, 0.34]	1.72 [0.51, 2.66]	10.46 [10.28, 10.59]
6	0.82 [0.59, 0.98]	3.22 [2.43, 4.43]	0.49 [0.00, 1.39]	4.33 [3.48, 5.35]	0.45 [0.34, 0.55]	0.17 [0.09, 0.30]	0.73 [0.00, 1.54]	1.21 [0.98, 1.39]

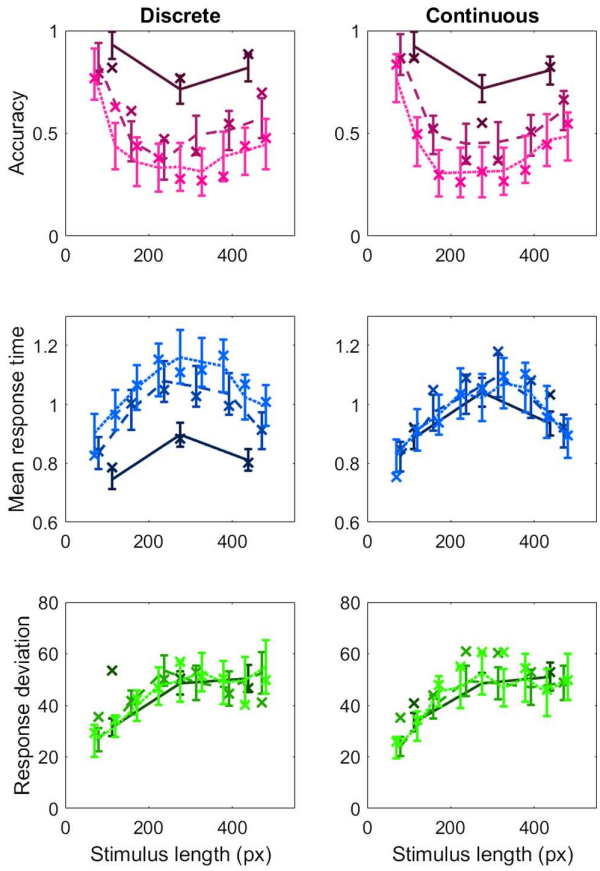
Note. Maximum a posteriori estimates (95% HDI) of the parameters for precision coefficient (c), accumulation rate scalar (δ), base threshold for discrete conditions (θ_0), threshold for continuous condition (θ_C), normalization factor controlling threshold adjustment for similarity in the discrete condition (K_D), nondecision time (τ), threshold adjustment for the number of options in the discrete condition (θ_N), and the theshold variability in the continuous condition (K_C) for each subject (S).

matjags (Steyvers, 2011). Each of the parameters shown in Table 1 was estimated for each individual simultaneously, constraining the individual-level estimates with a group-level prior as described above. In total, this resulted in estimates for nine parameters × 31 participants, plus 18 (9 group-level distributions × 2 hyperpriors) group-level parameters to fit $N = 15{,}575$ responses and associated response times. The JAGS code for fitting this model (and a version that can be used to fit individual participants) is on the Open Science Framework at https://osf.io/6d29q.

The model fit the general patterns in the average accuracy and response time data, as shown by the lines in Figure 5. It was able to accommodate the bow effects in accuracy and response times even through only thresholds differed as a function of number of responses. There was clear overestimation of accuracy in the second category of the set-size 3 continuous condition. Our best explanation for this result, and the reason that accuracy differed so much in this case between discrete and continuous conditions, is that there was no number in the continuous condition that allowed participants to match exactly the location of the middle of the scale. This may have resulted in poor accuracy in the continuous condition, where participants appear not have known that there were only three stimulus categories that could appear and thus made responses further from the middle of the scale. Because there was only one threshold for the continuous condition, the model predicted entirely overlapping mean response times for these conditions (middle right panel), a prediction that was largely reflected in the data as shown in Figure 5.

Figure 5

Data (Xs) and Model Predictions (Lines, Colored and Dashed as in Figure 4) for the Mean Response Times, Accuracy, and Mean Response Deviations Across the Conditions of the Experiment



Note. Bars indicate the 95% highest density intervals for the predicted mean based on simulations generated from the model, where the model produced one predicted response for every response in the data. See the online article for the color version of this figure.

The slowing captured by the model in response time with set size in the discrete condition (middle left panel) is due entirely to the difference in thresholds ($\theta_{3/6/9}$). For accuracy and response times in both discrete and continuous conditions, there is a clear asymmetry in the bow effects, where accuracy is higher and response times are faster at the short end of the scale. The asymmetry is well described by the addition of the single b parameter to the model, but it is not clear whether this will be a necessary ingredient to account for behavior in other mapping tasks. One possibility is that it is related to using a mouse to respond, as lateral biases in responses can sometimes be eliminated when a joystick is used (see Busemeyer, Kvam, et al., 2019).

A finer grained test of thresholds mediating set-size effects is provided by considering the fit to the entire distribution of response times, as thresholds have different effects on distribution shape than the other parameters such as drift rates or nondecision time (Ratcliff & Smith, 2004). Figure 6 shows the fits for the discrete condition (fits of the continuous condition are provided in Supplemental Materials, and are of similar quality). Consistent with a threshold effect, set-size conditions differ mainly in the leading edge and variability. Consistent with differences in drift rate, both variability and skew were increased toward the middle of the response scale.

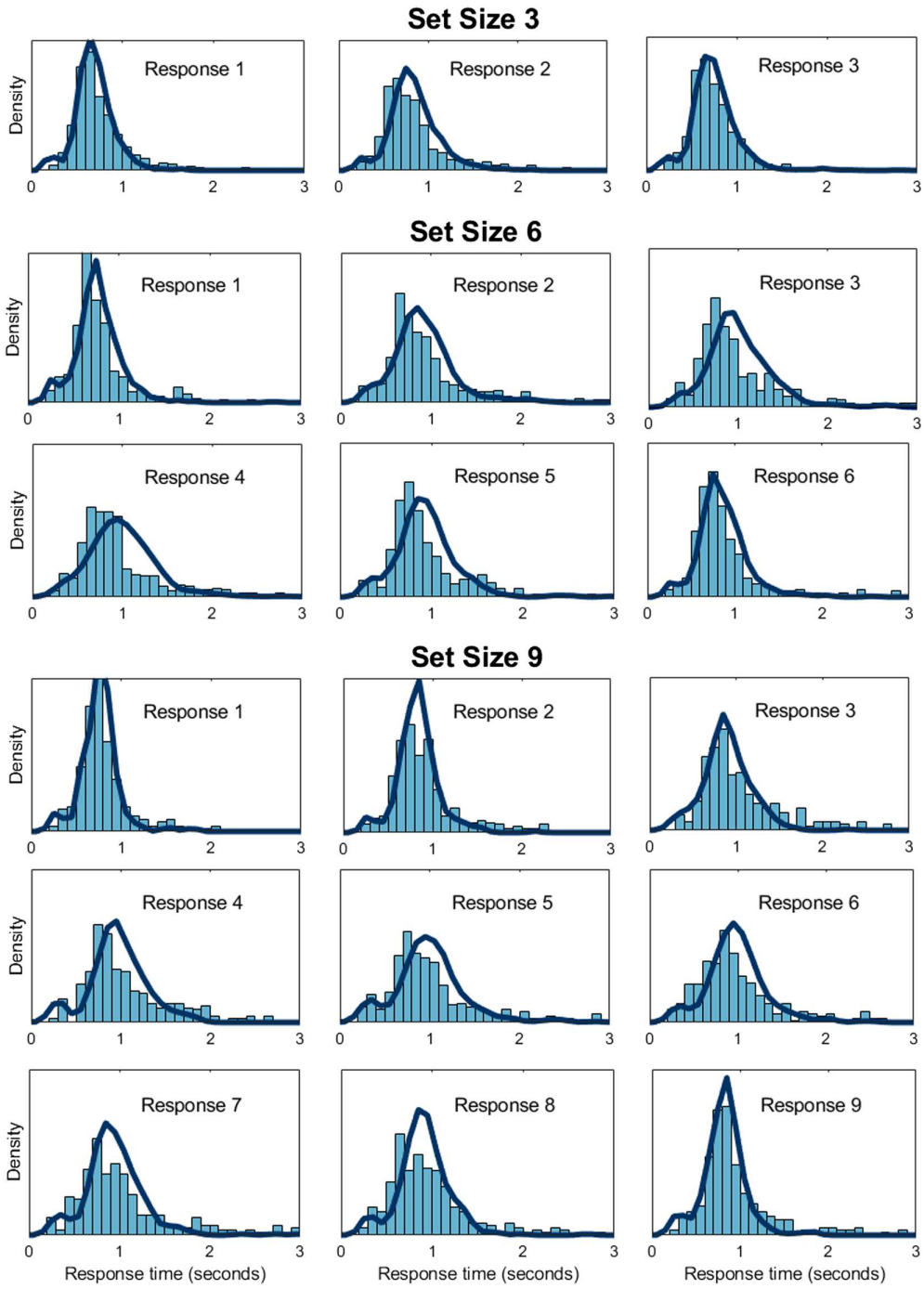
Table 1 shows the group-level mean parameter estimates, illustrating the central tendency of each one across individuals. Note that the δ, c, and b parameters combine to determine the drift rates for the short (v_L) and long (v_L) accumulators. The 95% credible intervals that accompany each parameter show that they were precisely estimated, consistent with the model's simple structure. The model is also about as simple as it can be without serious misfit: when we tried to simplify the model further by removing start-point variability, accuracy in the set-size 3 condition was drastically over estimated. The discrete-condition threshold parameters showed an increase in magnitude with set size and consistent with empirical results the difference between set sizes 3 and 6 was greater than that between 6 and 9. The overall increase suggests that participants responded more cautiously as set size increased in order to ameliorate the associated decrease in baseline accuracy. In agreement with this idea, the threshold for the continuous condition was very similar to that for the middle set size in the discrete condition. Consistent with the greater motor demands of using a mouse and thus greater motor preparation time (Fitts, 1954), the nondecision time estimate was greater than typically found in paradigms with a button-press response.

Out of Sample Predictions

The model was fit to response time and accuracy (i.e., whether each response was within 10 pixels of the correct value) data. However, in the continuous condition, a response is not simply correct or incorrect, it also has a magnitude, distributions which are shown in Figure 7 as histograms. Although the model was not fit to

Figure 6

Observed (Histogram) and Predicted (Lines) Distributions of Response Times Across Different Discrete Set Size Conditions of the Experiment

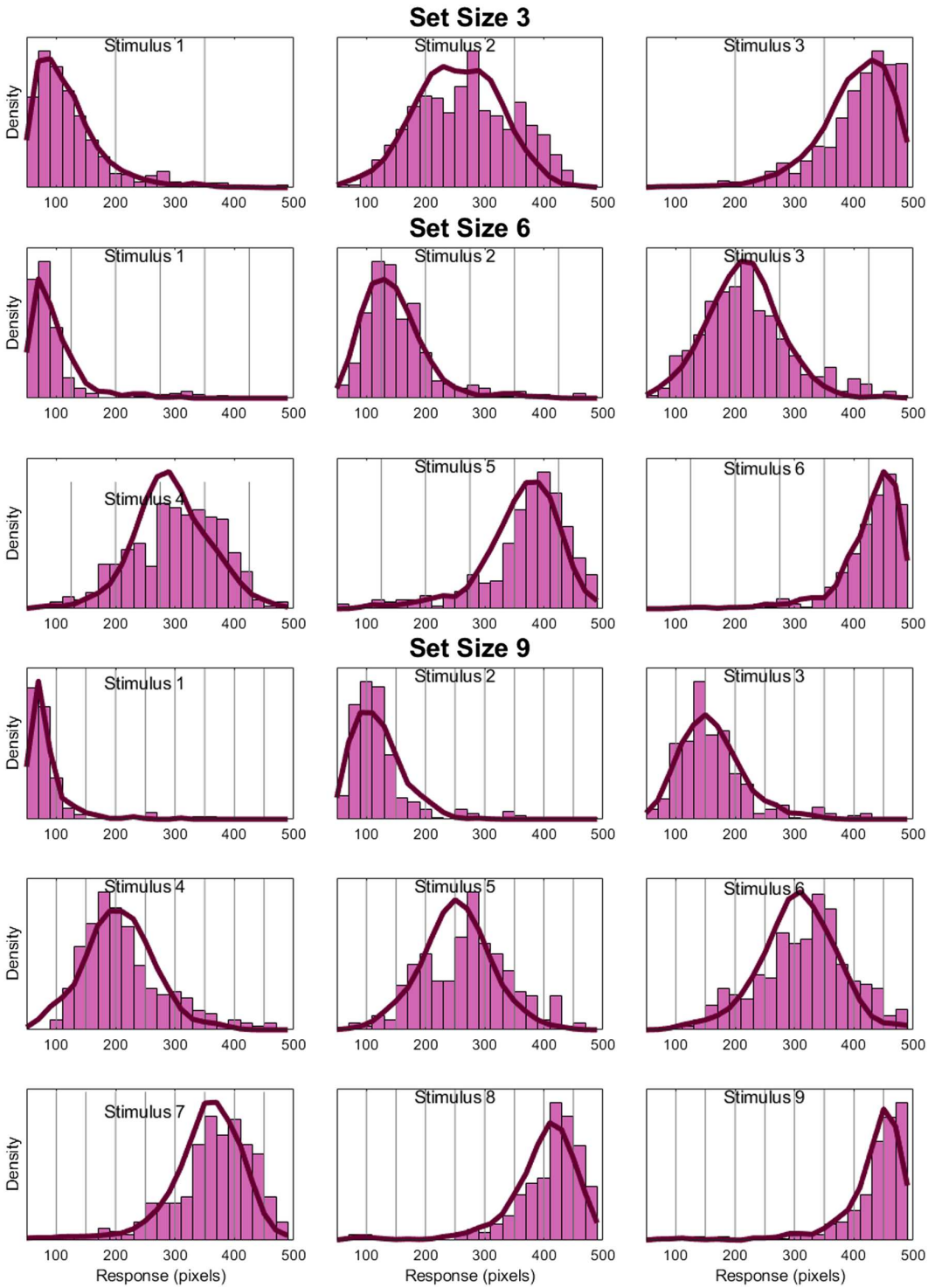


Note. See the online article for the color version of this figure.

this data, we performed a type of cross-validation test to determine how well it could predict it. To do so, we generated a single-simulated trial for each trial completed by each individual based on the model's maximum a posteriori parameters, estimated by passing a kernel density estimator over the individual-level parameter samples and linearly interpolating the maximum height for each participant. This allowed us to equitably represent the predicted data and the relative influence of each participant (and their corresponding model parameters) in the simulated data, with the resulting predicted distributions shown as densities in Figure 7.

Figure 7

Distribution of Responses Magnitudes in the Continuous Condition (Histograms) and Model Prediction Based on Fits to Accuracy and Response Time Data (Lines)



Note. Light gray vertical lines mark category boundaries. The histograms show aggregate data (collapsed across participants) from each condition, while the model predictions are a weighted average of the probability densities generated from simulated data for each participant. To generate the model predictions, we created an artificial data set matching the exact composition of the real sample. For example, if there were 50 trials from Participant 1 in the real data, we simulated 50 trials from Participant 1's maximum a posteriori parameters. Once simulated data was generated for every participant, we passed a kernel density estimator over it to approximate the density of responses (lines). See the online article for the color version of this figure.

Figure 7 shows that the observed distributions of response magnitudes had the pattern of skewness predicted by MAAT: for each stimulus below (resp., above) the middle category, the response distribution exhibits exhibits right (resp. left) skew. Furthermore, the skew is more extreme as the stimulus gets closer to the edges of the response scale. This skew is well accounted for by the model, which produced even the strong skew that was characteristic of stimuli in the minimum (1) and maximum (3/6/9) categories. It can do so because of the β distributions used to characterize the activation of the anchors. A strong activation for one anchor produces a β-distributed drift rate that is concentrated very close to 0 or 1, which is then scaled by the drift parameter δ. As a result, there are mainly large values, with occasional low values, for the drift of the strongly favored anchor and so the drift rate distribution itself exhibits skew when stimuli are close to either end of the scale.

Discussion

The key take-away from the modeling of the line-length experiment is that the same fundamental mechanisms can support both discrete and continuous responding. We were able to produce all of the most important phenomena in accuracy, response time, and response magnitude distributions (deviations) in both cases by varying only the response thresholds. In the discrete condition, slowing with increased set size was accounted for by increasing thresholds. Thresholds are typically thought of as being set in a strategic but slow manner, which is consistent with set-size differences in the discrete condition being evident to participants from the display (see Figure 3). The 10+ practice trials with each new display would easily allow participants to make the threshold adjustment. It appears that participants set higher thresholds in an attempt to ameliorate the decreased accuracy associated with larger set sizes. A separate threshold accounted for the continuous condition, which again seems reasonable given that it also had a distinctive display with corresponding practice trials. Although it is possible that participants might notice that different stimuli were used in different blocks of the continuous condition and adjust their thresholds, this does not seem to have been the case as quite good fits were obtained with the same threshold for all stimulus set sizes.

Our modeling results also indicated that the decrease in accuracy with increasing set size was due to increasing similarity between adjacent response options (Van Maanen et al., 2012). Increasing the number of response options while keeping the same span of stimuli constant as we did in the present experiment naturally means that adjacent stimuli are closer, and hence more similar, to each other. Therefore, the associated decrease in accuracy, which at least in the continuous conditions was not modulated by threshold differences according to our model, can be attributed to a "crowding" effect. However, set size might also fundamentally alter how the response alternatives are represented, as suggested by information theoretic accounts of multi-alternative choice (Hick, 1952). MAAT could allow for this possibility by letting the precision coefficient, c, which controls the precision with which response alternatives are represented, vary with set size. However, this was not necessary to obtain good fits, and is inconsistent with empirical findings about the precision of response magnitudes, which did not differ with set size. Rather, there seems to be a single representation of a stimulus that is common across different response conditions. However, the present design was limited in that the number of response options and similarity are confounded. In the next study, we varied similarity of response options within each set size.

Hue Identification

Responses on the hue wheel are a prevalent methodology in visual working memory tasks (Zhang & Luck, 2009, 2011) and have driven the development of continuous-response models (Ratcliff, 2018; Smith, 2016; Smith et al., 2020). It is particularly notable in these tasks that participants' responses tend group near "cardinal" hues—red, green, yellow, blue, magenta, and cyan—even when the stimuli are uniformly distributed across the hue wheel. As such, our model ought to be able to account for the concentration of responses around these locations in the continuous response task while simultaneously describing accuracy and response times in both discrete and continuous conditions.

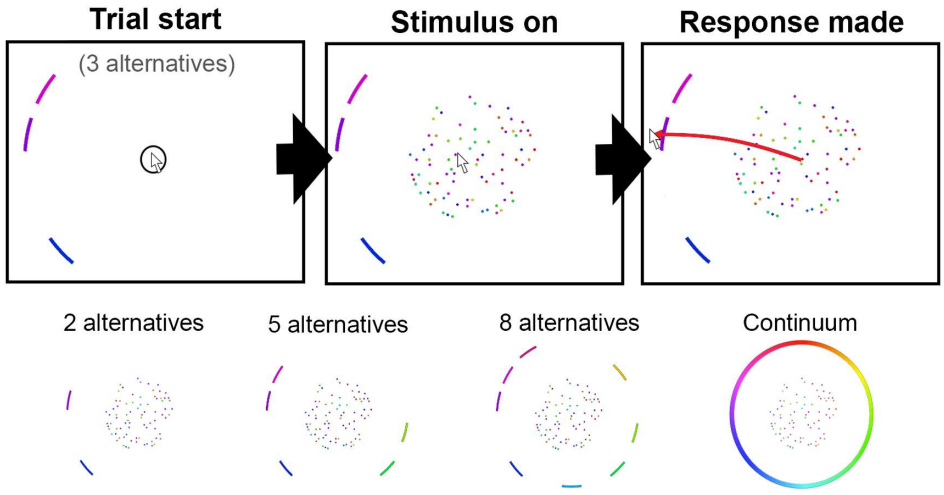
We obtained the grouping effect empirically, and used it to deepen our theory of the connection between discrete and continuous responses, using a hue identification task. In this task, an array of differently colored dots was shown on the screen and participants were asked to assess which color is most common. Saturation was set to 1.0 and the value was set to 0.8 in HSV color space, so this discrimination was based only on hue. In the discrete-response condition, a fixed set of response options was available in each block of trials, and so participants had to determine which of these possibilities was most consistent with the stimulus. In the continuous condition, they had to make a response on the hue wheel.

Figure 8 illustrates the identification displays for continuous and discrete conditions, which were run within subjects. Each participant worked with a wide variety of hue combinations in order to provide a rich and highly constraining data set. In order to model each person's behavior individually, there were few participants who all had high-quality data (i.e., each performed a large number of trials Smith & Little, 2018). As in Experiment 1, before blocks of trials in each condition, participants were given ample practice in order to adjust to each new response configuration. Details of stimulus and display construction are given in the Methods section below. Here, we provide an overview in order to set the stage our modeling choices, which are described in the next subsection.

The entire hue wheel was used for responding in the continuous condition. The stimulus to be judged consisted of a cloud of dots centred on the middle of the hue wheel. Each display had dots of 16 different hues. Half of the 16 were the same on all trials within a block, and half differed from trial to trial. On each trial, a dominant hue was chosen from the set of eight constant hues. The dominant hue occurred more often than the remaining hues, which occurred equally often. The participant's task was to identify the dominant hue. Each block in the discrete condition had either $N = 2$, 3, 5, or 8 responses displayed as disconnected arcs. Stimuli were constructed in the same way as for the continuous condition, except the dominant hue was randomly chosen from the response hues. Response hues were randomly chosen from the constant set of eight at the start of the block and were the same for all trials within a block.

For smaller set sizes, Hick's Law often provides a good account of the increase in mean response time with set size, at least when set size is not confounded with similarity. For example, van Ravenzwaaij et al. (2019) modeled data from Van Maanen et al.'s (2012) experiment where participants identified 3, 5, 7, or 9 movement directions that

Figure 8
The Decision Task in Study 2



Note. Depending on the condition, these were comprised of 2, 3, 5, 8, or a continuous span of alternatives. Participants responded by moving their mouse across the arc corresponding to their desired response. See the online article for the color version of this figure.

were equally spaced (i.e., the range of directions increased with set size). The data followed Hick's Law, a pattern that was fit by van Ravenzwaaij et al.'s (2019) ALBA (advantage linear ballistic Accumulator) model even when the same threshold was used for each response and set size. However, differences in thresholds with set size were required to provide a fine-grained account of set-size effects on accuracy. Usher et al.'s (2002) LCA (leaky competitive accumulator) model also predicts Hicks Law in the broad, but also best fits the fine-grained pattern of data when thresholds are allowed to vary with set size.

In our design, the average similarity among responses increases with set size because the range of possible response hues remains fixed. If responding slows with similarity, as is commonly observed with other manipulations that increase difficulty, response time should increase more quickly than predicted by Hick's Law. This deviation from Hick's Law might be accommodated in models such as the ALBA and LCA by an appropriate adjustment of thresholds with similarity. However, it is more difficult to see how these racing accumulator models would deal with the continuous case if they continue to follow Hick's Law as the number of accumulators is increased because in the limit of large set sizes this predicts response time will increase without bound, which is clearly unreasonable.

On the other hand, contemporary models built for continuous responses often assume a single threshold that is the same for all responses (Ratcliff, 2018; Smith, 2016; Smith et al., 2020). Such single threshold models may have trouble dealing with the unequally spaced responses, and hence variations in the similarity among responses, that occur in the discrete condition of our design. For example, in our set-size 3 condition, suppose there are two similar response options (e.g., pink and orange) and one dissimilar response option (e.g., cyan, see lower panel of Figure 9). A cyan stimulus is distinct from the other response alternatives, suggesting that participants may exercise less caution (i.e., use a lower threshold) for responding with the cyan option and enjoy greater accuracy when a cyan stimulus is presented. Conversely, they would experience greater difficulty when the stimulus is pink or orange and use a higher threshold in an attempt to compensate.

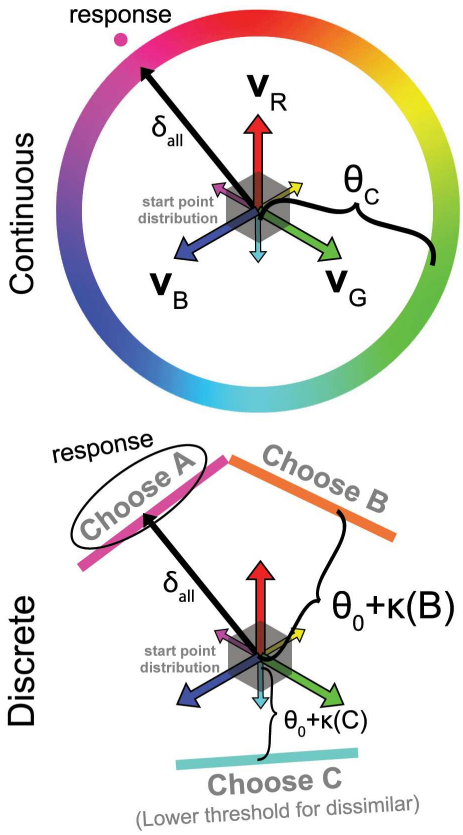
In other paradigms where difficulty varies among responses, participants sometimes adjust their decision thresholds, trading speed for accuracy in an attempt to compensate for the differences in difficulty. In our unified MAAT account of continuous and discrete tasks, we propose that thresholds are adjusted in two ways to provide a detailed account of set size and similarity effects in the hue task. In the discrete task, we propose a set of separate thresholds for each option, where the average threshold changes with set size and individual thresholds change based on their similarity to other options.

In the continuous condition, in contrast, thresholds do not change systematically with these factors, but can vary across trials, representing fluctuations in control, attention, perceived difficulty, or as a response to recent errors (Frank et al., 2005; Logan et al., 2014) We describe the mechanism by which this threshold change occurs, as well as the other details of the model, in the next section.

Hue Model

The color task model uses a trichromatic representation where responses are based on the amounts of red, green, and blue in the stimulus. This method of representing stimuli is based on the three (red, green, and blue) cone receptors in human color vision (Boynton, 1979; Schnapf et al., 1987). These three colors serve as the anchors in the MAAT model of Study 2, reflecting the idea that participants should have well-established exemplars for the colors red, green, and blue. The relative hue values, specified as R = red, G = green, and B = blue with each quantified on a 0 (*hue not present*) to 1 (*hue at maximum*) scale, drive three evidence

Figure 9
Diagram of the Model Used for the Color Task



Note. Stimuli are quantified in terms of drift rates for red (v_R), green (v_G), and blue (v_B) anchors. These drift rate components are then combined into an overall drift rate vector δ_{all} that characterizes the evidence accumulation process. See the online article for the color version of this figure.

accumulation process that anchor responses around these three hues (the large arrows in Figure 9).

As in the line length model where the long and short accumulators started with some activation, we assume that each of these new anchors (*R*, *G*, *B*) starts with a random degree of activation. The activation of each anchor is drawn as an independent uniform random variable on [0, 1]. The maximum value of this uniform distribution is fixed at 1 to set the scale of the model as in typical evidence accumulation models (Brown & Heathcote, 2008; Busemeyer, Gluth, et al., 2019; Ratcliff et al., 2016). Without such a scaling constraint, drift, threshold, and start points cannot be separately identified (i.e., different values can the same distribution of responses and response times, see Donkin, Brown, & Heathcote, 2009). This variability is shown as the shaded region in Figure 9.

Drift Rates

The rates for the three red, green, and blue anchors are specified using the corresponding relative hue (i.e., *R*, *G*, and *B*) values. Each of the drift rates is set according to a normal distribution:[3]

$$\begin{aligned} v_R &= \delta \cdot \text{Normal}\left(R, R\left(1-R\right)/c\right) \\ v_G &= \delta \cdot \text{Normal}\left(G, G\left(1-G\right)/c\right) \\ v_B &= \delta \cdot \text{Normal}\left(B, B\left(1-B\right)/c\right). \end{aligned} \tag{7}$$

The drift rate variance is set according to the uncertainty of a binomial random variable on [0, 1], which mimics the way in which variance changes in a β distribution (see Predictions section) by having the greatest variance when *R*, *G*, and *B* is close to 0.5 and the lowest variance when *R*, *G*, or *B* is close to 0 or 1. As with the line length model, drift rates are scaled based on the overall information sampling rate δ. The free parameter *c* corresponds to the precision coefficient for the system, in this case scaling the variance in *R*, *G*, and *B* accumulators as shown in the Normal distributions in Equation 7. The hue match values are analogous to short and long match for line length, being on the unit interval. They are also constrained to sum to 1 in this case because we used a hue wheel. In a hue wheel saturation and value in hue-saturation-value color space are constant, so a balance between colours is maintained with a constant sum of $R + G + B$ (see Figure 9). As a result, the amount of red in the options, for example, was a direct inverse of the amount of green plus blue. This is a simplified account of color perception; in the future, more nuanced theories of color vision could be introduced (including but not limited to tetrachromacy in humans and other animals—see Jameson et al., 2020).

The three drift rate scalars are combined to form the overall drift vector by multiplying each one by the direction in which the corresponding color is located. Red responses are in direction $\boldsymbol{d}_R = [0, 1]$ (i.e., at 12 o'clock). Green responses are 120° clockwise from red (i.e., at 4 o'clock) in direction $\boldsymbol{d}_G = \left[\frac{\sqrt{3}}{2}, -\frac{1}{2}\right]$, and blue responses are 120° counter-clockwise from red (i.e, at 8 o'clock) in direction $\boldsymbol{d}_B = \left[\frac{-\sqrt{3}}{2}, -\frac{1}{2}\right]$. The overall drift vector δ_{all}, where v_i, $i = R, G, B$, is given by Equation 7, is:

$$\delta_{all} = v_R \boldsymbol{d}_R + v_G \boldsymbol{d}_G + v_B \boldsymbol{d}_B. \tag{8}$$

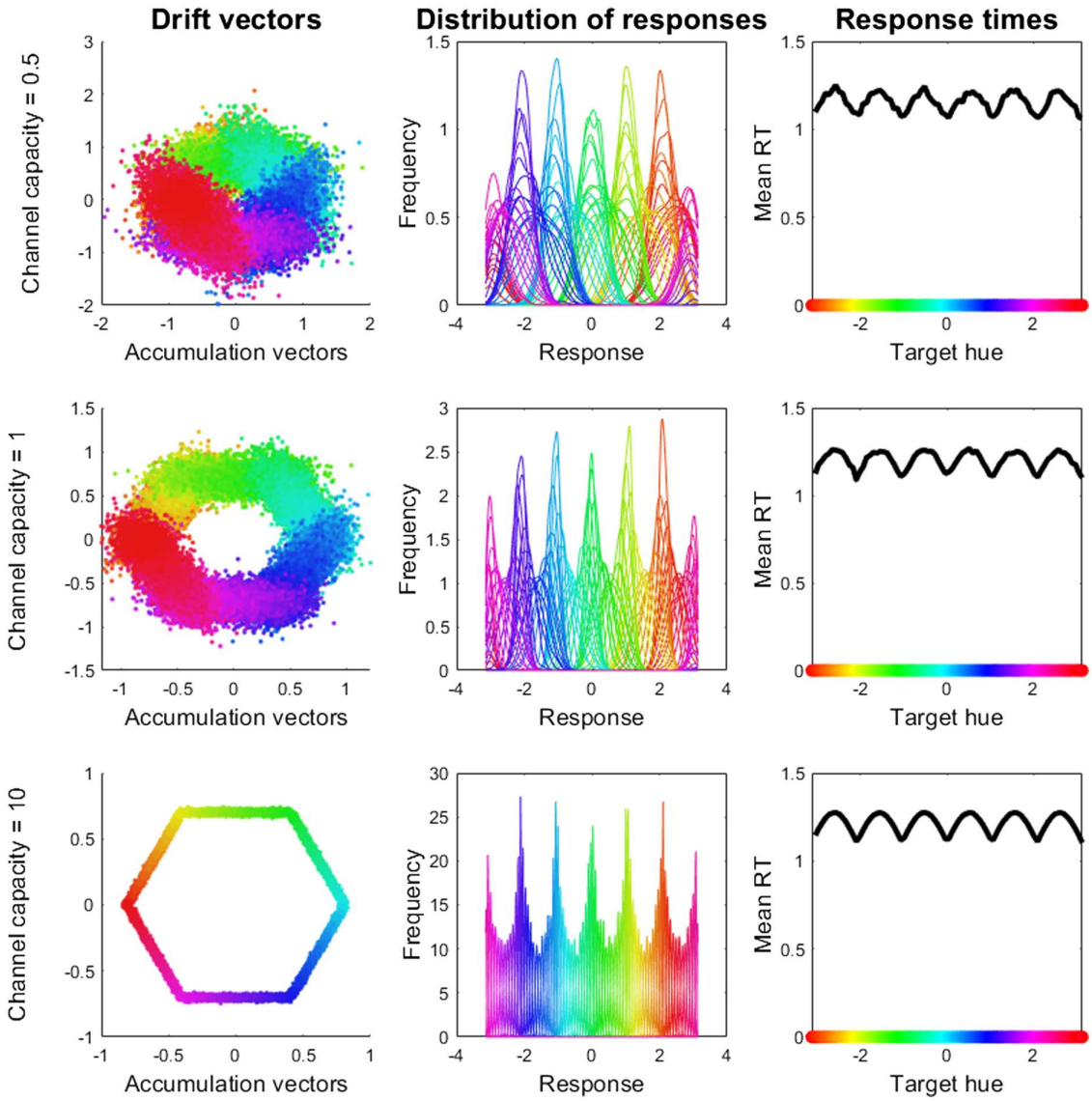
The combined drift rate δ_{all} describes the overall effect of the stimulus as a two-dimensional vector. The direction in which it points indicates the response that the stimulus favors, while its magnitude represents the rate of accumulation toward that response. Figure 10 illustrates drift vectors for different hue stimuli and capacities.

Drift rates are skewed toward the red, green, and blue, just as responses were skewed toward the upper and lower anchors in Study 1. This produces peaks in responding around red, green and blue. Secondary peaks also occur around cyan, magenta, and yellow (CMY, as shown in Figure 10) because these are local maxima of the drift vectors relative to the circular threshold: areas where the value

[3] We initially used a Beta distribution for these rates, but ran into problems with model recovery/chain convergence. This appeared to be due to high variance in the Beta distributions small values of *c* can result in drift distributions that are almost entirely concentrated at 0 or δ. These occurred in this study due to *R*, *G*, or *B* each being equal to zero for 1/3 of the stimuli, with at least one equal to zero for every stimulus, causing *c* to be hard to estimate in these cases. This was not an issue in Study 1 because Z_L and Z_R were never equal to zero.

Figure 10

Simulated Model Predictions for Drift Vectors (Left) Distributions of Evidence for Different Hues (Middle), and the Resulting Response Distributions (Right) for Three Different Levels of Precision c *(Increasing From Top to Middle to Bottom)*



Note. Note that responses are skewed toward primary and secondary colors as a result of the hexagonal shape of the drift vectors created from different color hues. See the online article for the color version of this figure.

of $R + G + B = 1$ (a linear constant) is larger relative to $R^2 + G^2 + B^2 = \theta$ (a quadratic constant), producing a shorter accumulation-to-threshold time for RGB and CMY stimuli.

These properties line up with previous work on continuous colors selections by both Ratcliff et al. (2018) and Smith et al. (2020) showing that responses tended to be concentrated on these six values. Our anchored-dimension representation provides a firm theoretical basis for *why* this should occur. Thus, there is no need for Ratcliff et al.'s sine-shaped thresholds or Smith et al.'s extra vector components added to drifts.

Smaller precision coefficients result in less ability to discriminate fine differences between hues, and generate responses that are more heavily biased toward the cardinal hues. As precision increases, this bias decreases and responses shift toward the true hues in the stimulus (bottom panels of Figure 10).

Our focus is on the ability of the model to account for the consequences of manipulations of the number and similarity of the response options. Allowing the anchor-based drifts to shoulder the burden of accounting for distributions of color responses permits us to shift focus toward the response sets instead. However, as elaborated below, model fit was improved by fine-tuning this representation by integrating the effects of differences in subjective similarity as assessed through a multidimensional scaling task performed by each participant. Similar approaches could be used to integrate more elaborate theories of color vision in order to provide a more complete account of perceptual grouping in these tasks.

Thresholds

As for the first study, thresholds play a key role in determining differences in behavior among conditions with different numbers of discrete alternatives. We slightly simplified the approach used in the line-length study by assuming that the thresholds increase linearly with set size (N) at a rate θ_N from a baseline θ_0. This increase compensates, at least in part, for the decrease in accuracy that naturally occurs as the number of responses-alternatives increases.

Simplicity also motivated our approach to the effect of similarity among responses on thresholds. Although the second study partially unconfounds the effects of set size and similarity by using unequally spaced stimuli, thresholds also play a key role in explaining similarity effects. When all responses have equal thresholds, MAAT necessarily predicts that response crowding reduces accuracy due to capture by nearby responses. However, as we show below, responses to options in a choice set that are dissimilar to (spaced further from) other alternatives are also faster as well as being more accurate than responses to options in the same set that have similar competitors. In order to capture the effect on speed in a simple manner, we assume that participants adjust their threshold for each option in the discrete condition as a linear function of its similarity to other options in the choice set. As in Experiment 1, this adjustment is plausible because the similarity among discrete responses was evident from the display and participants were afforded practice trials to make the adjustment.

Formally, we assume that the threshold θ_j required to select response j corresponds to a base value, θ_0, plus the adjustment for the number of response options θ_N, and the adjustment for the degree of similarity to other options in the choice set (formalized as the evidence it confers to all of the other responses):

$$\theta_j = \theta_0 + \theta_N \times N + \kappa \sum_{i \neq j} \boldsymbol{d}_i \cdot \boldsymbol{d}_j. \quad (9)$$

Here, K_D ($0 \leq K_D \leq 1$) is a weighting factor determining how responsive a participant's decision rule is to similarities between response alternatives. The similarities are quantified by the sum of the dot products between the directions for each response option. Hence, thresholds increase with similarity, slowing RT and mitigating the associated decrease in accuracy.

To illustrate threshold setting, suppose there are three responses that align with the directions of the anchors (i.e., $\boldsymbol{d}_R = [0,1]$, $\boldsymbol{d}_G = \left[\frac{\sqrt{3}}{2}, -\frac{1}{2}\right]$, and $\boldsymbol{d}_B = \left[\frac{-\sqrt{3}}{2}, -\frac{1}{2}\right]$). In this case, the sum's of the dot products are the same (-1) for every response, and so the thresholds are the same in every case. If instead, as is shown in Figure 9, two of the responses are more similar to each other (i.e., pink, $d_P = \left(-\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$, and orange, $d_O = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$) than to the third (i.e., cyan, $d_C = (0, -1)$) response, then the similar responses have smaller dot products (both -0.366) than the dissimilar response (-1.73). Hence, the threshold for cyan is smaller than the thresholds for pink and orange, as shown in Figure 9.

The value of K_D modulates the strength of the relative component and so determines the degree to which the threshold adjustment optimizes reward rate (i.e., minimizes response time for a given level of accuracy, see Bogacz et al., 2006, 2010; Kvam, 2019a; Tajima et al., 2019). In many reward-rate paradigms, it is found that threshold adjustment is too small to produce reward rate maximization, and we also found that the values of K_D that participants use are too small to be optimal. As a result, both accuracy and response time vary across manipulations of the stimuli and number of response options.

While the threshold in the continuous condition did not change systematically with the response options or set size as in the discrete condition, as the response options were always the same, we did allow it to vary around its mean estimate according to a normal distribution. This was done to compensate for the fact that the same continuous condition was repeated many times across many trials, blocks, and even sessions of the experiment. It is natural to expect that it would change based on fluctuations in cognitive control and attention (Logan et al., 2014) as well as trial-to-trial shifts in thresholds following errors (Frank et al., 2005; Navarro-Cebrian et al., 2016) and potentially even differences in perceived similarity of each color hue to its competitors. For these reasons, we include a parameter K_C that describes threshold variability across trials in the continuous condition. The threshold in the continuous condition was therefore drawn from a normal distribution $N(\theta_C, K_C)$ on each trial.

It is possible, although extremely rare based on the parameter estimates, for a start point to exceed the threshold for one or more of the choice options. When this happened, the choice option with the highest activation (greatest start point) was predicted as the option to be chosen and the response time for that trial was fixed at the value of nondecision time τ. The effects of manipulating each of the model's parameters is shown in Figure 11.

Study 2: Hue Discrimination

Participants each took part in six study sessions. In the first session, they rated the similarity on a scale of 0–100 of all pair-wise combinations of 30 equally spaced hues. In the remaining five sessions, they completed the decision task. The similarity ratings were used for two purposes. First, they were used to test for the dissimilarity advantage in response times in the decision task. Second, they were used to calculate subjective versions of the d vectors for each participant, which replaced the objective d values in Equation 9.

Method

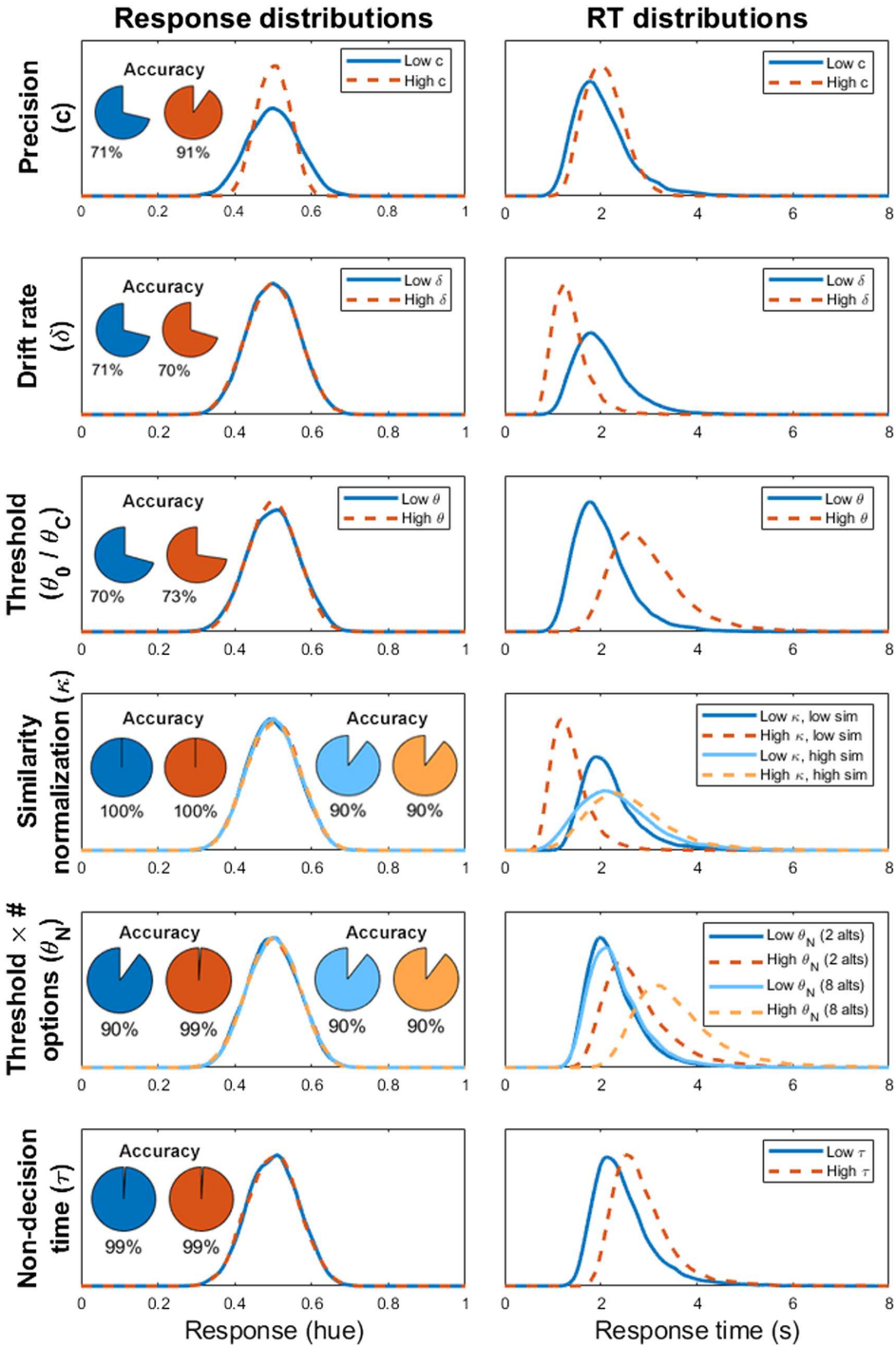
Six Michigan State University graduate students (4 female, 2 male, age range 22–30 years), completed the six sessions. Each participant completed approximately 500–1000 practice trials of the decision task, 1400–2300 experimental trials of the decision task, and 435 trials of the similarity rating task. Stimuli were generated and presented in MATLAB using Psychtoolbox 3 (Brainard, 1997; Kleiner et al., 2007). Analyses used the machine learning and circular statistics MATLAB toolboxes (Berens, 2009). All responses were recorded from the mouse.

Each session took approximately 1 hr to complete. Participants were paid $10 per session for participating and informed of their average accuracy at the end of each session. After completing informed consent, they were placed in a dark, windowless office and completed the similarity rating task in the first session. On later dates, they completed five sessions of the decision task in the same setting.

The similarity rating task was self-paced, so that participants could take as long as they wanted to make exact similarity

Figure 11

Demonstration of the Effects of Manipulating Each Parameter of the Model (Rows) on Distributions of Responses on a Continuum (Left Panels) as Well as Response Times (Right Panels)



Note. Accuracy for discrete conditions is inset as pie charts in the left panels. See the online article for the color version of this figure.

judgments and take breaks as they needed. They were encouraged to take a constant amount of time on each trial and to make sure that their judgments were internally consistent (e.g., a rating of 30 should indicate that a pair of colors is more similar than a rating of 25).

During the first session of the decision task, the experimenter demonstrated how to perform the task in both discrete and continuous conditions, emphasizing that mouse movements should be consistent and ballistic—that is, that participants should not

move the mouse until they were ready to respond, at which point they were to move the pointer directly to the response they wished to make. In addition to their initial briefing and demonstration, participants completed an extra 30 practice trials at the outset of the first session that covered all numbers of alternatives they might see during the task. In the first session, they then completed 10 or 15 blocks (dependent on time) of the decision task, including practice trials. In subsequent sessions, they completed 15 or 20 blocks of the decision task, including practice trials.

Once all six sessions were completed, participants were debriefed on the purpose of the study and the results of their performance, if desired.

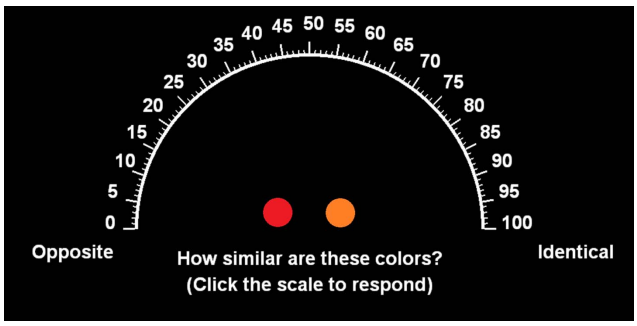
Similarity Rating Task

Figure 12 shows an example similarity rating display. Participants simply had to compare the two colors on the screen and assign a value from 0 (*opposite*) to 100 (*identical*) indicating how similar to one another they thought the colors were.

The colors used for the rating task were 30 hues equally spaced along the color wheel (see the large circle on the right of Figure 13). Participants were presented with each possible combination of two nonidentical colors exactly once, giving a rating for each pair-wise comparison. These pairwise ratings were used to populate the upper diagonal of a similarity matrix, which was used to generate a multidimensional scaling (MDS) solution that arranged the colors in two dimensions. This created an MDS arrangement for each participant that allowed us to personalize the model predictions: even for the same set of parameters, the model would make different predictions for participants with different MDS solutions.

This MDS step enables psychological similarity—as opposed to merely the distance in physical/stimulus space—to be used to predict behavior. Schurgin et al. (2020) showed in working memory tasks fixed-capacity models, where stimulus representations change with set size, are required to account for performance. However, when the subjective similarity between choice options is taken into account, performance can be explained by a single unitary signal detection framework with stimulus representations unaffected by load. Therefore, the version of MAAT that we we fit here uses the *subjective*/psychologically scaled similarity between response options to determined how thresholds are set, while allowing the actual stimulus itself (through the drift rates) to remain invariant to the relationships between stimuli. The exact procedure for setting these thresholds is described in the modeling section.

Figure 12
Layout of the Similarity Rating Task



Note. See the online article for the color version of this figure.

Decision Task

The decision task used in Study 2 is shown in Figure 8. Participants viewed displays of 78 dots scattered around a disc whose diameter subtended approximately 10° of visual angle. The dots varied in hue such that no two colors had a hue that was within 0.04 units of one another (with hue ranging from 0.0 to 1.0, wrapping around such that 0.0 = 1.0). In each display, there was a single dominant dot hue for exactly 18 dots. In addition to the dominant dot color, there were 15 other hues present in the display, with four dots of each. The participants' task was to identify which color was the dominant color in the display, match it to the alternatives shown surrounding the dot display (Figure 8, top right), and respond by moving their mouse to the corresponding hue in the display of alternatives.

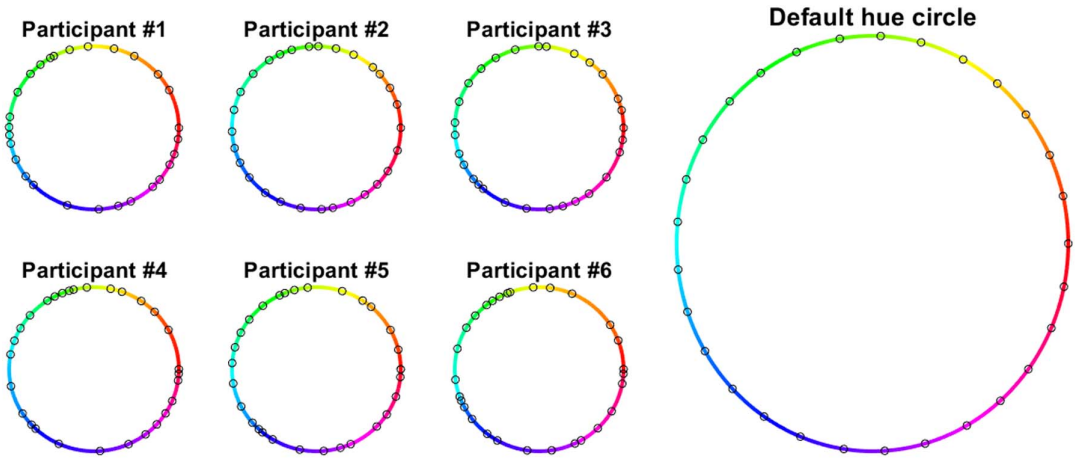
Of the 16 hues that would appear on each block of trials, eight were fixed across a block and eight were drawn randomly from trial to trial. The fixed hues depended on the available response colors—each of the possible response colors had to appear in the set of dots on every trial. The remaining eight nonfixed hues present in the dot display were drawn randomly on every trial subject to the restriction that no pair of hues in the display be closer than .04 hue units apart. The nonfixed hues were never the target color, so they served strictly as distractors or noise in the stimulus. In total, this yielded the 78 (18 target + 7×4 nontarget fixed + 8×4 nontarget random hues) dots in the display.

Each block consisted of 10 practice trials and 30 trials of the decision task. The set of response options was held constant across all 40 trials. A random alternative out of those available was chosen as the dominant color on each trial. The alternatives available to a participant were placed around the edges of a circle at approximately 20 visual degrees from the center, as shown in Figure 8. Each response alternative took up a 14-degree arc along the edge of this circle in the discrete condition (top/left panels). In the continuous condition (bottom right panel), a hue circle was shown where every degree of the circle was a different hue, approximating a continuous gradient of hues. Across trials, this method of displaying the response alternatives ensured that all such alternatives were equidistant from the center of the screen (where the mouse began the trial) and equal in size. Therefore, motor difficulty was matched across conditions, so differences in accuracy and response time were not attributable to motor demands (i.e., Fitt's law Fitts, 1954; MacKenzie & Buxton, 1992).

One issue that arose in Study 1 is stimulus bias, which could result from either a bias toward responding on the "short" side of the scale or result from most participants being right-handed (and thus their responses following a curved trajectory when responding on the left side of the scale; see Kvam & Busemeyer, 2020). To avoid a similar effect in the data from Study 2, we randomly flipped and/or rotated the response scale between sessions. To do so, a random orientation from 0° to 360° was draw, and a binomial random variable (0 or 1) was drawn to determine whether the scale would also be mirrored at the beginning of each session. Participants were oriented to the alignment of the scale with at least 20 practice trials at the beginning of each session, in addition to the 10 practice trials preceding each block of the task. This ensured that response location was not confounded with the hue of participants' responses.

Figure 13

Multidimensional Scaling Solutions for Each Participant (Left) Compared to the Objective Hue Circle (Right)



Note. Locations of colors on the standard hue color wheel (right) compared to the multidimensional scaling solutions for the locations of these colors based on each participant's subjective similarity ratings (left). See the online article for the color version of this figure.

At the beginning of a trial, a participant saw the available response options but not the stimulus. They began the trial by clicking within a small white circle in the middle of the screen, at which point their mouse cursor was centered and the stimulus appeared. Once a trial had started, a response was entered by moving the mouse across the edge of the circle or arcs on which the alternatives were shown. As soon as the mouse cursor crossed this boundary, their response time was recorded and their response was graded as correct or incorrect. In order to match the accuracy criterion across all conditions, responses were considered correct if they were within 7° of the center of the location of the true dominant dot color. In the discrete case, this meant that responses were correct if they crossed the arc colored in the true dominant dot hue. Participants were informed of this grading criterion prior to beginning the study.

Within a session, the number of response alternatives in a block was evenly split between 2, 3, 5, 8, and a continuum of alternatives. Participant saw two blocks of each type per session, with the block order randomly shuffled. At the end of each decision trial, participants received feedback on whether or not their choice was correct in the form of 100 (*correct*) or 0 (*incorrect*) points for that trial. Similarly to Study 1, ballistic mouse movement was encouraged by penalizing participants for straying between the dot display and available alternatives. This penalty was 1 point per every 20 ms above 300 ms from when the cursor started to move to when it indicated a response. For example, a trial where a participant spent 360 ms moving the cursor from its initial position in the center to its final position at the response location incurred a penalty of 3 points.

Practice Trials

In order to ensure that response times were affected as little as possible by practice effects, the physical locations of alternatives, and the time it took to make a ballistic movement to the edge of the circle, there were a minimum of 10 practice trials before every block of decision trials. Each practice trial was similar to the decision trials, except that a single large, colored dot was shown rather than a noisy multicolored dot display. Instead of picking the dominant hue out of the display, participants simply had to match the hue shown in the center of the screen to the alternatives available by moving their mouse through the arc for the corresponding hue. As in the decision task, accuracy and response time were recorded.

The alternatives shown during the practice trials were the same as those in the succeeding decision trials. One of the goals of the practice was to make sure participants knew exactly where each of the alternatives was before they began the decision task. Therefore, for each choice option present in the display of alternatives (2, 3, 5, or 8 for the discrete conditions), a participant saw at least two instances of the corresponding color appear in the center for them to match (meaning there were 16 practice trials in the 8-alternative condition). In the continuous condition, they saw 10 or more random hues appear in sequence in the center. Anytime a participant made an incorrect assignment during the practice trials, an additional practice trial was added.

After each practice trial, the participant received immediate feedback on their accuracy, including the hue they chose, the location of their response on the screen (in terms of degrees around the circle), the correct hue, the correct response's location on the screen, and how far away in degrees their response was from the center of the correct response.

Model Estimation

The model was fit using a standard Metropolis–Hastings algorithm for Markov chain Monte Carlo sampling. For the starting point of each chain, we used a point that was randomly jittered (multivariate normal with standard deviation of .1, .1, .05, .1, .05, .05, .05, and .05 for parameters c, δ, θ_0, θ_C, K_D, τ, θ_N, and K_C, respectively) around the maximum likelihood estimate, which was obtained from a Nelder–Mead simplex algorithm (fminsearch in MATLAB; Lagarias et al., 1998). This used five chains, each of whose length was 1,000 samples, with 300 burn-in samples. Each step in the chain was drawn

from a multivariate normal distribution with standard deviations of $\sigma = [.1, .1, .05, .1, .05, .05, .05, .05]$.

We included two modifications to the standard Metropolis–Hastings MCMC algorithm. First, the likelihood of the data given each set of model parameters was recomputed after every three rejected steps. This was necessary to ensure that the sampler would not get "stuck" at locations where the simulation-based likelihood was unusually high (Holmes, 2015), which can occur when the simulated data that is generated at a particular combination of model parameters happens to line up unusually well with the real data. Second, we included a migration step on every 10 samples, where the value of σ was multiplied by 5 on every 10th step in the chain. This helps the sampler avoid getting stuck at local minima and explore more of the parameter space than including only smaller steps (Turner et al., 2013).

Chains were visually inspected for convergence, and are shown in the Supplemental Materials. The $\hat{r}$ statistics were computed for each participant, where values close to 1 indicated good convergence between chains (Gelman & Rubin, 1992; Robert & Casella, 2010; Roy, 2020). These values were 1.004, 1.01, 1.02, 1.002, 1.002, and 1.01 for Participants 1–6, respectively, indicating good mixing across chains.

This study was not preregistered. Study materials including data, analysis code, and additional figures can be found on the Open Science Framework at https://osf.io/6d29q.

Results

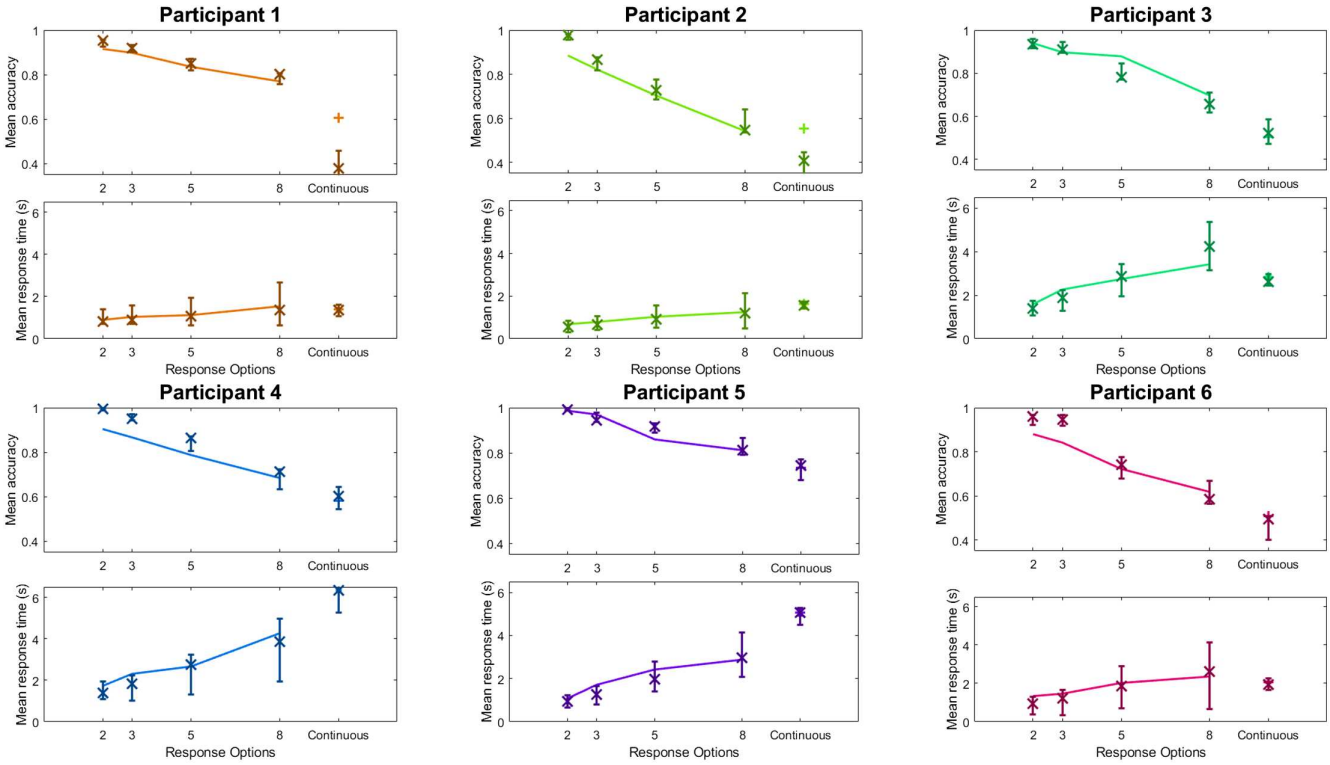
We first report two descriptive analyses. The first tests the form of the set-size effect on response time (i.e., Hick's Law) and accuracy, and the second compares response time in discrete and continuous conditions. Specifically, the second tests whether response time increases for responses that are more similar to other responses. These results are key to evaluation of the model, as they focus on two phenomena that are not predicted by other approaches to modeling continuous-outcome responses. We then report the results of model fitting.

Descriptive Analysis

Figure 14 (data shown as lines) illustrates that for all participants mean response times increased with set size, and mean accuracy decreased. In most cases, response time does not follow Hick's Law, which is predicted by MAAT due to the entanglement between effects of similarity and set size. In fact, overall, a hierarchical Bayesian model predicting mean response time as a function of the number of alternatives showed better fit with a linear link between number of options and RT than one which predicted response times as the $\log_2$ of the number of alternatives, DIC(linear) = 22,176 < DIC($\log_2$) = 24,290.

Figure 14

Predicted and Observed Performance of Each Participant in Study 2



Note. Relationship between the number of alternatives and accuracy (top panels) and and number of alternatives and response time (bottom panels) for each of the participants in the study, including participants 1 (top left, orange), 2 (top middle, green), 3 (top right, turquoise), 4 (bottom left, blue), 5 (bottom middle, purple), and 6 (bottom right, red). Empirical results for each participant are shown as the lines, and best-fit predictions from the model for each participant are shown as Xs in the corresponding color. Bars correspond to the model predictions derived from the 95% most likely parameter estimates (HDIs). See the online article for the color version of this figure.

More problematic for Hick's Law, or indeed a linear increase, is the pattern of response times that appears in the continuous condition. For half of the participants, mean response times in this condition are contained within the range of those produced by discrete numbers of alternatives (usually between 5 and 8). Whether the relationship between the number of response options and mean response times is linear or log-linear, any monotonically increasing effect would predict much longer response times in the continuous than in any of the discrete-alternative conditions simply because there are many more options available.

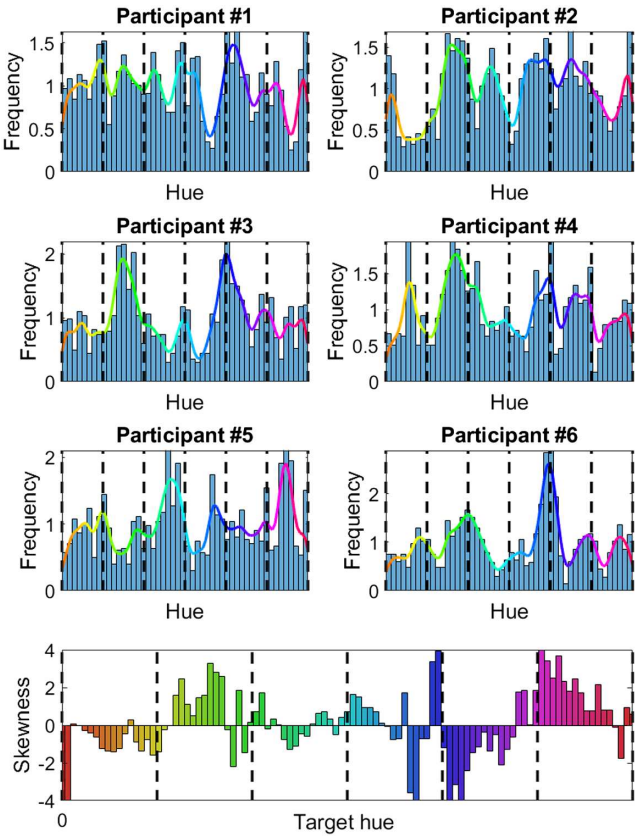
We next evaluated the effect of similarity between a target and distractors and the time it took to respond to the target on each trial. Similarity was measured in terms of the rated similarity between the target hue and its nearest distractor hue, linearly interpolated for hues in between the hues used in the similarity rating task described above. Response times on each trial were rank transformed relative to the block of trials to remove the skew. Response times were nested within conditions and conditions within participants in a hierarchical Bayesian model assuming Gaussian error, allowing us to estimate the within-condition relationship between target-distractor similarity and response time as a random effect. The average slope across each of these relationships between target-distractor distance and response time was −.18 (95% HDI = [−.24, −.12]), indicating that the closer an alternative was to other competitors, the longer participants took to respond to it.[4]

We also examined the distribution of responses in the continuous condition and their skew, as MAAT predicts skewed distributions of responses for target hues adjacent to one of the anchors (i.e., primary or secondary colors on the hue wheel). As described above, the response location was the exact spot that the mouse crossed the edge of the response circle. MAAT also predicts that these anchor-based responses should themselves be more frequent. The result is shown in Figure 15. Skewness was quantified by the third central moment of all responses divided by the cube of its standard deviation, for each group. The groups of responses were created by dividing responses into 100 categories according to the target hue on that trials (all responses are divided into target hues between n and n + .01, for n = 0, .01, .02, . . . , .99). Similar results are obtained with robust (median or mode based) skewness measures. In the observed data, responses tended to group near the primary and secondary colors (top panels). This resulted from a tendency for responses to be heavily skewed toward these responses, reflected in a marked shift in skew from negative to positive when the target shifted from below the anchor to above the anchor (where anchors are indicated by dotted vertical black lines in Figure 15).

We tested this more formally using simple linear correlations of the distance between the target and the nearest anchor [target-anchor distance] and the degree of skew of the responses when a particular hue was the target. This allowed us to evaluate whether the closeness to an anchor affected the skew of distributions. It did, as indicated by a negative correlation between target-anchor distance and response skewness: M = −.24, 95% HDI = [−.38, −.11]. Responses were more frequent near the anchors, as indicated by a negative correlation between the frequency of responses in each bin (out of the 100) and the distance between the target hue and the nearest of the six anchors: M = −.23, 95% HDI = [−.37, −.10] As in Study 1, the results of Study 2 support the prediction of the model, shown in Figure 10, that responses will be skewed toward the anchors, resulting in more frequent responses at these values.

Figure 15

Patterns of Responses for Each Participant, and the Overall Skew (Bottom) Compared to Model Predictions



Note. Distribution of responses for each participant (top six panels; bars are data, lines are kernel-smoothed data superimposed with corresponding hues) across all continuous-condition trials of the experiment and the skewness of these distributions based on the target hue (bottom panel). Dotted vertical lines correspond to anchors (primary/secondary hues) in all plots. See the online article for the color version of this figure.

Model Analysis

The empirical phenomena strongly suggest that an approach like MAAT, where responses are skewed toward anchors and where thresholds are set separately for each option in discrete choice, will out-perform any approach that fails to include these elements. Beyond this, it is still important to evaluate the absolute fit of the model to ensure that it captures not only the qualitative effects but the quantitative patterns in the data.

[4] The objective distance between the target and its competitors additionally affected responses, but only accounted for approximately 0.8% more of the variance in response time above and beyond subjective similarity (R^2 = .101 with only subjective similarity, R^2 = .109 with objective similarity included) and 0.6% of the variance in accuracy above and beyond subjective similarity (R^2 = .048 with only subjective similarity, R^2 = .054 with objective distance added) in regression models, implemented using default priors in JASP (Consonni et al., 2018; JASP Team, 2022). This suggests that the motor difficulty introduced by having alternatives close together did matter somewhat, but its effect was dwarfed by the effect of subjective similarity on performance.

In contrast to Study 1, where the the models were fit based on accuracy (i.e., whether or not a response fell within 7° of the true stimulus hue), we fit the models to the complete distribution of responses in Study 2. This was done for two reasons. First, there are many potential applications of MAAT where there is no "correct" response—such as judgments of price, confidence, preference, or Likert-style ratings. Second, a key element of the response data is their skew. While predicting the skew as an out-of-sample exercise in Study 1 shows that MAAT can capture the skew in principle, fitting the model only to accuracy throws away valuable information that can be used to inform the model fits, especially in the continuous conditions. Therefore, we fit the data from Study 2 using a multinomial likelihood for the discrete conditions and a continuous probability density approximation likelihood in the continuous condition.

The distribution of responses and response times was computed by evaluating the intersection of a ray, projecting from the starting point to the thresholds in either the discrete or continuous condition. For the discrete condition, this can be simplified by computing the path of an accumulator for each option. This is obtained by taking the component of the starting point along each option comp_{vj} (s) (intercept of each accumulator) and the component of the drift along each option comp_{vj} (δ_{all}) (drift of each accumulator), and comparing their values to the threshold for each option θ from Equation 9.

The continuous condition is somewhat more complex to derive distributions of responses. For this condition, we must solve an equation relating the linear path of evidence accumulation to the circular response boundary. As with the line length model, the full solution is presented in the Supplemental Materials.

In both conditions, response time is given by the distance between the start point and the threshold (θ) divided by the rate of accumulation toward the corresponding response option ($\delta_{\text{all}} \cdot \boldsymbol{v}_i$), plus a fixed nondecision time (τ). The likelihood of the data for a particular trial was obtained by generating 500 simulated trials from the proposed set of parameters (at each step in the MCMC chain) for every real trial and calculating the predicted joint distribution of responses and response times by passing a kernel density estimator over the response-RT data to perform probability density approximation (Holmes, 2015; Lin et al., 2019; Turner & Sederberg, 2014).

Incorporating Subjective Similarity

To incorporate the ratings from the similarity task (Figure 12), we remapped the locations of the different color hues on the circle using a circular multidimensional scaling (MDS) procedure (Cox & Cox, 1991; Kvam & Turner, 2021). The results for each participant are shown in Figure 13. The participant-level solutions in this figure show the best MDS solution, so that the change in perceived similarity between any two adjacent dots is the same. As we might expect, colors near the center of the green, blue, and red portions of the color wheel were grouped closer together, indicating that they were perceived to be more similar than colors that were in between the anchors. These subjective similarities were used to inform the estimates of thresholds in the decision tasks: if participants had two hues that appeared very similar (e.g., two green hues), then their thresholds would be higher than if they had two hues that appeared to them very different (e.g., a yellow and orange hue) even if the objective distance between those hues in HSV color space was the same. Formally, the location of each hue in Figure 13 for each participant was substituted for the values of $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ in Equation 9.

To evaluate whether the addition of subjective similarity ratings helped account for behavior on the task, we fit the model with and without this similarity transformation included. The addition of subjective similarity ratings based on the MDS solutions did not add any additional parameters, so the models can be compared based on their raw log likelihoods. For all but one participant (Participant #2), the log likelihood from the model fit improved substantially with the inclusion of the subjective similarity ratings (all log likelihood differences >1,000). Participant #2 did not appear to be sensitive even to objective similarity between options in the choice set, as indicated by the estimates of K_D, so insensitivity to subjective similarity is not too surprising. The Bayes factors for other participants suggest that they were responsive to the perceived similarity between the options in their choice set when they set their thresholds above and beyond the distances between these colors on the raw hue color wheel. The fact that thresholds are responsive to subjective similarity will not be surprising to vision scientists, as it is well-documented that discriminability is not uniform across the hue color wheel (Ohta & Robertson, 2006; Wyszecki & Stiles, 1982). However, it provides further evidence of the importance of subjective similarity to the modeling of both discrete and continuous response tasks (Schurgin et al., 2020). In light of these findings, the results discussed in the next section are from the model that used subjective similarity.

Model Results and Discussion

Maximum a posteriori (MAP) parameter estimates along with the 95% HDIs are presented in Table 2. The mean accuracy and response times predicted by the model for each participant are shown as Xs in Figure 14 and observed versus predicted response (location at which the mouse crossed the response circle) and response time quantiles for both discrete (red) and continuous conditions (yellow) are shown in Figure 16.

Model posterior predictions were generated by simulating 100 trials of artificial data for every real data point from the MAP estimates for each participant, and the 95% HDIs were generated based on the resulting mean accuracy/RT estimates from these 95% most likely parameter values. Detailed results for one example participant are shown in Figure 17, with similar plots for the other participants provided in Supplemental Materials.

The model succeeds in providing a relatively good fit to response times, both in terms of the mean response times shown in the bottom panels of Figure 14, quantiles of the response and RT distributions shown in Figure 16, and in terms of full RT distributions aggregated over participants separately for discrete and continuous conditions in the bottom panel of Figure 17. The one area where improvement could potentially be made is in the tails of the response time distributions, as shown by the model consistently under-estimating the 90th percentile of response times (Figure 16). These response times are quite long, on the order of 2–10 s, which is outside the typical range that perceptual evidence accumulation models typically predict. These may also include trials where participants' attention lapsed or where they had a particularly challenging pair

Table 2

Summary of the Parameters Used in the Model of Behavior in Study 1 (Line Lengths) With Mean of Posterior Estimates Averaged Over Participants and Corresponding 95% Highest-Density Intervals in Square Brackets

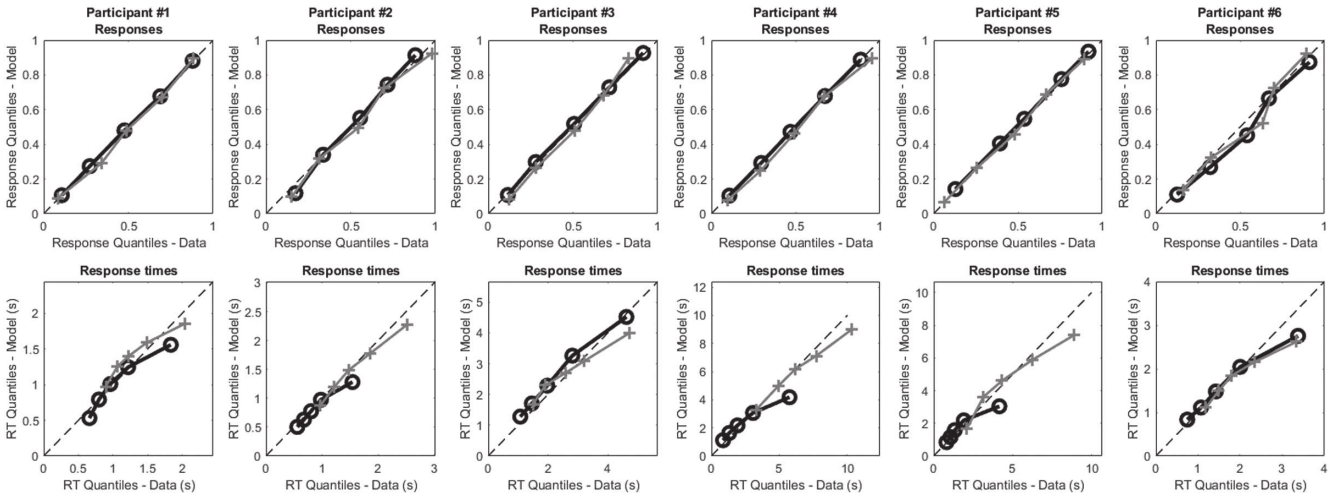
Parameter	Value	Name	Function
s	0.044 [.034, .054]	Start-point variability	Controls distribution of starting points. Higher = greater noise, faster responses, lower accuracy
c	2.61 [2.53, 2.68]	Precision coefficient	Controls the precision of the line length representations relative to the anchors. Higher = greater discrimination, accuracy
δ	14.37 [13.82, 14.89]	Drift rate	Controls the rate of evidence accumulation for all anchors. Higher = faster responses.
θ_3	8.50 [8.20, 8.83]		Controls the amount of evidence needed to make a decision
θ_6	9.48 [9.11, 9.75]	Threshold	Higher = slower but more accurate responses
θ_9	9.97 [9.66, 10.34]		3/6/9 = discrete set size
θ_c	9.40 [9.10, 9.71]		c = continuous
b	0.0098 [0.0096, 0.01]	Bias	Adjusts the stimulus bias to respond toward short versus long responses. Higher = greater bias toward short responses
τ	0.57 [0.43, 0.69]	Non-decision time	Time required (seconds) for nondecision processes (e.g., encoding the stimulus and response production)

of stimuli. We did not include drift rate variability in the model, striving to go as far as we could with threshold adjustments alone, but adding this variability (signifying fluctuations in attention or capacity from trial to trial) may also help the model account for the long tail of response time distributions.

A key factor in the model's success is allowing thresholds to change based on the perceived similarities among a participant's response options. This is indicated by K_D estimates in Table 2, which are much greater than zero for all but one participant. We tested the importance of this component by evaluating the likelihood of a nested model where similarity does not affect thresholds (i.e., $K_D = 0$). To do so, we calculated the Savage–Dickey Bayes factor (Wagenmakers et al., 2010), evaluating the height of the prior at $K_D = 0$ against the height of the posterior at the same point. Here, we present the log Bayes factors: positive values indicate support for threshold changes across sets of options, while negative values indicate that the data provide support against participants shifting their thresholds. The prior for the values of K_D were uniform on [0, 1]. For Participants 1–6, respectively, the log Bayes factors in favor of similarity-based adjustments to thresholds are −0.83 (inconclusive), −1.59 (weakly favoring no adjustment), >1,000 (strongly favoring adjustment), 390.8 (strongly favoring adjustment), >1,000 (strongly favoring

Figure 16

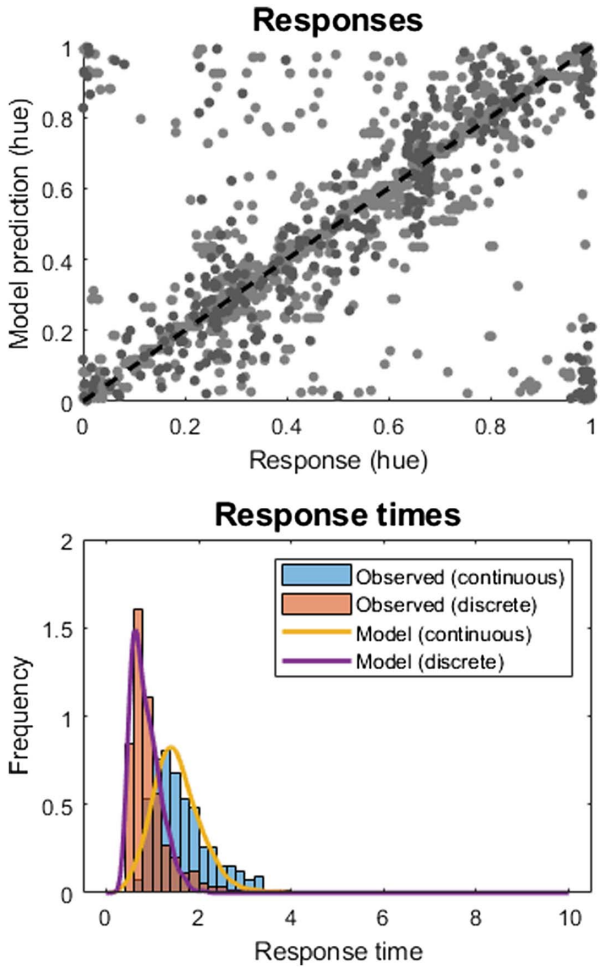
Quantile–Quantile (Q–Q) Plots of the Observed (x-Axis) Versus Prediction of the Model (y-Axis) for the 10th, 30th, 50th, 70th, and 90th Quantiles



Note. These are included for both response/hue (top panels) and response times (bottom panels), for both discrete (black/o) and continuous (gray/+) conditions, and for each participant (columns 1–6, respectively). Points along the dotted diagonal line indicate perfect fit of a particular quantile.

Figure 17

Observed Data and Posterior Model Predictions for the Distribution of Responses (Top) and Response Times (Bottom: Histograms and Lines, Respectively) for Participant 2



Note. Plots for other participants can be found in the supplementary materials. Note that the clumps of points on the top left and bottom right occur because the colour scale wraps around. See the online article for the color version of this figure.

adjustment), and 5.92 (strongly favoring adjustment). Although mixed across participants, this generally supports the inclusion of threshold changes based on similarity in the discrete condition.

Likewise, we can examine the degree of support for threshold variability in the continuous condition by computing the Savage–Dickey Bayes factor on K_C. Since the value of K_C can be any positive value, we used an exponential distribution for the prior, $\Pr(K_C) \sim \text{Exp}(5)$. This resulted in a height of the prior of 0.2 at $K_C = 0$. As before, we report the log Bayes factor for each participant, where positive values support threshold variability and negative values support no threshold variability. The results strongly supported threshold variability in the continuous condition for all but one participant, with Participants 1–6, respectively, having log Bayes factors of −3.43 (strongly disfavoring variability), 120.40 (strongly favoring variability), 422.14 (strongly favoring variability), >1,000 (strongly favoring variability), >1,000 (strongly favoring variability), and 37.90 (strongly favoring variability).

Table 2 also shows that the effect of set size on thresholds (θ_N) varied across participants, suggesting substantial individual differences in the way participants modulated their thresholds as a function of set size. Three participants even had 95% HDIs on θ_N that included zero, indicating that it is credible that they did not change their thresholds based on the number of options on the screen. Much of the variability across set size can be accommodated through the changes in similarity and K_D (because the response space gets more "crowded" Van Maanen et al., 2012), meaning that this parameter only indexes the changes in thresholds with set-size over and above this effect.

The model also provides a relatively good fit to mean accuracy (Figure 14) and distributions of responses (Figures 16 and 17, top panels). It does, however, have a tendency to slightly overestimate the accuracy in the 2- and 3-alternative conditions, perhaps not capturing motor variability that could have led participants to accidentally miss the arc as participants could be incorrect either by selecting a different option or by missing the arcs altogether.

Although the observed responses lined up with those predicted from the model, there were occasional exceptions, particularly in the discrete choice case (e.g., top panel Figure 17, light gray dots) where the model predicted a correct response but participants chose a response alternative in their choice set that was far away from the target choice alternative. However, note that the clusters of points near the top left and bottom right do not represent large misfit, but occur because the response scale wraps around near the extremes.

In summary, the MAAT model provides a parsimonious account that captures the main trends in the distributions of responses and response times across five conditions and a variety of similarity relations with only seven parameters. Notably, the changes in behavior across the number and similarity of response options were handled purely by changes to the thresholds in the model, so that the fundamental perceptions and representations of the stimuli are unaffected by manipulations of response options.

General Discussion

The results of these experiments add to a growing body of work on continuous response tasks, which has discovered skewed responses on bounded scales (Kvam & Busemeyer, 2020), used changes in two-dimensional drift to predict response distributions (Ratcliff, 2018; Smith et al., 2020), and proposed that the mapping from stimulus to response is the main mechanism distinguishing between continuous and discrete response tasks (Smith, 2016).

We evaluated four hypotheses that should hold if our anchor-based modeling framework is valid: (a) bow effects in unidimensional judgments (Study 1); (b) response skew toward anchors (Studies 1 and 2); (c) slower and less accurate responses with an increasing number of response options (Studies 1 and 2); and (d) higher threshold setting for responses with more similar competitors (Study 2). In each case, the empirical phenomena supported the hypotheses and the MAAT model was able to account for the qualitative patterns in the data.

By connecting the discrete and continuous responding we were able to disentangle the perceptual processes related to representing stimuli from the response processes related to generating a decision. MAAT makes strong and plausible selective influence assumptions:

dynamic (drift) components of these tasks are entirely ascribed to stimulus-driven factors—the length or color of the stimulus—while decision thresholds account for response-related factors, such as whether responding is discrete or continuous and in the discrete case the number and similarity of responses options. Approaches that have focused purely on the discrete case have taken a different approach. Usher et al. (2002) propose that an increase of drift rates due to lateral inhibition plays a large part in explaining set-size effects and (Van Maanen et al., 2012) attribute that similarity effects only impact drift rates. van Ravenzwaaij et al. (2019), like us, attribute set-size effects to the rule that determines when accumulation terminates; in their framework through an increase in the number of accumulators that have to reach threshold to trigger a response, rather than in the threshold itself.

Although our selective influence assumptions have some intuitive appeal, these assumptions may be violated in some circumstances such as strategic modulation of attention corresponding to change rates, as has been found in traditional binary-choice paradigms (e.g., Rae et al., 2014). Hence, future applications of MAAT to new paradigms should consider this issue.

Regardless of the mechanism accounting for similarity effects, our results emphasize that any decision model must take into consideration the similarity between target and distractor alternatives in the choice set (Kvam, 2019a; Van Maanen et al., 2012). Further, we showed that incorporating subjective similarity judgments by using multidimensional scaling (Shepard, 1962; Treat et al., 2002) can improve the modeling results for perceptual stimuli. In other domains, approaches like latent semantic analysis (Bhatia, 2013, 2017; Deerwester et al., 1990) for linguistic or more conceptual stimuli may yield similar improvements. Certainly the further development of dynamic models of multidimensional and continuous choice will require quantitative accounts of (stimulus and/or response) similarity and its potential effects on representations and/or decision rules (Schurgin et al., 2020).

Similarity as a consideration in decision-making is not altogether a new idea, but it is one that has been under-emphasized in continuous models (Kvam & Turner, 2021). Critically, threshold values that are sensitive to stimulus similarity lead to a qualitative violation of the predictions of the circular diffusion model (Smith, 2016). By definition, all of the points on a circular boundary are equidistant from the center, meaning that all thresholds must be equal. Therefore, our finding that thresholds can be set separately for different choice options (based on their similarity to other options) conflicts with the circular boundary proposal that is at the heart of that model. It seems likely that this sort of effect will be at play in most discrete-choice scenarios, as similarity is a fundamental part of multi-alternative choice as revealed through context effects (Busemeyer, Gluth, et al., 2019; Sherif et al., 1958; Trueblood et al., 2014).

Although highly flexible, the simulation-based approach that we implemented here makes it much more computationally demanding than the simple analytic likelihood of the circular diffusion model. Once the assumption of a single circular boundary is removed, we are forced either to develop different analytic likelihoods or to approximate the likelihood using simulation-based approaches (Holmes, 2015; Lin et al., 2019; Turner & Sederberg, 2014). Using simple linear ballistic models like the ones developed here greatly reduces the computational burden of a simulation-based approaches, as they require only a small, fixed set of random variables to model across-trial variability as opposed to a large and inconsistently sized set of random variables that is required for models using within-trial variability (Ratcliff et al., 2016). New approaches using machine learning for near-instantaneous estimation of simulation-based models (Radev et al., 2020) or at least for much faster likelihood approximation and MCMC/importance sampling (Fengler et al., 2021) appear promising for alleviating this problem.

Whether response-similarity effects are accommodated through different stopping rules or evidence accumulation rates for the different response options, a decision is triggered when the support for one option minus the (average) support for other options exceeds a criterion. If normalization is performed on the thresholds as in the geometric approach we took here, then the stopping rule automatically reflects a difference between support for different options. Conversely, if normalization is performed on the evidence itself, then the support for an option is redefined as the difference between the evidence for one option minus evidence for the others (normalizing the evidence to sum to zero on a step-by-step basis, as in models like Ratcliff, 2018; Ratcliff & Starns, 2013). Neural and neuroeconomic models tend to pursue the latter route, where *divisive normalization* is applied to the evidence for different responses, which allows the models to predict context effects (Louie et al., 2011; Olsen et al., 2010; Steverson et al., 2019; see also Gluth et al., 2020; Webb et al., 2021). Normalization appears to be an element of the optimal strategies for multi-alternative choice (Tajima et al., 2019), and it can be mimicked by directional thresholds that are adjusted for similarity, as we have done here (Kvam, 2019a). Divisive normalization therefore also provides a neural basis for the similarity-based stopping rule we have proposed.

The connection to evidence representations in binary and multi-alternative choice tasks suggests that selection in discrete and continuous response tasks may involve common neural representations, such as those in LIP and other cortical areas (Churchland et al., 2008). The accumulation process laid out here and shown in Figure 1 can be mimicked by a correlated multiple-accumulator LBA (see Kvam, 2019a, Figure 9), which can be implemented by a simple neural circuit (Tajima et al., 2019, supplementary note 3). Certainly, the model can be informed by a better understanding of the similarity relations between the available responses, which may suggest that (for example) edge categories are more distinct or that the underlying construct does not map onto a purely unidimensional representation (Dodds et al., 2012; Kvam & Turner, 2021). Converging evidence from behavioral approaches like multidimensional scaling (Shepard, 1962) and neural similarity measures like representational similarity analysis or neural decoding rates (Kriegeskorte et al., 2008; Raizada & Connolly, 2012) should shed light on better ways to translate physical stimuli into psychological representations that can be incorporated into a model like MAAT.

Intersecting Paradigms

The continuous version of these tasks reflects response processes that are similar to perceptual and memory based absolute production tasks (Zotov et al., 2010), where participants are asked to produce a (continuous) stimulus based on a previously seen stimulus or a cued category. In the present experiment, instead of a category prompt, participants are attempting to assign a spatial location to the stimulus, but we might expect similar phenomena to be observed in converging response paradigms. Absolute production tasks result in

similar phenomena to those of absolute identification (Dodds et al., 2011), so it should perhaps come as no surprise that a similar model is able to account for behavior in both paradigms. This is true of the present model as well. We have shown that behavior on both tasks—encompassing responses where a stimulus must be mapped to a number or category, and ones where it must be reproduced continuously—can be explained using a single underlying psychological process by simply remapping representations to responses in the different conditions or tasks.

Another domain where continuous responses are commonly required is in continuous function learning tasks (Busemeyer et al., 1997; Koh & Meyer, 1991). In these paradigms, extrapolation (assigning a response to a stimulus outside the experienced range) and interpolation (assigning a previously unseen stimulus inside the experienced range) are key capacities for a model. Interpolation in this type of paradigm is quite simple, and something participants would likely have run into during the task—stimuli that are within the typical range would simply be mapped onto drift rates by substituting the values of the stimulus (line length in Study 1, or RGB value of the dominant color in Study 2) using Equations 4, 5, and 7 to specify the evidence accumulation process. Extrapolation is somewhat more difficult, especially since the color space is inherently circular and so does not have clear "ends" that a new stimulus could go beyond. Even with line length, extrapolation would require recalibration of the anchors in the task. In principle, the stimuli could span only a portion of the quadrant we have used thus far (e.g., 200–300 pixels) and use only part of the response scale. Expected responses to stimuli outside this range could be determined by a simple linear mapping from stimulus to response.

Of course, linear relations between stimulus and response are not the only functions that people can learn and use (though they do seem biased toward these relations Kalish et al., 2007). Further work might examine how people map stimuli to the scale when these relations are nonlinear, such as logarithmic or quadratic mappings. The pattern of response skew, bow effects, and response times in these kinds of tasks would almost certainly provide insight into how people learn to assign internal representations to external scales.

Extensions

One component of the tasks that we did not consider closely was the range of stimuli, which was fixed within both experiments (50–500 pixels in Study 1, and the circular hue range in Study 2). Even though we did not examine how manipulating the range of stimuli and responses affects performance, we can still make predictions about these scenarios. Our results suggested that c and δ do not change with the number or responses, and the same seems likely to be the case with range given these parameters are construable as a measure of the participant's information processing capacity and not something determined by the stimuli and responses. Hence, having a wider range (more extreme upper/lower anchors) will not necessarily result in worse overall performance relative to a narrow range when the number of the categories remains constant. However, it is also possible that a change in the range will affect objective and subjective similarity relationships and so affect threshold adjustments. Clearly, more research is needed to see if MAAT can produce range effects like those observed in typical absolute judgment tasks (Brown et al., 2008; Hutchinson, 1983; Luce et al., 1982; Nosofsky, 1983).

A natural extension of MAAT is to the Likert-style rating tasks that are widely used in other areas of psychology and the social sciences (Mignault et al., 2009). Although activation relative to anchors is an extremely useful tool for generating the drift rates produced by the model, it is not necessary to quantify evidence in terms of support for the longest end and support for the shortest end of the scale. As in the circular diffusion model (Smith, 2016; Kvam, 2019b), the two dimensions of drift can be transformed into a drift *direction* ϕ and a drift *magnitude* $|\delta|$, describing the favored response and the rate of accumulation, respectively. If a modeler elects to use this parameterization of the model, they have to develop a theory that connects stimuli to ϕ and $|\delta|$.

For example, using an opponent-process approach to setting the drift rates, these two parameters are given by simple transformations of the overall drift vector δ_{all} specified in Equations 4 and 5 (where $\delta_{all} = \delta_L + \delta_R$) and 1:

$$\text{Drift direction } \phi = \tan^{-1}(\delta_{all}(2)/\delta_{all}(1)),$$

$$\text{Drift magnitude } ||\delta|| = \sqrt{\delta_{all}(1)^2 + \delta_{all}(2)^2}.$$

In a Likert-style rating task, the drift direction would be mapped onto a distribution over ratings based on which threshold is reached, or by divvying up a single response boundary into discrete responses (as in Figure 1). Across trials, the drift direction and magnitude will naturally vary.

The key to modeling performance on any particular scale is to determine exactly how these two quantities change based on the stimuli presented, and how the participant represents the stimulus relative to the response scale. It is possible that there exist general principles that can determine the drift rate for any given stimulus, but at present it seems more effective to leverage task or paradigm-specific theory to set the drift rates, as in our double-anchor unidimensional account of line length and our trichromatic account of color perception.

Conclusions

The goal of this article was to explore the relationship between discrete and continuous response tasks, with a unified modeling framework used to disentangle the representation of stimuli from the response processes involved in making a decision. Across two experiments, we showed that the same underlying perceptual processes and corresponding accumulation rates can be assumed invariant to the number and similarity of response alternatives. Effects of the number of response options and the similarity between them were reflected in decision rule differences mediated by thresholds changes. We built domain-specific theory into each model, using a double anchor approach for line length and the tri-chromatic theory of color vision for hue, but it seems that the framework developed here should be generally useful in modeling a variety of tasks spanning absolute judgments, absolute production, Likert scale ratings, confidence, and probability, pricing and preference judgments (Kvam & Busemeyer, 2020). We have extended dynamic models that describe decision-making as an accumulation-to-bound process (Busemeyer, Gluth, et al., 2019; Ratcliff et al., 2016) and applied them to more complex mapping tasks while maintaining the important components that allow those models to account well for performance on binary choice tasks (Brown & Heathcote, 2008). This provides the basis for a widely

applicable, dynamic model of perceptually based decision-making among continuous or discrete sets of choice alternatives. We are hopeful that it will pave the way for further advances in other domains like categorization, memory retrieval, and ratings on tasks like preferential choice and confidence (Bhatia & Pleskac, 2019).

References

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*(2), 310–329. https://doi.org/10.1006/jmps.1998.1249

Baranski, J. V., & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 929–945. https://doi.org/10.1037//0096-1523.24.3.929

Berens, P. (2009). CircStat: A Matlab toolbox for circular statistics. *Journal of Statistical Software*, *31*(10), 1–21. https://doi.org/10.18637/jss.v031.i10

Bhatia, S. (2013). Associations and the accumulation of preference. *Psychological Review*, *120*(3), 522–543. https://doi.org/10.1037/a0032457

Bhatia, S. (2017). Associative judgment and vector space semantics. *Psychological Review*, *124*(1), 1–20. https://doi.org/10.1037/rev0000047

Bhatia, S., & Mullett, T. L. (2016). The dynamics of deferred decision. *Cognitive Psychology*, *86*(C), 112–151. https://doi.org/10.1016/j.cogpsych.2016.02.002

Bhatia, S., & Pleskac, T. J. (2019). Preference accumulation as a process model of desirability ratings. *Cognitive Psychology*, *109*, 47–67. https://doi.org/10.1016/j.cogpsych.2018.12.003

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002

Boynton, R. M. (1979). *Human color vision*. Holt, Rinehart and Winston.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Brown, S. D., Marley, A. A. J., Donkin, C., & Heathcote, A. (2008). An integrated model of choices and response times in absolute identification. *Psychological Review*, *115*(2), 396–425. https://doi.org/10.1037/0033-295X.115.2.396

Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462. https://doi.org/10.1016/j.jmp.2009.09.002

Busemeyer, J. R., Byun, E., Delosh, E. L., & McDaniel, M. A. (1997). Learning functional relations based on experience with input-output pairs by humans and artificial neural networks. In K. Lamberts (Ed.), *Knowledge, concepts, and categories*. MIT Press.

Busemeyer, J. R., Gluth, S., Rieskamp, J., & Turner, B. M. (2019). Cognitive and neural bases of multi-attribute, multi-alternative, value-based decisions. *Trends in Cognitive Sciences*, *23*(3), 251–263. https://doi.org/10.1016/j.tics.2018.12.003

Busemeyer, J. R., Kvam, P. D., & Pleskac, T. J. (2019). Markov versus quantum dynamic models of belief change during evidence monitoring. *Scientific Reports*, *9*(1), Article 18025. https://doi.org/10.1038/s41598-019-54383-9

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*(6), 693–702. https://doi.org/10.1038/nn.2123

Consonni, G., Fouskakis, D., Liseo, B., & Ntzoufras, I. (2018). Prior distributions for objective bayesian analysis. *Bayesian Analysis*, *13*(2), 627–679. https://doi.org/10.1214/18-BA1103

Cox, T. F., & Cox, M. A. (1991). Multidimensional scaling on a sphere. *Communications in Statistics - Theory and Methods*, *20*(9), 2943–2953. https://doi.org/10.1080/03610929108830679

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Dodds, P., Brown, S., Zotov, V., Shaki, S., Marley, A., & Heathcote, A. (2011). Absolute production and absolute identification. *Revisiting Miller's Limit: Studies in Absolute Identification*, 148–175.

Dodds, P., Rae, B., & Brown, S. D. (2012). Perhaps unidimensional is not unidimensional. *Cognitive Science*, *36*(8), 1542–1555. https://doi.org/10.1111/cogs.2012.36.issue-8

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*(6), 1129–1135. https://doi.org/10.3758/PBR.16.6.1129

Donkin, C., Brown, S. D., Heathcote, A., & Marley, A. (2009). Dissociating speed and accuracy in absolute identification: The effect of unequal stimulus spacing. *Psychological Research PRPF*, *73*(3), 308–316. https://doi.org/10.1007/s00426-008-0158-2

Donkin, C., Chan, V., & Tran, S. (2015). The effect of blocking inter-trial interval on sequential effects in absolute identification. *The Quarterly Journal of Experimental Psychology*, *68*(1), 129–143. https://doi.org/10.1080/17470218.2014.939665

Dotan, D., Meyniel, F., & Dehaene, S. (2018). On-line confidence monitoring during decision making. *Cognition*, *171*, 112–121. https://doi.org/10.1016/j.cognition.2017.11.001

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*(6), 763–771. https://doi.org/10.3758/PBR.17.6.763

Fengler, A., Govindarajan, L. N., Chen, T., & Frank, M. J. (2021). Likelihood approximation networks (lans) for fast inference of simulation models in cognitive neuroscience. *eLife*, *10*, Article e65074. https://doi.org/10.7554/eLife.65074

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*(6), 381–391. https://doi.org/10.1037/h0055392

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron*, *47*(4), 495–501. https://doi.org/10.1016/j.neuron.2005.06.020

Friedman, J., Brown, S. D., & Finkbeiner, M. (2013). Linking cognitive and reaching trajectories via intermittent movement control. *Journal of Mathematical Psychology*, *57*(3–4), 140–151. https://doi.org/10.1016/j.jmp.2013.06.005

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. https://doi.org/10.1214/ss/1177011136

Gluth, S., Kern, N., Kortmann, M., & Vitali, C. L. (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. *Nature Human Behaviour*, *4*(6), 634–645. https://doi.org/10.1038/s41562-020-0822-0

Guest, D., Kent, C., & Adelman, J. S. (2018). The relative importance of perceptual and memory sampling processes in determining the time course of absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(4), 615–630. https://doi.org/10.1037/xlm0000438

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hawkins, G. E., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012a). Context effects in multi-alternative decision making: Empirical data and a

bayesian model. *Cognitive Science*, *36*(3), 498–516. https://doi.org/10.1111/cogs.2012.36.issue-3

Hawkins, G. E., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012b). An optimal adjustment procedure to minimize experiment time in decisions with multiple alternatives. *Psychonomic bulletin & review*, *19*(2), 339–348. https://doi.org/10.3758/s13423-012-0216-z

Hawkins, G. E., Marley, A. A. J., Heathcote, A., Flynn, T. N., Louviere, J. J., & Brown, S. D. (2014). Integrating cognitive process and descriptive models of attitudes and preferences. *Cognitive Science*, *38*(4), 701–735. https://doi.org/10.1111/cogs.2014.38.issue-4

Heathcote, A., Holloway, E., & Sauer, J. (2019). Confidence and varieties of bias. *Journal of Mathematical Psychology*, *90*, 31–46. https://doi.org/10.1016/j.jmp.2018.10.002

Heathcote, A., & Matzke, D. (in press). Winner takes all! What are race models, and why and how should psychologists use them? *Current Directions in Psychological Science*.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. https://doi.org/10.1080/17470215208416600

Hollands, J., & Dyre, B. P. (2000). Bias in proportion judgments: The cyclical power model. *Psychological Review*, *107*(3), 500–524. https://doi.org/10.1037/0033-295X.107.3.500

Holmes, W. R. (2015). A practical guide to the probability density approximation (pda) with improved implementation and error characterization. *Journal of Mathematical Psychology*, *68*, 13–24. https://doi.org/10.1016/j.jmp.2015.08.006

Hutchinson, J. (1983). On the locus of range effects in judgment and choice. *Advances in Consumer Research*, *10*, 305–308.

Jameson, K. A., Satalich, T. A., Joe, K. C., Bochko, V. A., Atilano, S. R., & Kenney, M. C. (2020). *Human color vision and tetrachromacy*. Cambridge University Press.

JASP Team. (2022). *JASP* (Version 0.16.1) [Computer software]. https://jasp-stats.org/

Kalish, M. L., Griffiths, T. L., & Lewandowsky, S. (2007). Iterated learning: Intergenerational knowledge transmission reveals inductive biases. *Psychonomic Bulletin & Review*, *14*(2), 288–294. https://doi.org/10.3758/BF03194066

Kent, C., & Lamberts, K. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 289–305. https://doi.org/10.1037/0278-7393.31.2.289

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16. https://nyuscholars.nyu.edu/en/publications/whats-new-in-psychtoolbox-3

Koh, K., & Meyer, D. E. (1991). Function learning: Induction of continuous stimulus-response relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 811–836. https://doi.org/10.1037//0278-7393.17.5.811

Koop, G. J., & Johnson, J. G. (2011). Response dynamics: A new window on the decision process. *Judgment & Decision Making*, *6*(8), 750–758. http://journal.sjdm.org/11/m29/m29.html

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, Article 4. https://doi.org/10.3389/neuro.06.004.2008

Kvam, P. D. (2019a). A geometric framework for modeling dynamic decisions among arbitrarily many alternatives. *Journal of Mathematical Psychology*, *91*, 14–37. https://doi.org/10.1016/j.jmp.2019.03.001

Kvam, P. D. (2019b). Modeling accuracy, response time, and bias in continuous orientation judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 301–318. https://doi.org/10.1037/xhp0000606

Kvam, P. D., & Busemeyer, J. R. (2020). A distributional and dynamic theory of pricing and preference. *Psychological Review*, *127*(6), 1053–1078. https://doi.org/10.1037/rev0000215

Kvam, P. D., Busemeyer, J. R., & Pleskac, T. J. (2021). Temporal oscillations in preference strength provide evidence for an open system model of constructed preference. *Scientific Reports*, *11*(1), Article 8169. https://doi.org/10.1038/s41598-021-87659-0

Kvam, P. D., & Heathcote, A. (2022). *A unified theory of discrete and continuous responding*. https://osf.io/6d29q/

Kvam, P. D., & Turner, B. M. (2021). Reconciling similarity across models of continuous selections. *Psychological Review*, *128*(4), 766–786. https://doi.org/10.1037/rev0000296.

Lacouture, Y., Li, S.-C., & Marley, A. (1998). The roles of stimulus and response set size in the identification and categorisation of unidimensional stimuli. *Australian Journal of Psychology*, *50*(3), 165–174. https://doi.org/10.1080/00049539808258793

Lacouture, Y., & Marley, A. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science*, *3*(4), 401–433. https://doi.org/10.1080/09540099108946595

Lacouture, Y., & Marley, A. A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology*, *39*(4), 383–395. https://doi.org/10.1006/jmps.1995.1036

Lacouture, Y., & Marley, A. A. J. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & psychophysics*, *66*(7), 1206–1226. https://doi.org/10.3758/BF03196847

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147. https://doi.org/10.1137/S1052623496303470

Lamberts, K. (2000). An information-accumulation theory of speeded categorization. *Psychological Review*, *107*, 227–260. https://doi.org/10.1037/0033-295X.107.2.227

Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: How action influences perceptual decision making. *PLoS Computational Biology*, *11*(4), Article e1004110. https://doi.org/10.1371/journal.pcbi.1004110

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *140*, 1–55.

Lin, Y.-S., Heathcote, A., & Holmes, W. R. (2019). Parallel probability density approximation. *Behavior Research Methods*, *51*(6), 2777–2799. https://doi.org/10.3758/s13428-018-1153-1

Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, *121*(1), 66–95. https://doi.org/10.1037/a0035230

Longstreth, L. E. (1988). Hick's law: Its limit is 3 bits. *Bulletin of the Psychonomic Society*, *26*(1), 8–10. https://doi.org/10.3758/BF03334845

Louie, K., Grattan, L. E., & Glimcher, P. W. (2011). Reward value-based gain control: Divisive normalization in parietal cortex. *Journal of Neuroscience*, *31*(29), 10627–10639. https://doi.org/10.1523/JNEUROSCI.1237-11.2011

Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-worst scaling: Theory, methods and applications*. Cambridge University Press.

Luce, R. D. (1997). Several unresolved conceptual problems of mathematical psychology. *Journal of Mathematical Psychology*, *41*(1), 79–87. https://doi.org/10.1006/jmps.1997.1150

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. https://doi.org/10.3758/BF03202769

MacKenzie, I. S., & Buxton, W. (1992). Extending fitts' law to two-dimensional tasks. In P. Bauersfeld, J. Bennett, & G. Lynch (Eds.) *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 219–226). Association for Computing Machinery.

Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology*, *37*(2), 136–151. https://doi.org/10.1111/j.2044-8317.1984.tb00797.x

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychophysical judgements applied to magnitude estimation. *Journal of Mathematical Psychology*, *30*(4), 339–390. https://doi.org/10.1016/0022-2496(86)90016-7

Matthews, W. J., & Stewart, N. (2009). The effect of interstimulus interval on sequential effects in absolute identification. *The Quarterly Journal of Experimental Psychology*, *62*(10), 2014–2029. https://doi.org/10.1080/17470210802649285

Mignault, A., Bhaumik, A., & Chaudhuri, A. (2009). Can anchor models explain inverted-u effects in facial judgments? *Perceptual and motor skills*, *108*(3), 803–824. https://doi.org/10.2466/pms.108.3.803-824

Miletić, S., Boag, R. J., Trutti, A. C., Stevenson, N., Forstmann, B. U., & Heathcote, A. (2021). A new model of decision processing in instrumental learning tasks. *eLife*, *10*, Article 309. https://doi.org/10.7554/eLife.63055

Miller, G. A. (1953). What is information measurement? *American Psychologist*, *8*(1), 3–11. https://doi.org/10.1037/h0057808

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. https://doi.org/10.1037/h0043158

Navarro-Cebrian, A., Knight, R. T., & Kayser, A. S. (2016). Frontal monitoring and parietal evidence: Mechanisms of error correction. *Journal of Cognitive Neuroscience*, *28*(8), 1166–1177. https://doi.org/10.1162/jocn_a_00962

Nosofsky, R. M. (1983). Information integration and the identification of stimulus noise and criterial noise in absolute judgment. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(2), Article 299. https://doi.org/10.1037//0096-1523.9.2.299

Nosofsky, R. M. (1997). An exemplar-based random-walk model of speeded categorization and absolute judgment. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of r. duncan luce* (pp. 347–365). Lawrence Erlbaum.

Ohta, N., & Robertson, A. (2006). *Colorimetry: Fundamentals and applications*. Wiley.

Olsen, S. R., Bhandawat, V., & Wilson, R. I. (2010). Divisive normalization in olfactory population codes. *Neuron*, *66*(2), 287–299. https://doi.org/10.1016/j.neuron.2010.04.009

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press.

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. https://doi.org/10.1037/0033-295X.112.2.383

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. https://doi.org/10.1037/A0019737

Plummer, M. (2003). Jags: A program for analysis of bayesian graphical models using gibbs sampling. In K. Hornik & F. Leisch (Eds.), *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, pp. 1–10). Achim Zeileis.

Radev, S. T., Mertens, U. K., Voss, A., & Köthe, U. (2020). Towards end-to-end likelihood-free inference with convolutional neural networks. *British Journal of Mathematical and Statistical Psychology*, *73*(1), 23–43. https://doi.org/10.1111/bmsp.v73.1

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. D. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1226–1243. https://doi.org/10.1037/a0036801

Raizada, R. D., & Connolly, A. C. (2012). What makes different people's representations alike: Neural similarity space solves the problem of across-subject fmri decoding. *Journal of Cognitive Neuroscience*, *24*(4), 868–877. https://doi.org/10.1162/jocn_a_00189

Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review*, *125*(6), 888–935. https://doi.org/10.1037/rev0000117

Ratcliff, R., & McKoon, G. (2020). Decision making in numeracy tasks with spatially continuous scales. *Cognitive Psychology*, *116*, Article 101259. https://doi.org/10.1016/j.cogpsych.2019.101259

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Ratcliff, R., & Starns, J. J. (2009). Modeling Confidence and Response Time in Recognition Memory. *Psychological Review*, *116*(1), 59–83. https://doi.org/10.1037/a0014086

Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, *120*(3), 697–719. https://doi.org/10.1037/a0033152

Ratcliff, R., Voskuilen, C., & Teodorescu, A. (2018). Modeling 2-alternative forced-choice tasks: Accounting for both magnitude and difference effects. *Cognitive Psychology*, *103*, 1–22. https://doi.org/10.1016/j.cogpsych.2018.02.002

Reynolds, A., Garton, R., Kvam, P., Sauer, J., Osth, A. F., & Heathcote, A. (2021). A dynamic model of deciding not to choose. *Journal of Experimental Psychology: General*, *150*(1), 42–66. https://doi.org/10.1037/xge0000770

Reynolds, A., Kvam, P. D., Osth, A. F., & Heathcote, A. (2020). Correlated racing evidence accumulator models. *Journal of Mathematical Psychology*, *96*, Article 102331. https://doi.org/10.1016/j.jmp.2020.102331

Robert, C. P., & Casella, G. (2010). *Introducing Monte Carlo Methods with R* (Vol. 18). Springer.

Roy, V. (2020). Convergence diagnostics for markov chain monte carlo. *Annual Review of Statistics and Its Application*, *7*, 387–412. https://doi.org/10.1146/statistics.2020.7.issue-1

Schnapf, J., Kraft, T., & Baylor, D. (1987). Spectral sensitivity of human cone photoreceptors. *Nature*, *325*(6103), 439–441. https://doi.org/10.1038/325439a0

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature Human Behaviour*, *4*(11), 1156–1172. https://doi.org/10.1038/s41562-020-00938-0

Seibel, R. (1963). Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology*, *66*(3), 215–226. https://doi.org/10.1037/h0048914

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika*, *27*(3), 219–246. https://doi.org/10.1007/BF02289621

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, *55*(2), 150–155. https://doi.org/10.1037/h0048784

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425–451. https://doi.org/10.1037/rev0000023

Smith, P. L. (2019). Linking the diffusion model and general recognition theory: Circular diffusion with bivariate-normally distributed drift rates. *Journal of Mathematical Psychology*, *91*, 145–158. https://doi.org/10.1016/j.jmp.2019.06.002

Smith, P. L., & Corbett, E. A. (2019). Speeded multielement decision-making as diffusion in a hypersphere: Theory and application to double-target detection. *Psychonomic Bulletin & Review*, *26*(1), 127–162. https://doi.org/10.3758/s13423-018-1491-0

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8

Smith, P. L., Saber, S., Corbett, E. A., & Lilburn, S. D. (2020). Modeling continuous outcome color decisions with the circular diffusion model: Metric and categorical properties. *Psychological Review*, *127*, 562–590. https://doi.org/10.1037/rev0000185

Smith, P. L., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, *53*(2), 293–315. https://doi.org/10.1348/000711000159349

Steverson, K., Brandenburger, A., & Glimcher, P. (2019). Choice-theoretic foundations of the divisive normalization model. *Journal of Economic Behavior & Organization*, *164*, 148–165. https://doi.org/10.1016/j.jebo.2019.05.026

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*(4), 881–911. https://doi.org/10.1037/0033-295X.112.4.881

Steyvers, M. (2011). *Matjags 1.3: A matlab interface for jags*. https://github.com/msteyvers/matjags

Tajima, S., Drugowitsch, J., Patel, N., & Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, *22*(9), 1503–1511. https://doi.org/10.1038/s41593-019-0453-9

Teodorescu, A. R., Moran, R., & Usher, M. (2016). Absolutely relative or relatively absolute: Violations of value invariance in human decision making. *Psychonomic Bulletin & Review*, *23*(1), 22–38. https://doi.org/10.3758/s13423-015-0858-8

Teodorescu, A. R., & Usher, M. (2013). Disentangling decision models: From independence to competition. *Psychological Review*, *120*(1), 1–38. https://doi.org/10.1037/a0030776

Townsend, J. T. (2008). Mathematical psychology: Prospects for the 21st century: A guest editorial. *Journal of Mathematical Psychology*, *52*(5), 269–280. https://doi.org/10.1016/j.jmp.2008.05.001

Townsend, J. T., & Wenger, M. J. (2004). A theory of interactive parallel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, *111*(4), 1003–1035. https://doi.org/10.1037/0033-295X.111.4.1003

Treat, T. A., McFall, R. M., Viken, R. J., Nosofsky, R. M., MacKay, D. B., & Kruschke, J. K. (2002). Assessing clinically relevant perceptual organization with multidimensional scaling techniques. *Psychological Assessment*, *14*(3), 239–252. https://doi.org/10.1037/1040-3590.14.3.239

Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*(1), 68–111. https://doi.org/10.1037/0033-295X.91.1.68

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*(2), 179–205. https://doi.org/10.1037/a0036137

Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review*, *21*(2), 227–250. https://doi.org/10.3758/s13423-013-0530-0

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*(3), 368–384. https://doi.org/10.1037/a0032222

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology*, *46*(6), 704–715. https://doi.org/10.1006/jmps.2002.1420

Van Maanen, L., Grasman, R. P., Forstmann, B. U., Keuken, M. C., Brown, S. D., & Wagenmakers, E.-J. (2012). Similarity and number of alternatives in the random-dot motion paradigm. *Attention, Perception, & Psychophysics*, *74*(4), 739–753. https://doi.org/10.3758/s13414-011-0267-7

van Ravenzwaaij, D., Brown, S. D., Marley, A. A. J., & Heathcote, A. (2020). Accumulating advantages: A new conceptualization of rapid multiple choice. *Psychological Review*, *127*(2), 186–215. https://doi.org/10.1037/rev0000166

Vickers, D. (2001). Where does the balance of evidence lie with respect to confidence? In E. Sommerfeld, R. Kompass, & T. Lachmann (Eds.), *Proceedings of the 17th Annual Meeting of the International Society for Psychophysics* (pp. 148–153). Pabst.

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179–197. https://doi.org/10.1016/0001-6918(82)90006-3

Wagenmakers, E. J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001

Ward, L. M., & Lockhead, G. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, *84*(1), 27–34. https://doi.org/10.1037/h0028949

Webb, R., Glimcher, P. W., & Louie, K. (2021). The normalization of consumer valuations: Context-dependent preferences from neurobiological constraints. *Management Science*, *67*(1), 93–125. https://doi.org/10.1287/mnsc.2019.3536

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 385–398. https://doi.org/10.1037/a0034851

Wyszecki, G., & Stiles, W. S. (1982). *Color science* (Vol. 8). Wiley.

Yu, S., Pleskac, T. J., & Zeigenfuse, M. D. (2015). Dynamics of postdecisional processing of confidence. *Journal of Experimental Psychology: General*, *144*(2), 489–510. https://doi.org/10.1037/xge0000062

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*(4), 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x

Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, *22*(11), 1434–1441. https://doi.org/10.1177/0956797611417006

Zotov, V., Shaki, S., & Marley, A. (2010). Absolute production as a possible method to externalize the properties of context dependent internal representations. In A. Bastianelli & G. Vidotto (Eds.), *Proceedings of the 26th annual meeting of the international scoeity for psychophysics* (pp. 203–208). The International Society for Psychophysics.

Received July 27, 2021
Revision received May 8, 2022
Accepted May 21, 2022 ■