in press. JEP:General

Further tests of sequence-sensitive models in a modified Garner task using separable dimensions

Deborah J. Lin & Daniel R. Little
The University of Melbourne

Abstract

In the study of perceptual categorization, a key distinction is made between separable and integral dimensions. Separable dimensions are easy to attend in isolation, while integral dimensions are not. Little, Wang, and Nosofsky (2016) showed that when trial-by-trial responses are analyzed, a consistent pattern of sequential effects was found in a modified Garner paradigm using integral-dimension stimuli. The present experiments investigate whether these pronounced sequential effects are also found with separable-dimension stimuli. Four experiments using two different types of separable dimensions were conducted. The results indicate that similar patterns of sequential effects were present for separable-dimension stimuli, but, unlike for integral dimensions, the effect of a change in the irrelevant dimension in the filtering task was not found. Further, for separable dimensions, the overall pattern of sequential effects did not vary between the Garner tasks (i.e., control, correlated, and filtering). To explain these results, we fit a sequence-sensitive exemplar model and compared the fits of this model to a novel sequence-sensitive feature model, in which only the relevant feature influences the categorization decision. We find that the full exemplar model provides a more compelling account of both our separable dimension data and the integral dimension data of Little et al. (2016). These findings provide a more complete understanding of perceptual categorization and add to the growing body of literature on the prevalence and critical implications of strong sequential effects in cognitive tasks.

Portions of this paper were completed as part of partial fulfillment of an honours degree project by Deborah J. Lin. This work was supported by ARC Discovery Project Grant DP160102360 and a Melbourne Research Grant Support Scheme Grant to Daniel R. Little. We would like to thank Ariel Goh for assistance with data collection. Correspondence concerning this article should be addressed to Daniel R. Little, School of Psychological Sciences, University of Melbourne, Melbourne, Victoria 3010, Australia. E-mail: daniel.little@unimelb.edu.au. Data and code are available from: https://github.com/knowlabUnimelb/MODGARNER

Nearly every choice we make can be interpreted as a categorization judgment. For example, choosing which pastry to buy at the local bakery could entail categorization as sweet or not sweet, deciding whether to approach someone at an event could require a categorization judgment of their friendliness, and deciding which COVID vaccine to get would involve considering relevant medical advice. In these scenarios, dimensional information such as the savouriness of a pastry, the warmth in a facial expression, or the cost/benefit analysis associated with a vaccine could be used to inform these judgments. Categorization often involves combining information across several dimensions. The details of this combination depend on, among other things, the type of features being integrated along with one's existing knowledge and recent experience of a feature's relevance. The purpose of this paper is to examine the influence of recent categorization judgments on the categorization of perceptual stimuli comprised of separable and integral dimensions.

The distinction between separable and integral dimensions is of foundational importance for the understanding of information integration (Garner, 1974; Shepard, 1964, 1987). Separable dimensions can be easily attended in isolation without interference from other dimensions. Examples of separable dimensions include the size and shape of an object or the location and orientation of a line. By contrast, integral dimensions are difficult to attend to independently of one another (Lockhead, 1966). An example of integral dimensions are the brightness and saturation of a color; identifying whether a change in the color is due to a change in brightness or a change in saturation is difficult (Fifić, Nosofsky, & Townsend, 2008; Little, Nosofsky, Donkin, & Denton, 2013; Shepard, 1987).

Whether a pair of dimensions are separable or integral can be determined using Garner's (1974) set of selective attention tasks (see also Algom & Fitousi, 2016; Garner & Felfoldy, 1970; Gottwald & Garner, 1972). In these tasks, subjects are required to repeatedly sort stimuli that vary on multiple dimensions into one of two possible groups as quickly and accurately as possible based on the values of a single relevant dimension. Three tasks – control, correlated, and filtering – are used; these tasks vary in the structure of the stimulus space but maintain a constant group membership along a single relevant dimension, as illustrated in the top panel of Figure 1.

In the control or baseline task, two stimuli vary only along the relevant dimension (i.e., dimension X in Figure 1), such that stimulus 1 belongs to one category and stimulus 2 belongs to the other. In the correlated task, two stimuli vary along both the relevant dimension and a second irrelevant dimension (i.e., dimension Y). In the filtering task, four stimuli vary along both the relevant and an irrelevant dimension, but the response is determined solely by the stimulus' value along the relevant dimension. In all conditions, participants are instructed to identify the value of the stimulus along the relevant dimension according to this common rule; that is, subjects are instructed to attend only to the relevant dimension and to ignore any variation in the irrelevant dimension.

With separable dimensions, a robust finding is that response times (RTs) are invariant across the control, correlated, and filtering tasks (Garner, 1974; Garner & Felfoldy, 1970) suggesting that selective attention limits any influence of the irrelevant dimension. In contrast, with integral dimensions, participants tend to respond faster in the correlated task, suggesting a facilitation effect, and slower in the filtering task, suggesting an interference effect (Garner & Felfoldy, 1970; Grau & Nelson, 2016; Nosofsky & Palmeri, 1997a; Shepp & Swartz, 1976). These findings provide a set of benchmark results for theories of attention and categorization.

Formal models of the Garner task

Several formal models have been proposed to explain these results. For instance, models based on general recognition theory (GRT; Ashby & Townsend, 1986) - a multivariate generalization of signal detection theory (Green & Swets, 1966), and the associated decision-bound theory (DBT; Ashby & Gott, 1988), hereafter GRT-DBT, posit that each stimulus is represented by a multivariate probability density. This density provides the probability of sampling a percept with specific values on both dimensions. A decision boundary is established by the participant to separate different response regions. A response is made by sampling a percept and determining the region in which the sampled percept falls.

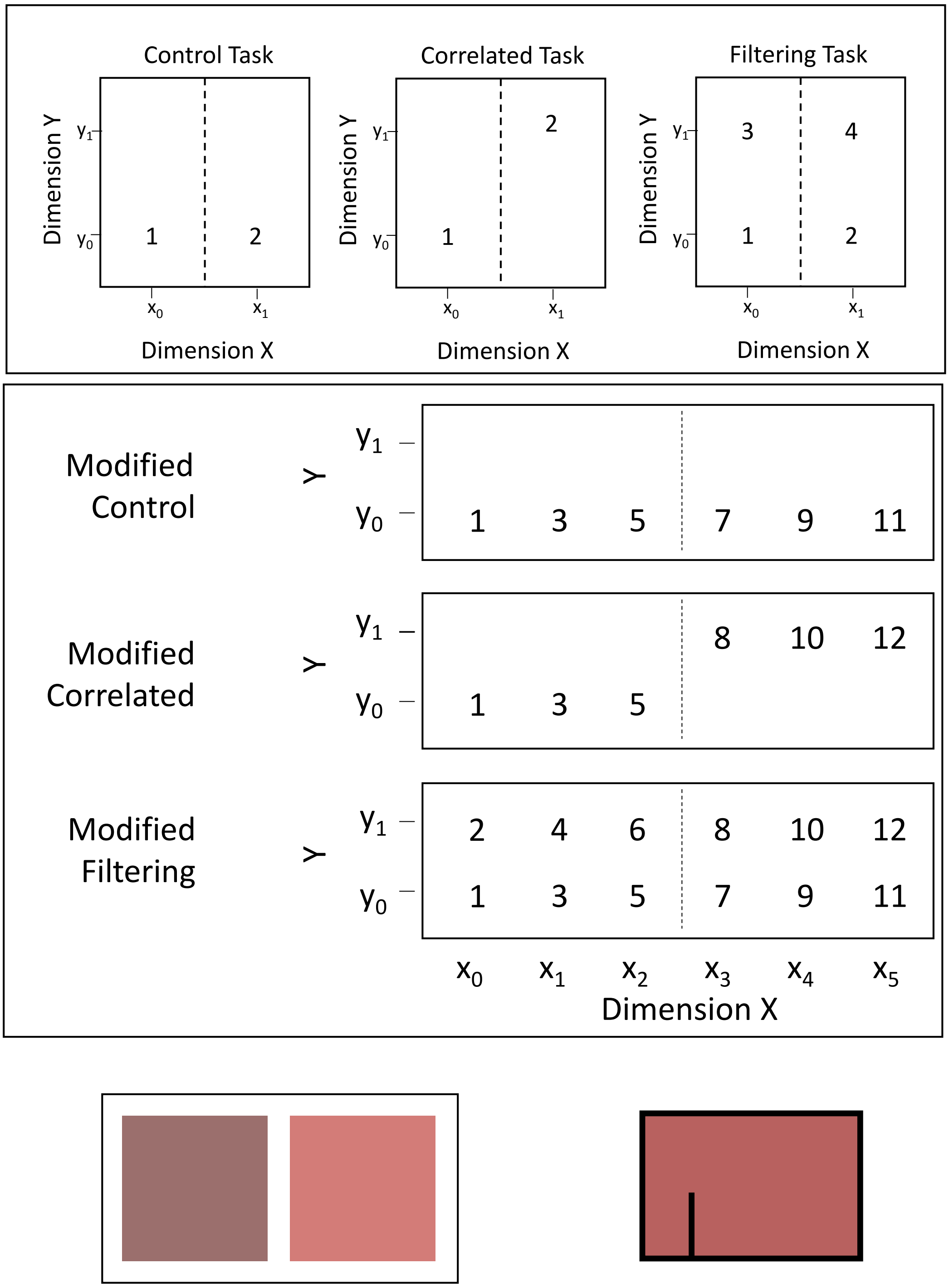

Figure 1. Top panel: Standard Garner task (see e.g., Garner, 1974). Middle panel: Modified Garner task from Little, Wang & Nosofsky (2016). Bottom panels: Left - Example stimulus from Experiments 1 & 3. Right - Example stimulus from Experiments 2 & 4. Top panel adapted from Little, Wang & Nosofsky (2016).

Additional assumptions have been used to predict RTs making the model applicable to the standard Garner tasks. One assumption was that RT decreased monotonically with the distance of a sampled percept from the category boundary (Ashby, 1989, 2000; Ashby, Boynton, & Lee, 1994; Maddox & Ashby, 1996). However, when coupled with other assumptions, the RT-distance hypothesis makes predictions which are not observed in data. [1]

To explain the separable condition results, GRT-DBT assumes that the means and variances of the item distributions on the relevant dimension do not change as the irrelevant dimension is varied and that the decision boundary is orthogonal to the relevant dimensional axis (Ashby et al., 1994; Maddox & Ashby, 1996). The implication is that the density in each response region is equivalent across conditions resulting in equivalent predictions for performance. By contrast, to explain the integral condition results, GRT-DBT assumes that the locations or variances of the item distributions differ across items and/or that the decision boundary is not orthogonal to the relevant dimensional axis (Maddox, 1992).

An alternative, exemplar-based approach, assumes that the presented stimulus activates previously stored items which race to be retrieved at rates determined by their similarity to the current item (Nosofsky, 1986). In the separable condition, more weight is given to the relevant dimension, consequently making items which share values of the relevant dimension more similar to each other and less similar to other items. For integral dimensions, the weights are distributed approximately equally across dimensions capturing a lack of ability to selectively attend to integral dimensions (Nosofsky, 1987). The implication is that, since the dimensions are weighted equally, differences in the distances between stimuli (see Figures 1; top row) drive differences in similarity, which in turn produce different RT predictions across conditions.

[1] The RT-distance hypothesis is typically coupled with the assumption that the filtering condition has increased perceptual variability compared to the control condition. This assumption accurately predicts that the mean RT in the filtering condition will be longer than the mean RT in the control condition; however, this assumption also leads to an erroneous prediction that the fastest responses in the filtering condition should be faster than fastest responses in the control condition (Nosofsky & Palmeri, 1997a). This latter prediction is not observed in the empirical data (Little, Wang, & Nosofsky, 2016; Nosofsky & Palmeri, 1997a).

Sequential effects

One challenge to these theoretical accounts is the presence of marked sequential effects. For instance, in the original Garner task, stimulus repetitions are responded to faster than non-repetitions (Felfoldy, 1974; Garner, 1974). One explanation holds that filtering interference arises because there are fewer exact repetitions in the filtering task compared to the control or correlated task (Garner, 1974). Furthermore, there is an additional cost when the stimulus changes but the response remains the same (compared to the control task). A formal theory of these effects was recently proposed and tested using a modified version of the Garner task (see the middle panel of Figure 1).

Little, Wang, and Nosofsky (2016) established that large trial-by-trial sequential effects arise during the categorization of integral dimensions. Their modified tasks extended the standard Garner stimulus space by increasing the number of stimuli along the relevant dimension (see Figure 1). This manipulation was designed to reduce the number of direct stimulus repetitions. This was an important modification because in the standard Garner task, the control and correlated tasks can be accurately completed using a "bypass" strategy that uses only information from the previous item (Fletcher & Rabbitt, 1978; Krueger & Shapiro, 1981): if the stimulus on the current trial was repeated from the previous trial then make the same category response as the previous trial; otherwise switch responses. One explanation for the slower RTs in the filtering task compared to the control task is that participants use different strategies in each: a bypass strategy in the control task and some different strategy in the filtering task. The modified task removes this confound because a bypass strategy does not work for any of the tasks. The results using the modified task nevertheless indicated that the similarity and category membership of the preceding stimulus exerted a strong influence on the RT and accuracy of the current trial. Hence, other types of sequential effects were evoked by use of the modified task.

To summarize Little et al.'s (2016) results, for items two or three steps from the category boundary, the item presented on the previous trial had a negligible effect on the current response since these items were near the floor for RT and near the ceiling for accuracy. However, the similarity of the previous item had large effects on items near the category boundary. Those effects, and their relation to standard Garner task sequential effects, can be described as follows:

1. a *near-boundary repetition effect* in which accuracy was higher and RTs shorter when items near the decision boundary were repeated (e.g., item 5 preceded by item 5 in Figure 1). This is similar to the repetition effect seen in all three of the standard Garner task conditions.

2. a *same-category far-item "pushing" effect* in which RTs were longer and accuracy was lower for the categorization of near items immediately preceded by far items of the same category (e.g., item 5 preceded by item 1 in Figure 1). The near and far items have comparatively low similarity within the context of the task, and a relatively large change in the relevant dimension value would typically suggest that the current item falls into the opposite category. This effect is analogous to the category contrast effect observed by Stewart, Brown, and Chater (2002) for unidimensional stimuli.

   In the standard Garner filtering task, there is an RT cost associated with trials on which the stimulus changes but the response remains the same (Garner, 1974). This cost likely represents the same type of effect as the pushing effect that we observe in the modified task. In the standard task, the cost due to a same category item change is not as large as the cost due to a change in both the response and the item. Hence, one cannot account for the effect simply by assuming a single mechanism such as a propensity to switch responses when the stimulus changes (Fletcher & Rabbitt, 1978; Krueger & Shapiro, 1981).

3. an *opposite-category adjacent-item "pulling" effect* in which RTs were longer and accuracy lower for near items that were immediately preceded by a near item in the opposite category (e.g., item 5 preceded by item 7 in Figure 1). This effect is similar to assimilation effects in unidimensional identification tasks (Lockhead & King, 1983; Stewart et al., 2002; M. Treisman & Williams, 1984; Ward & Lockhead, 1970). In

contrast to the pushing effect, the intuition is that highly similar items often belong to the same category, but for this effect, relying on similarity alone can be misleading because although the items are highly similar, they belong to different categories. Both the pushing and pulling effects introduce uncertainty about the category membership of items near the boundary since the similarity between the preceding and current item suggests a different response to the actual category membership. As indicated above, in the standard Garner task, there is a large cost to changing both the item and the response from the previous to the current trial. This could represent an effect analogous to Little et al.'s (2016) pulling effect.

Though these effects were anticipated by similar results in unidimensional categorization tasks (Stewart et al., 2002; Stewart & Brown, 2004), there were additional differences between the modified Garner tasks when integral dimensions were used. These differences can be understood by considering how changes in the distance between stimuli change the perceived similarity between stimuli. For example, all of the effects were generally attenuated in the correlated task relative to the control task. In the correlated task, the distances between opposite category items are larger than in the control task. This would result in a smaller pulling effect because of the reduction in similarity at larger distances. The pushing effect would also be reduced because the near items are less similar to the opposite category items in the correlated condition than in the control condition.

In the filtering task, there were also changes in the sequential effects depending on whether the irrelevant dimension was the same or different from the previous trial (e.g., a change from item 5 to item 6; see Modified Filtering task in Figure 1). Changing the irrelevant dimension, along with either holding the value of the relevant dimension constant (in a repetition trial) or changing the value on the relevant dimension (in a pushing or pulling effect trial), acted to increase the distance between the previous and current item. Little et al. observed that when the irrelevant dimension varied in the filtering task, the strengths of the repetition and pulling effects were attenuated while the strength of the pushing effect was enhanced in line with the similarity-distance account above.

This explanation was instantiated formally in a sequence-sensitive exemplar-based (sequential EB) model that generalized Nosofsky and Palmeri's (1997b) Exemplar-Based Random Walk (EBRW) model to account for the between-task differences along with the sequential effects. For example, EBRW assumes that when a to-be-categorized item is presented, exemplars stored in memory race to be retrieved. When a category A or category B item is retrieved, a random walk counter is incremented toward the category A or B threshold, respectively. When the random walk hits either threshold, the decision ends, and the corresponding response is made. The sequential EB model assumes that, in addition to this "long-term" exemplar retrieval, there are additional boosts to the decision process arising from the retrieval of recently presented items. That is, when presented with an item to be categorized, the preceding item is retrieved with some probability. The previous item's *category* receives a boost to the extent that the preceding item is similar to the current item. A second mechanism introduced in the sequential EB model is a bias to switch responses when the current stimulus is different from the previous stimulus.[2]

Garner (1974) reported that, in the filtering task, there was little difference between trials in which there was a change along the relevant dimension (and hence a change in the response) and trials in which there was a change along both the relevant and irrelevant dimensions. Any effects of similarity between successive stimuli in the standard filtering tasks appear to be swamped by the large effect of changing the response. (Alternatively, with only a few stimuli, the differences in discriminability may already be very large even within a dimension so that any additional change in discriminability has little effect). Our modified design is more sensitive to the effects of similarity between the previous and current stimulus presumably since stimuli vary across a larger range. These similarity-based sequence effects turn out to be important for distinguishing our sequence-sensitive exemplar-based model from a sequence-sensitive rule-based account.

Little et al. (2016) introduced a second sequence-sensitive model that generalized the GRT-DBT account of the Garner task by coupling the GRT-DBT with a linear ballistic

[2] For computational reasons, the sequence-sensitive model was implemented using the continuous time linear ballistic accumulator (LBA; Brown and Heathcote, 2008) rather than a random walk.

accumulator (LBA; Brown and Heathcote, 2008) to explain RTs and by adding two mechanisms to explain the sequential effects. In this rule-based model, hereafter the sequence sensitive *GRT-LBA*, it was assumed that when an item is repeated, the perceptual variability of that item is reduced leading to faster responding. The GRT-LBA model also assumed that the decision boundary on the current trial shifted to increase or decrease the region of the category space occupied by the previous item's category. The implication is that the sequential effects are not based on item-specific similarity effects like in the exemplar-based model but on category-specific boundary effects. A consequence of these mechanisms is that the GRT-LBA model does not predict a pushing effect per se, but the difference between a repeated item and a near item preceded by any member of the same category should be equivalent. Likewise, the sequence-sensitive GRT-LBA model also predicted that the pulling effect of an adjacent opposite category near item preceding a near item should be equivalent to the pulling effect of a middle or far opposite category item. An illustrative schematic of these predictions is presented alongside the predictions of the sequence-sensitive exemplar model in Figure 2. A further short-coming of the sequential rule-based model is that it does not predict any effect of changing the value on the irrelevant dimension in the filtering task; however, for integral dimensions, the repetition, pushing, and pulling effects all vary depending on whether or not the irrelevant dimension is the same or different from the previous trial. Little et al. (2016) found that a few subjects had results in line with the sequence sensitive GRT-LBA predictions, but most did not.

One untested prediction that follows from the selective attention mechanism in the sequential EB-LBA model is that, for separable dimensions, we should not find any differences between tasks; however, because the sequential effects are determined by different, independent mechanisms, we should see similar patterns of sequence effects as those observed by Little et al. (2016) for integral dimensions. While Garner (1974) did not examine sequential effects from the separable dimension condition in the standard task, it is expected that these effects exist. For instance, even if selective attention renders the correlated, control and filtering conditions psychologically equivalent to one another, one would still expect to see

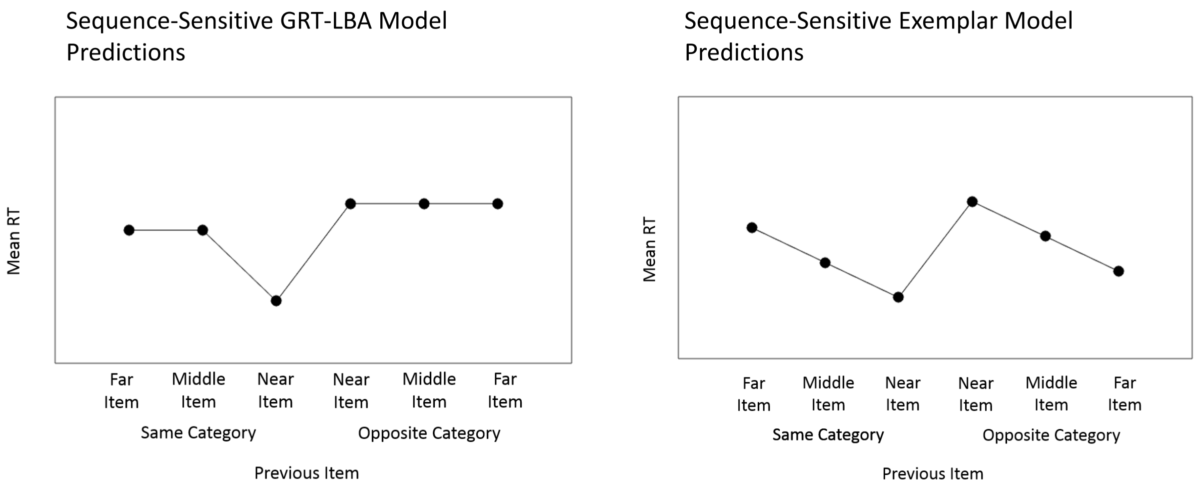

Figure 2. Illustrative predictions for near item performance as a function of the previous item for the sequence-sensitive GRT-LBA model (left) and the sequence-sensitive exemplar model (right).

repetition effects as observed in the integral *control* task. In the standard Garner task, one would also likely expect complex patterns of repetitions and alternations as demonstrated in several 2AFC tasks (Cho et al., 2002; Gökaydin, Navarro, Ma-Wyatt, & Perfors, 2016; Jentzsch & Sommer, 2002; Jones, Curran, Mozer, & Wilder, 2013; Soetens, Boer, & Hueting, 1985). In those tasks, there are two items, each associated with a separate response; the results are well-explained by assuming that people simultaneously learn about the base rates of each stimulus along with the rate of alternations (Jones et al., 2013). In our modified tasks, where there are more than two stimuli, we might expect participants to learn about the additional information provided by the ordinal relations between successive items. We would expect this to result in more complex pushing and pulling effects as observed in our modified control task with integral dimension as well as observed in unidimensional tasks involving simple auditory or simple visual stimuli (Stewart et al., 2002).

The question remains as to whether graded effects of similarity to the previous item will be observed with separable dimensions. The implication would be that participants are sensitive not only to the ordinal relations between stimuli but also to the metric distances between successive items. Previous investigations using separable dimensions have found that people are sensitive only to the ordinal difference between items but not the magnitude of the change from the previous to the current trial (Stewart et al., 2002). A critical distinction from the present work is that in Stewart et al. (2002) the values of their separable dimensions, semicircles varying in diameter and the angle of a radial line, were perfectly correlated with one other during the experiment. Consequently, either of the dimensions may have been used independently or redundant changes on both dimensions might have been used rendering the task effectively unidimensional. Our modified task provides a compelling test of the differences across tasks allowing for differential predictions of the sequence-sensitive exemplar and GRT-LBA models (see Figure 2).

We test these predictions using two types of separable dimension stimuli: saturation and brightness varied in two adjacent color patches (Experiments 1 and 3; see Figure 1, bottom left) and the saturation and inset line position of a rectangle (Experiments 2 and 4; see Figure 1, bottom right; hereafter termed *boxcars*). The former were confirmed as separable dimensions in Garner's (1974) initial studies[3]; the latter have been confirmed as evoking independent dimensional processing in a related task (Fifić et al., 2008; Little et al., 2011). To foreshadow our results, we observe data in all of our conditions consistent with the predictions of the exemplar-based model but not the GRT-LBA model.

In addition to the experiments, we present a hierarchical Bayesian implementation of the sequential exemplar model to the data from these four experiments and also to the integral dimension data reported in Little et al. (2016).[4] The posterior parameter distributions of these fits indicated that the parameters governing the sequential effects take on similar values for the integral and separable dimension data. Notably, the parameters governing the representational aspects of the stimuli differ across conditions for the integral

[3]We are not arguing that separability is isomorphic to spatial separation, though separation does facilitate selective attention in many circumstances. Spatially separated features may still be processed in a dependent or pooled manner (e.g., if the features belong to the same "feature map"; Blunden, Howe, & Little, 2020; A. Treisman & Gelade, 1980) . By contrast, spatially overlapped dimensions may be processed independently (Little, Nosofsky, & Denton, 2011; Moneer, Wang, & Little, 2016).

[4]Although we use a small-N design (Smith & Little, 2018) and collect a large number of trials for each participant, once we collapse the data by looking at the effect of the previous item, the amount of trials per relevant condition is reduced dramatically. The hierarchical Bayesian model allows us to pool across participants in a principled way. A second benefit of the Bayesian approach is that it allows us to account for uncertainty in our estimates of summary statistics and latent model parameters. Where there are strong qualitative contrasts between models - as between the exemplar and GRT-LBA model, we report maximum likelihood fits for efficiency.

dimensions but not the separable dimensions. The selective influence of these parameters allows the model to predict invariant RTs across tasks in the separable dimension data while simultaneously predicting a complex pattern of sequential effects.

Dimensions vs Exemplars

One potential challenge to the exemplar-based explanation of separable-dimension stimuli is evidence showing that some separable dimensions are not pooled into a single representation as assumed by exemplar models, and instead, independent decisions are made about each of the dimensions. Garner (1974) posited that the observed invariance for separable-dimension stimuli was consistent with the assumption that the dimensions of the separable stimulus were processed serially. There are now several compelling demonstrations using more modern designs showing that decisions about separable dimensions are indeed made in serial, or in some cases in parallel, in a manner inconsistent with the predictions of the exemplar-based model.

To explain, Fifić et al. (2008) tested whether stimulus dimensions were processed in serial, in parallel, or pooled (i.e., into a coactive coactive architecture) using Systems Factorial Technology (SFT; Townsend and Nozawa (1995); Townsend and Wenger (2004); reviewed in the recent volume by Little, Altieri, Fifić, and Yang (2017)). Fifić et al. (2008) showed that integral dimensions are consistent with pooling or coactivity, as assumed by the EBRW (Nosofsky & Palmeri, 1997a) and the GRT-DBT model (Ashby & Gott, 1988), but separable dimensions are not. This difference has now been demonstrated over a large number of dimensional types (Blunden et al., 2020; Blunden, Wang, Griffiths, & Little, 2015; Cheng, McCarthy, Wang, Palmeri, & Little, 2018; Little et al., 2011, 2013; Moneer et al., 2016). We refer the reader to Griffiths, Blunden, and Little (2017) for a review of these results. The take-home message is that whereas integral dimensions are pooled into a whole object representation, separable dimensions are processed independently in different channels. Consequently, in a unidimensional task using separable dimensions like our modified Garner task, only the relevant dimension may be processed at all.

Some ambiguity arises because in unidimensional tasks, where only a single dimension is relevant, like the Garner task and the modified Garner task, an exemplar model can provide a good account of the data simply by tuning its attention weight solely to the relevant dimension. On the other hand, selective attention may not only act spatially in changing the similarity between objects but also temporally such that dimensions with larger weights are processed first (Lamberts, 1995, 1998, 2000) and independently of other dimensions. To test this account in our modified Garner task, we also instantiated a new *feature-based* similarity model, which only processes the relevant dimension in each of the tasks. We fitted this model hierarchically to both the integral data of Little et al. (2016) and the separable data presented in the experiments below. We then used competitive model selection to determine whether the additional information provided by the irrelevant dimension in the sequential exemplar model provides any improvement in fit (offset by a penalty for the additional complexity of the model) relative to a model which only processes the relevant dimension. As shown in the work that follows, we find that the sequential exemplar model provides a better account of not only the integral dimension data from the modified Garner task but also the separable dimension data reported here.

General Method

Overview

We used the modified Garner paradigm developed by Little et al. (2016). Similar to the standard Garner paradigm, there were three tasks: control, correlated, and filtering. In each of these tasks, participants are required to make quick and accurate categorization judgments. In contrast to the standard Garner paradigm (Garner, 1974; Garner & Felfoldy, 1970), there are six levels of stimulus variation on the relevant dimension and two values of stimulus variation on the irrelevant dimension (see Figure 1).

We used a small number of participants, each of whom completed a large number of sessions (see Little & Smith, 2018; Smith & Little, 2018), each individual participant can be thought of as an individual replication of the experiment. The extent to which the results of each participant agree then provides an indication of the stability of the result.

Human testing was approved by the Melbourne School of Psychology HEAG (Approval Number: 1034866).

Participants

Experiment 1: Brightness and Saturation in Different Locations. Eight participants from the University of Melbourne were randomly assigned to either the brightness (N = 4) or saturation (N = 4) conditions.

Experiment 2: Saturation and Line Position. Experiment 2 was conducted after Experiment 1. Eight new participants from the University of Melbourne were randomly assigned to either the saturation (N = 4) or line-position (N = 4) conditions. All participants reported normal or corrected to normal visual acuity and color vision. Participants received \$12 plus an additional \$3 bonus for accurate performance ($\geq$ 95%) for each session.

Stimuli

Experiment 1: Brightness and saturation in different locations. The stimuli were pairs of color squares (100 $\times$ 100 pixels) varying in brightness (value) and saturation (chroma). Each pair of color squares was presented within a rectangular frame. Each color was selected from the Munsell hue 5R. For each participant, the relevant dimension was initially set to either the right or left square for all six sessions, while the irrelevant dimension was varied by changing the value of the other square. For the brightness condition, the full set of stimuli was created by combining six levels of brightness (Munsell values 3, 4, 5, 6, 7, 8) varying in one square (with a fixed Munsell chroma of 4) and two levels of saturation (Munsell chroma 6, 8) varying in the other square (with a fixed Munsell brightness of 6).

For the saturation condition, the full set of stimuli was created by combining six levels of saturation (Munsell chroma 4, 6, 8, 10, 12, 14) varying in one square (with a fixed Munsell value of 6) and two levels of brightness (Munsell value 5, 6) varying in the other square (with a fixed Munsell saturation of 4). The increments in value for brightness and chroma for saturation were chosen such that they would be equally discriminable based on a psychological scaling study by Nickerson (1936), showing that one unit of value is equivalent to two units of chroma in discriminability. This assumption is supported by multidimensional scaling presented in Little et al. (2016). The stimuli were presented on a monitor resolution of 1280 $\times$ 1024, subtended a visual angle of about 4.70 degrees. All stimuli were centrally presented on a calibrated CRT monitor.

Experiment 2: Saturation and line position. The boxcar stimuli were color rectangles (170 $\times$ 255 pixels) with a line (100 $\times$ 10 pixels) positioned along the base of the rectangle. The color was chosen from Munsell hue 5R with fixed brightness (Munsell value = 5) but varying saturation. The line varied in its horizontal position along the base of the rectangle in increments of 20 pixels from the left side of the rectangle. In the saturation condition, the full set of stimuli was created by combining six levels of saturation (Munsell chroma = 4, 6, 8, 10, 12, 14) and two line positions (60, 80 pixels from the left). In the line-position condition, the full set of stimuli was created by combining six line positions (20, 40, 60, 80, 100, 120 pixels from the left) and two levels of saturation (Munsell chroma = 8, 10). The stimuli were presented on a monitor resolution of 1280 $\times$ 1024, subtended at a visual angle of about 4.70 degrees.

For both experiments, the xyY coordinates corresponding to the Munsell value and chroma were converted to RGB values by first converting them to CIE XYZ color-space coordinates using standard transformations (Travis, 1991).

Procedure

Participants completed six one-hour categorization sessions. The first two sessions were always scheduled on the same day or on consecutive days. In each session, participants completed all three tasks. In each block, save the second filtering task block, there were 24 practice trials and 120 experimental trials. There were two different versions of the control (each utilizing one of the two values of the irrelevant dimension) and correlated tasks (each using different combinations of the relevant and irrelevant dimensions).

The two versions of the control task were presented in consecutive blocks, as were the two versions of the correlated task. As the filtering task consisted of 12 stimuli, the filtering task was repeated over two consecutive blocks to ensure an equal number of repetitions of each stimulus across tasks. Due to a bug in the experimental code, only three conditional orderings were presented: control-correlated-filtering, control-filtering-correlated, and correlated-control-filtering. (This is corrected in the subsequent Experiments 3 and 4.) The order of presentation of the individual stimuli was randomized within each block.

At the beginning of each session, participants were instructed to read the instructions displayed on the monitor. The presented instructions can be found in Appendix A.

On each trial, a fixation cross was presented for 1500 ms, followed by the stimulus. The participant had to decide whether the stimulus belonged to category A or B. Response choice and RT were recorded via button press of a calibrated RT box (Li, Liang, Kleiner, & Lu, 2010). The stimulus remained on screen until a button press was made. For the first session, full feedback was provided (i.e., “correct” or “wrong”) for each response. In subsequent sessions, feedback was only provided for incorrect responses. If a response was not made before 5000 ms, the feedback “too slow” was shown. Feedback remained on screen for 2000 ms. Participants were instructed to categorize each stimulus into either Category A or Category B as accurately and quickly as possible.

Data Analysis

All of the analyses and exclusion criteria were chosen prior to data analysis. The entire first session was considered practice and excluded from further analysis. Trials with RTs greater than 3000 ms or less than 200 ms were also excluded. Less than 5% of the total data was excluded.

For comparison to previous reports of the standard Garner task, we report analyses of overall mean RTs and item-averaged mean RTs for each condition and for each subject using Bayesian analyses implemented in JAGS (Plummer, 2003). Details can be found in Appendix B.

To analyze the sequential effects, RTs and accuracy data were averaged across subjects and plotted in each condition and task as a function of the immediately preceding item. For a parsimonious representation of sequential effects across individuals and to increase the power for each item condition, we employed a hierarchical Bayesian framework. We evaluated these effects using hierarchical Bayesian linear regression conducted using R (3.4.2, R Core Team, 2017) with the brms package (1.10.0, Bürkner, 2017) which uses the No U-Turn Sampling algorithm implemented in rstan (2.16.2, Rstan Development Team, 2017).

Results

Experiment 1 & 2: Average RT for each experimental condition

To assess whether we reproduced the expected invariance across tasks with separable dimensions, we computed the average correct RT for each subject in each of the control, correlated, and filtering tasks. We then subtracted these average RTs from the overall RT for each subject so that each subject provides a deviation for each condition. This ensures that the results are not obscured by potential differences in the overall average RT and improves the convergence of the MCMC sampler (J. Kruschke, 2014).

The expectation is that if RT is invariant across tasks then these deviations should be near zero. For comparison, we also present the same analysis for the integral data from Little et al. (2016). The average RTs from integral-dimension stimuli are expected to show a control deviation near zero, a correlated deviation less than zero, and a filtering deviation greater than zero, reflecting the expected ordering of the tasks arising from correlated facilitation and filtering interference. The results are shown in Figure 3. The posteriors of the deviations in each condition show that while the integral conditions have the expected ordering of tasks in each condition, the separable conditions are invariant in Experiment 1 (to the extent that the 99% HDI's overlap), though we note that in the saturation condition of Experiment 1, the ordering of the RTs has pattern similar to the integral dimension data. In Experiment 2, the saturation condition shows the expected invariance, but in the line position condition, the correlated condition has longer RTs than the control or filtering condition. We next examine this at the level of individual participants.

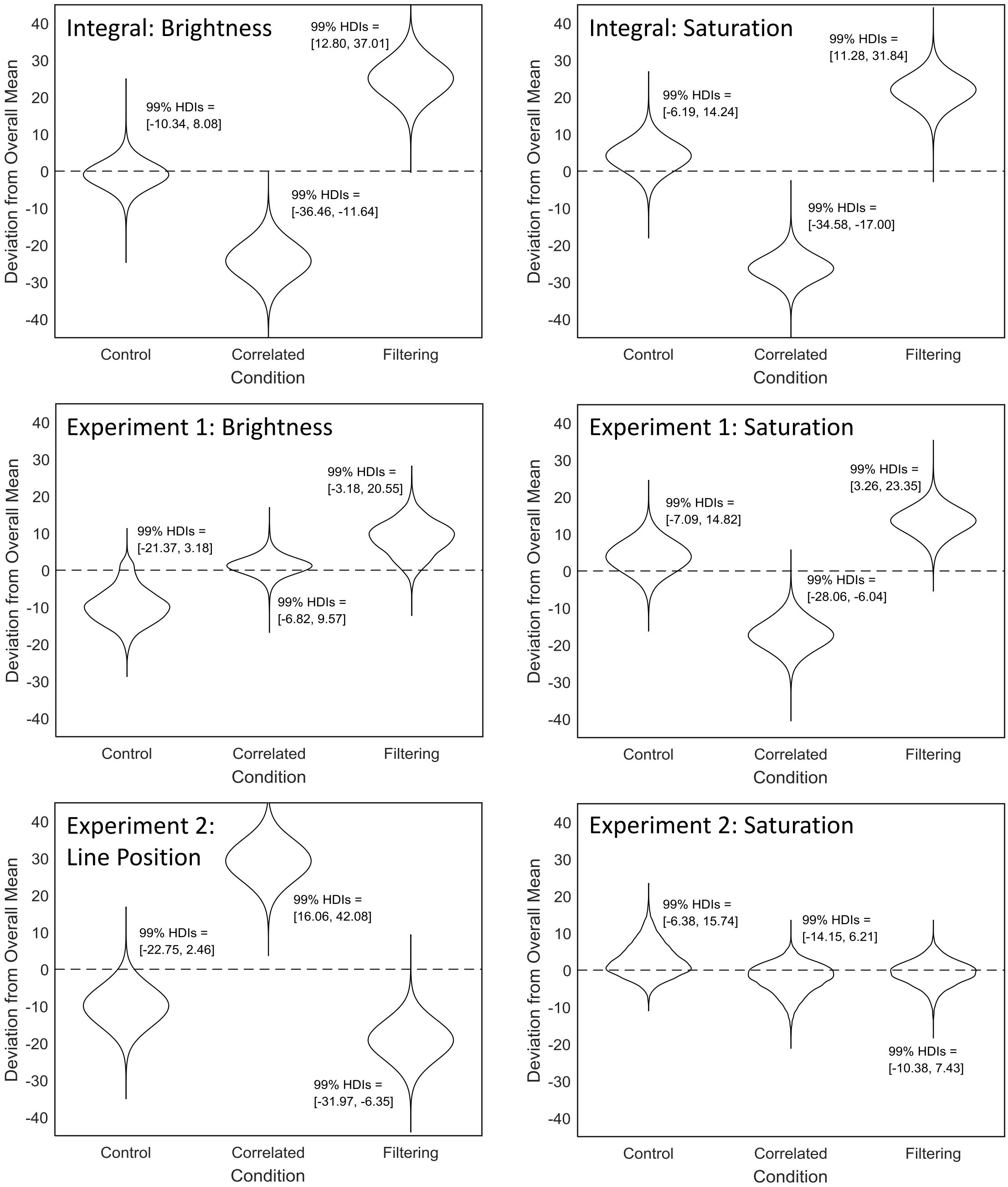

Figure 3. Posterior deviations from the overall average RT across the control, correlated, and filtering conditions in the integral Brightness (N = 4, Top Left) and integral Saturation (N = 4, Top Right) conditions from Little, Wang & Nosofsky (2016), Experiment 1: separable color Brightness (N = 4, Middle Left) and separable color Saturation (N = 4, Middle Right), and Experiment 2: separable boxcar Line Position (N = 4, Bottom Left) and separable boxcar Saturation (N = 4, Bottom Right).

Experiment 1 & 2: Individual average RT analyses

To estimate the mean RTs in each condition for each subject, we again applied a Bayesian analysis using JAGS (see Appendix B). The posterior estimates of the average RTs for each item condition for each subject are shown in Table 1. Most of the subjects have near items which have larger RTs than the middle and far items as expected.

There are some individual differences in the ordering of the mean RTs for the near items across tasks. Looking at the 99% HDI's for the near items, all of the participants in the brightness condition of Experiment 1, save one, and the saturation condition of Experiment 2 show invariant RTs across the three tasks. That is, the 99% HDI's overlap substantially allowing for an inference that the means of these tasks are effectively equal. Two participants in the saturation condition of Experiment 1 and three participants in the line position condition of Experiment 2 have at least one task which varies substantially from the control condition. In most cases, it appears the correlated condition is either faster than the control condition (Experiment 1, S2 and S3; Experiment 2, L2) or slower than the control condition (Experiment 2, L1 and L3). In two cases, the filtering condition is also slower than the control condition (Experiment 1, B1; Experiment 2, L1).

Interim Summary - Average RT Results for Experiments 1 and 2

The analysis of the overall average RTs provided evidence that the stimulus dimensions used in our experiments were treated as separable. However, we found deviations from invariance for some participants across the control, correlated, and filtering conditions. One argument against integrality is that these differences were idiosyncratic. We suspect that these differences arose because participants were not instructed to attend to the relevant stimulus dimension as is common in applications of the standard Garner task. This would have allowed for individual difference to arise in how attention was allocated to different stimulus dimensions. For example, in the correlated condition, where both dimensions are relevant, differing allocations of attention could have resulted in RT differences between observers. To address this, we replicated Experiments 1 and 2 using the same modified

Table 1
Estimates of the Posterior Mean RTs (and 99% Highest Density Intervals) for each task, item, and subject from Experiments 1 and 2.

	Experiment 1 Brightness				Experiment 2 Line Position		
	Near	Middle	Far		Near	Middle	Far
Subject B1				Subject L1			
Control	530 (511, 549)	421 (410, 434)	408 (398, 419)	Control	689 (649, 732)	492 (472, 513)	463 (447, 479)
Correlated	556 (536, 576)	445 (434, 459)	420 (409, 432)	Correlated	805 (758, 859)	614 (585, 643)	562 (534, 593)
Filtering	596 (572, 621)	451 (438, 465)	426 (416, 438)	Filtering	803 (755, 858)	620 (589, 659)	593 (563, 627)
Subject B2				Subject L2			
Control	504 (481, 528)	382 (372, 393)	370 (362, 379)	Control	887 (836, 935)	605 (584, 631)	552 (529, 574)
Correlated	507 (483, 531)	387 (375, 398)	363 (354, 371)	Correlated	774 (725, 822)	579 (554, 608)	517 (500, 536)
Filtering	497 (479, 518)	370 (361, 379)	366 (358, 374)	Filtering	882 (837, 933)	598 (575, 621)	521 (504, 537)
Subject B3				Subject L3			
Control	503 (481, 531)	444 (426, 462)	424 (405, 443)	Control	1087 (1006, 1184)	909 (839, 983)	864 (801, 937)
Correlated	542 (513, 572)	452 (435, 474)	440 (423, 461)	Correlated	1208 (1103, 1344)	1084 (992, 1180)	959 (886, 1042)
Filtering	548 (521, 580)	448 (429, 470)	459 (438, 482)	Filtering	952 (885, 1031)	704 (666, 749)	617 (589, 649)
Subject B4				Subject L4			
Control	610 (582, 642)	480 (464, 495)	489 (473, 506)	Control	866 (820, 917)	619 (592, 652)	574 (547, 601)
Correlated	625 (596, 655)	502 (486, 519)	484 (469, 500)	Correlated	831 (782, 874)	608 (578, 640)	574 (547, 602)
Filtering	651 (620, 687)	518 (501, 535)	502 (485, 518)	Filtering	869 (821, 929)	720 (675, 768)	616 (583, 656)
	Experiment 1 Saturation				Experiment 2 Saturation		
	Near	Middle	Far		Near	Middle	Far
Subject S1				Subject S1			
Control	629 (592, 666)	462 (445, 479)	450 (434, 468)	Control	488 (473, 504)	426 (417, 435)	413 (405, 423)
Correlated	623 (589, 655)	476 (460, 493)	445 (431, 460)	Correlated	501 (484, 519)	441 (431, 451)	419 (410, 429)
Filtering	665 (626, 702)	496 (479, 516)	487 (467, 508)	Filtering	488 (474, 503)	428 (419, 438)	426 (417, 435)
Subject S2				Subject S2			
Control	749 (703, 799)	519 (501, 537)	471 (458, 483)	Control	724 (687, 763)	554 (533, 575)	513 (495, 532)
Correlated	668 (635, 713)	477 (463, 493)	456 (442, 470)	Correlated	658 (628, 691)	508 (492, 525)	489 (475, 502)
Filtering	775 (725, 835)	492 (475, 510)	467 (455, 481)	Filtering	672 (645, 700)	524 (509, 540)	489 (476, 503)
Subject S3				Subject S3			
Control	955 (896, 1020)	549 (531, 568)	523 (509, 539)	Control	755 (719, 796)	534 (513, 554)	500 (484, 517)
Correlated	824 (774, 882)	541 (523, 559)	550 (533, 567)	Correlated	756 (716, 803)	519 (504, 537)	505 (488, 523)
Filtering	927 (873, 982)	572 (553, 593)	528 (511, 545)	Filtering	730 (692, 770)	506 (490, 522)	487 (472, 502)
Subject S4				Subject S4			
Control	604 (571, 638)	422 (410, 436)	411 (399, 423)	Control	816 (772, 877)	490 (475, 505)	452 (440, 464)
Correlated	560 (529, 593)	419 (406, 433)	409 (398, 421)	Correlated	769 (722, 823)	488 (474, 503)	451 (441, 463)
Filtering	609 (575, 647)	418 (404, 430)	423 (407, 438)	Filtering	813 (752, 869)	527 (507, 546)	466 (454, 479)

Garner design and stimuli but with explicit instructions to attend to a single relevant dimension. Before turning to the sequential effects, we first report the average results of these experiments to alleviate lingering concerns about the separability of our stimulus dimensions.

Experiments 3 & 4: Attentional Instructions

Participants

Experiment 3: Brightness and Saturation in Different Locations. Six new participants from the University of Melbourne were randomly assigned to either the brightness (N = 3) or saturation (N = 3) conditions.

Experiment 4: Saturation and Line Position. Six new participants from the University of Melbourne were randomly assigned to either the saturation (N = 3) or line-position (N = 3) conditions. All participants reported normal or corrected to normal visual acuity and color vision. Participants received \$12 plus an additional \$3 bonus for accurate performance ($\geq$ 95%) for each session.

Stimuli, Design, Procedure & Data Analysis

All aspects of Experiments 3 and 4 were identical to Experiments 1 and 2, respectively, with two exceptions. First, we corrected the task counterbalancing across sessions. Second, at the beginning of each session, participants received explicit instructions to attend to the relevant stimulus dimension for their condition. For instance, in Experiment 3, the following instructions were added: "You should only pay attention to the colour on the left. The colour on the right will vary but you should ignore changes to this colour." Analogous instructions were used if the relevant dimension was located on the right. In Experiment 4, the additional instruction read: "You should only pay attention to the position of the line. The colour will vary but you should ignore changes to the colour." Again, analogous instructions were used when saturation was the relevant dimension.

Experiment 3 & 4: Average RT for each experimental condition

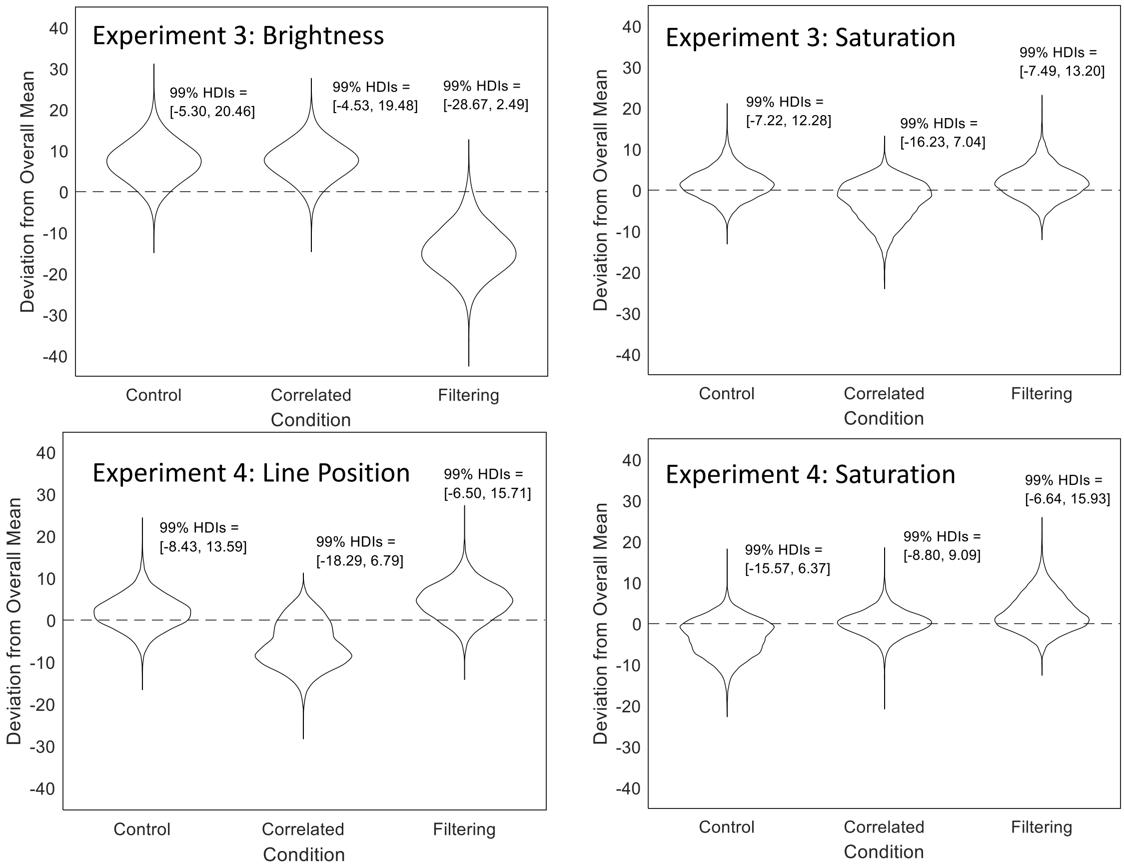

Figure 4. Posterior deviations from the overall average RT across the control, correlated, and filtering conditions from Experiment 3: separable color Brightness (N = 4, Middle Left) and separable color Saturation (N = 4, Middle Right), and Experiment 4: separable boxcar Line Position (N = 4, Bottom Left) and separable boxcar Saturation (N = 4, Bottom Right).

The average RTs from each condition in Experiments 3 and 4 show the expected invariance across tasks (see Figure 4).

Experiment 3 & 4: Individual average RT analyses

The posterior estimates of of the average RTs for each item condition for each subject are shown in Table 2. For most of the participants, the near item RTs were invariant across tasks. The sole exception was Subject BI2 in the Experiment 3 Brightness Relevant condition, who had shorter RTs in the filtering condition than the control condition.

Table 2

Estimates of the Posterior Mean RTs (and 99% Highest Density Intervals) for each task, item, and subject from Experiments 3 and 4.

	Experiment 3 Brightness				Experiment 4 Line Position		
	Near	Middle	Far		Near	Middle	Far
Subject BI1				Subject LI1			
Control	473 (459, 490)	388 (380, 397)	380 (372, 388)	Control	590 (563, 623)	479 (458, 502)	435 (418, 454)
Correlated	472 (458, 488)	396 (387, 405)	381 (371, 390)	Correlated	557 (535, 585)	457 (439, 475)	419 (407, 434)
Filtering	475 (458, 492)	399 (389, 408)	385 (376, 395)	Filtering	589 (563, 616)	469 (450, 490)	453 (436, 476)
Subject BI2				Subject LI2			
Control	962 (916, 1016)	729 (698, 764)	702 (672, 733)	Control	487 (468, 509)	393 (383, 404)	374 (367, 383)
Correlated	920 (866, 975)	714 (674, 756)	714 (676, 752)	Correlated	498 (478, 520)	394 (384, 406)	375 (367, 384)
Filtering	867 (826, 911)	704 (673, 738)	660 (631, 687)	Filtering	516 (496, 541)	411 (399, 422)	391 (382, 400)
Subject BI3				Subject LI3			
Control	640 (606, 676)	509 (491, 530)	491 (475, 510)	Control	542 (511, 578)	383 (373, 396)	353 (344, 361)
Correlated	663 (628, 700)	522 (501, 544)	497 (477, 517)	Correlated	491 (464, 520)	375 (365, 386)	351 (343, 359)
Filtering	612 (582, 642)	480 (465, 497)	471 (454, 487)	Filtering	508 (480, 539)	383 (372, 394)	353 (344, 361)
	Experiment 3 Saturation				Experiment 4 Saturation		
	Near	Middle	Far		Near	Middle	Far
Subject SI1				Subject SI1			
Control	633 (600, 664)	497 (481, 514)	476 (460, 491)	Control	360 (350, 371)	317 (311, 322)	310 (304, 316)
Correlated	584 (557, 614)	482 (464, 498)	454 (440, 468)	Correlated	369 (359, 379)	329 (322, 336)	324 (318, 330)
Filtering	639 (607, 672)	489 (472, 506)	478 (459, 497)	Filtering	357 (348, 367)	325 (318, 333)	315 (310, 321)
Subject SI2				Subject SI2			
Control	485 (461, 513)	395 (384, 406)	381 (373, 390)	Control	474 (455, 493)	418 (406, 429)	401 (391, 414)
Correlated	476 (457, 499)	399 (389, 411)	384 (375, 396)	Correlated	474 (453, 495)	411 (397, 424)	408 (394, 421)
Filtering	521 (492, 550)	405 (392, 420)	389 (378, 401)	Filtering	496 (474, 523)	418 (406, 430)	408 (396, 420)
Subject SI3				Subject SI3			
Control	488 (470, 506)	469 (454, 487)	470 (456, 487)	Control	518 (490, 545)	396 (385, 408)	374 (365, 383)
Correlated	486 (468, 502)	459 (445, 472)	458 (443, 472)	Correlated	532 (506, 559)	409 (397, 422)	384 (373, 394)
Filtering	474 (457, 492)	461 (445, 479)	455 (440, 471)	Filtering	565 (532, 595)	422 (409, 437)	397 (386, 408)

Interim Summary - Average RT Results for Experiments 3 and 4

The average RTs for each experimental condition and for each individual (save one) were invariant across the modified Garner tasks as expected for separable dimension stimuli. Evidently, instructing participants to pay attention to the relevant dimensions eliminated any variability in the mean RT across tasks for each individual. This provides further confirmation that our stimulus dimensions were indeed separable. It is important to emphasize that in averaging across the previous item, this analysis obscures the systematic sequential effects which are the main focus of the paper.

Sequential Effects

Sequential Effects - Experiment 1: Brightness and Saturation in Adjacent Colors

RT and accuracy data as a function of the immediately preceding item were averaged across subjects, separately for the brightness condition and saturation condition, across logically equivalent category A and category B items (see Figure 5). For each task, the three lines refer to whether the current item was a Near item (to the boundary), Middle item, or Far item, represented by circles, diamonds, and squares, respectively. For the filtering condition, an additional line indicates whether the immediately preceding Near item had a different irrelevant dimension value (filled circles)[5]. Similar patterns of sequential effects as found for integral-dimension stimuli in Little et al. (2016) appear to arise for separable dimension stimuli in the present experiment.

[5]The effect of changing the irrelevant dimension on the Middle and Far items was minimal so these trials are averaged with trials in which the irrelevant dimension was the same. These are the lines shown in Figure 5.

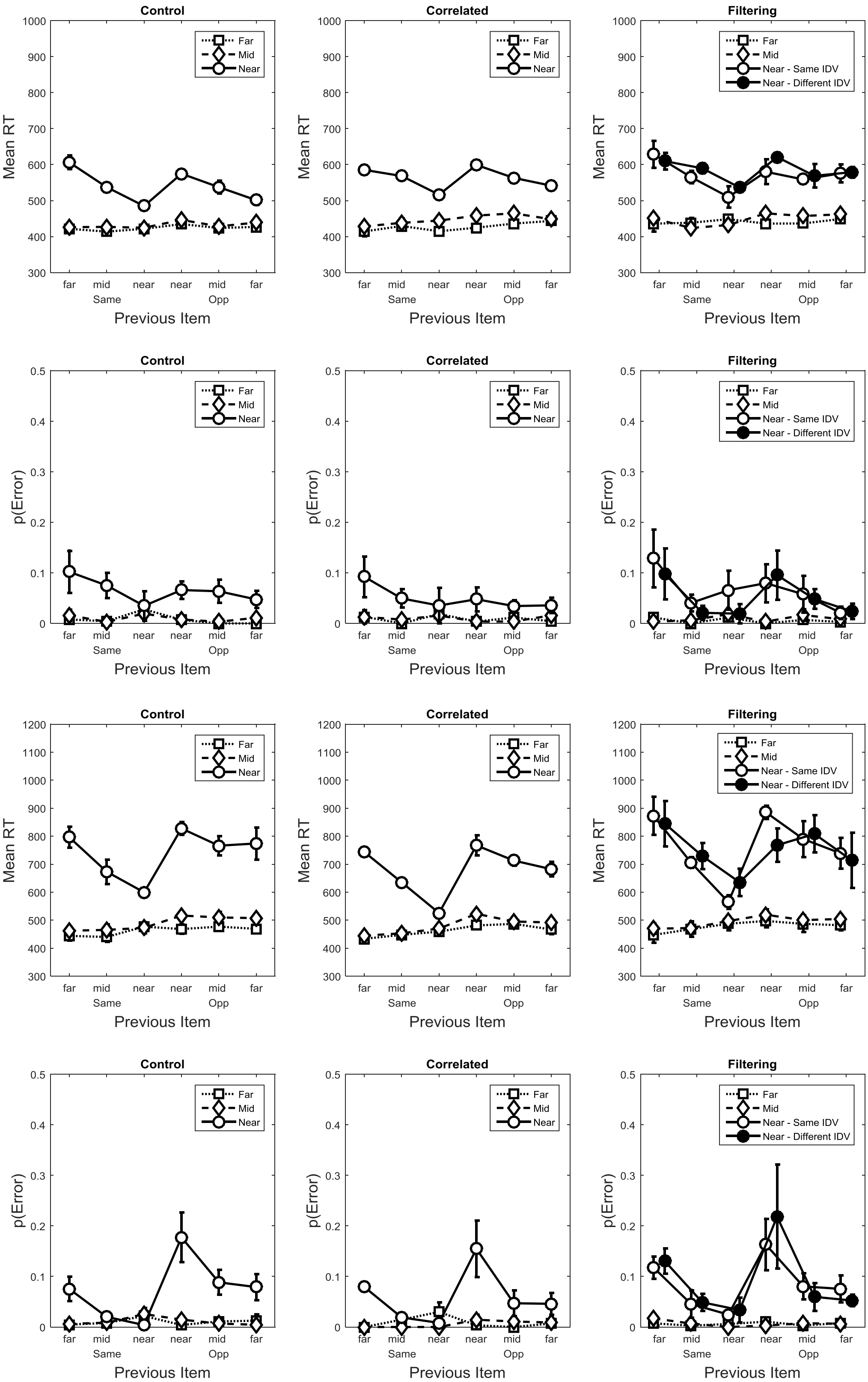

Figure 5. Experiment 1: Mean RT and error rates as function of the immediately preceding item in the (a) Brightness Group (top two rows) and (b) Saturation Group (bottom two rows). Error bars are one-standard error.

To analyze the sequential effects, we used a hierarchical Bayesian linear regression. The posterior estimate of the coefficient on the group effect can be used to determine the size of the sequential effect. This analysis is detailed in Appendix C. We report Bayes Factors (BF) indicating the support for the effect vs a null effect. A BF less than 1 can be interpreted as support for the null hypothesis; a BF greater than 1 indicates support for the alternative hypothesis. The magnitude of the BF can be interpreted using the evidence categories provided by Jeffreys (1961): $1 < BF < 3$ = Anecdotal; $3 < BF < 10$ = Moderate; $10 < BF < 30$ = Strong; $30 < BF < 100$ = Very Strong; $BF > 100$ = Extreme.

Across all tasks and in both the Brightness and Saturation conditions, there was moderate to extreme evidence for the repetition effect, pushing effect, and pulling effect in the RTs (see the Mean RT - BF_{dir} column in Table C1. There was only anecdotal evidence for the pulling effect in the filtering task in the Brightness condition (see the Filtering - Adj Opp row, $BF_{dir} = 1.42$). We infer that the sequential effects were all confirmed to occur in the predicted direction.

One important prediction which follows from the use of separable-dimension stimuli is that there should be no difference between the effects in the filtering task when the irrelevant dimension changes from the previous to the current trial. Indeed, there was strong support for the null hypothesis of no difference between the repetition, the pushing, and the pulling effects (see the Mean RT - BF_{10} column in Table C1 in the Irrelevant Dimension Change rows). This null effect supports the idea that for separable dimension there is little effect of varying the irrelevant dimension.

The sequential effects were also found in the predicted directions for accuracy for most of but not all of the tasks. For instance, in the Brightness condition, although there was very strong evidence for the pushing effect in all three tasks and moderate-to-strong evidence for the repetition effect on accuracy in the control and correlated tasks, there was anecdotal evidence against the repetition effect in the filtering task ($BF_{dir} = 0.94$). For the pulling effect, there was only weak evidence for the effect in the filtering task ($BF_{dir} = 2.53$), but weak evidence against the effect in the control and correlated tasks ($BF_{dir} = 0.70$ and

$BF_{dir} = 0.74$, respectively). Evidently, the RTs are more sensitive to the sequential effects than error rates in the Brightness condition, perhaps due to a floor effect on the error rate.

For the Saturation condition, there was very strong evidence for the repetition and pulling effects on accuracy in all tasks (see the Error Rate - BF_{dir} columns in Table C1). There was also moderate evidence for the pushing effect but only in the correlated and filtering tasks. There was anecdotal evidence supporting the null pushing effect in the Saturation control task.

For accuracy, there was also support for a null difference in the sequential effects when the irrelevant dimension changed in the filtering task in both the Brightness and Saturation conditions (see Error Rate - BF_{10} column). The sole exception to this was the repetition difference in the Brightness condition. Here, there was anecdotal evidence for a negative difference such that the repetition effect error rate was lower when the irrelevant dimension changed from the previous trial to the current trial compared to when the irrelevant dimension did not change ($BF_{10} = 1.27$).

Sequential Effects - Experiment 2: Saturation and Line Position

The results are shown in Figure 6. As in Experiment 1, the sequential effects were similar to the integral dimension results from Little et al. (2016).

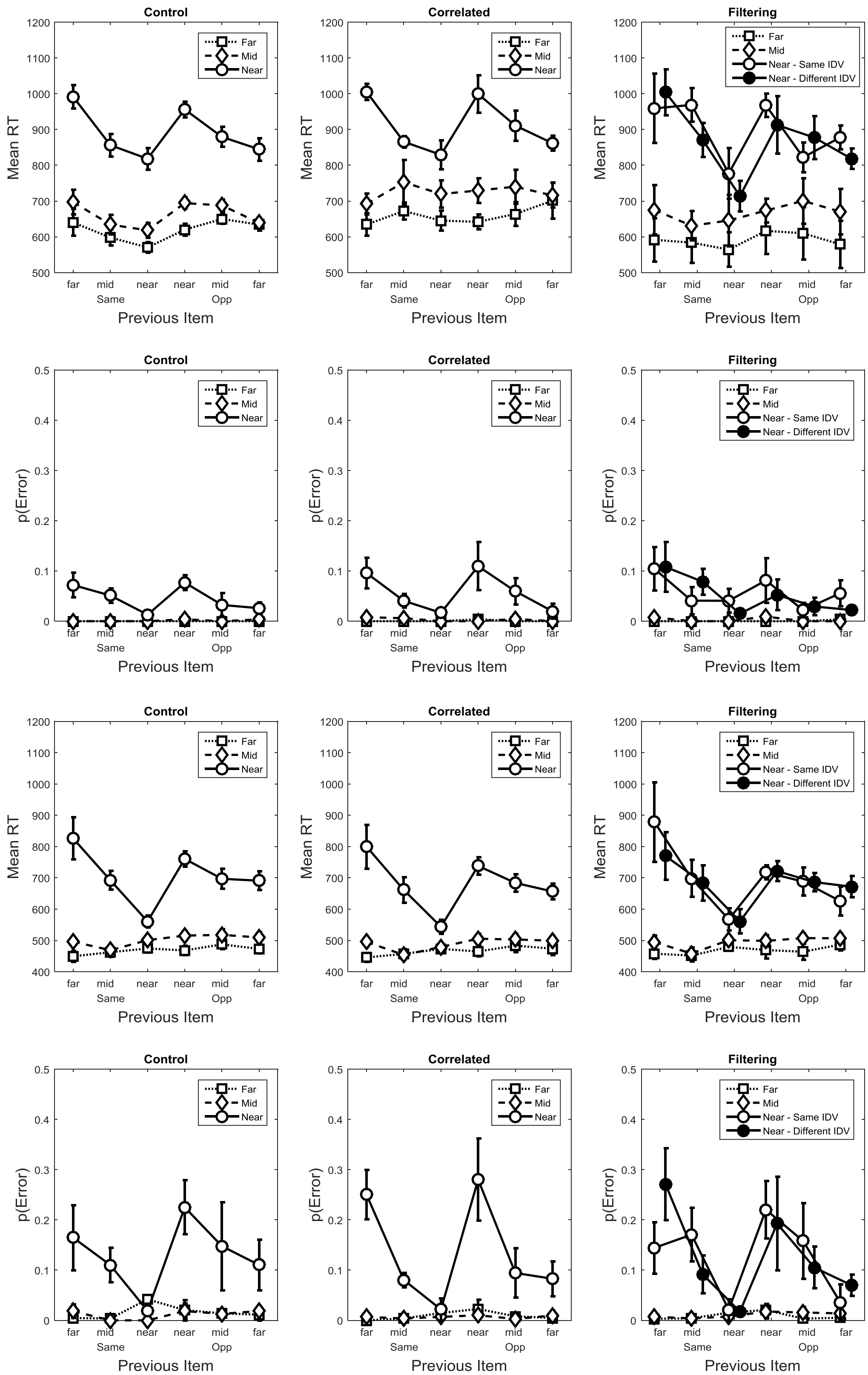

Figure 6. Experiment 2: Mean RT and error rates as function of the immediately preceding item in the (a) Line Position Group (top two rows) and (b) Saturation Group (bottom two rows). Error bars are one-standard error.

Looking at the BF_{dir} in Table C2, as in Experiment 1, there was moderate to extreme evidence for the repetition effect, the pushing effect, and the pulling effect on the RTs in both conditions. There was also very strong to extreme evidence for the repetition and pulling effects on accuracy in all conditions. There was strong evidence for the pushing effect on accuracy only in the correlated condition. In the control and filtering conditions, the evidence for pushing effect was either only anecdotal (in the control condition) or in favor of the null (in the filtering condition). The Bayes Factors favored the null hypothesis of no difference between the effects in the filtering condition when the irrelevant dimension changes from the previous to the current trial. This was also true for accuracy, except for the difference in the pushing effect; here, there was anecdotal evidence for a difference between the size of the pushing effect when the irrelevant dimension changed ($BF_{10} = 4.00$).

Sequential Effects - Experiment 3: Brightness and Saturation

The sequential effects in the Mean RT and error rates for Experiment 3 are shown in Figure 7. Recall that in this experiment, participants were instructed to attend to the relevant dimension, either brightness or saturation, when varied in two locations.

For the Mean RTs, in both the Brightness and Saturation conditions, there was Strong to Extreme evidence for a repetition effect in all three tasks. There was mixed evidence for the far same pushing effect and near opposite pulling effect across the Brightness and Saturation conditions, likely resulting from the relatively faster responding across all sequential pairs in the Saturation condition compared to the Brightness conditions (see Figure 7). But importantly, there was evidence of no difference across levels of the irrelevant dimensions change across both the Brightness and Saturation conditions.

More specifically for the pushing and pulling effects, for the Brightness condition, there was Extreme evidence for a pushing effect in the Control and Correlated tasks but only anecdotal evidence for a pushing effect in the Filtering condition. In the Saturation condition, there was evidence in favour of a null pushing effect across all tasks. For the opposite category adjacent-item pulling effect, there was strong evidence in the Brightness

Filtering condition but evidence in favor of a null effect in the Brightness Control and Correlated conditions. However, for the Saturation conditions, there was Strong to Very Strong evidence of a pulling effect in the Control and Correlated conditions but evidence in favor of a null pulling effect in the Filtering condition. For the error rate data, there was Moderate to Extreme evidence of a repetition effect and near opposite pulling effect across all tasks in both the Brightness and Saturation conditions. There was moderate evidence for a pushing effect in Brightness Control and Correlated conditions, but in the remaining conditions, the evidence was anecdotal or favoured the null. For most conditions, there was evidence in favour of no difference between change in the irrelevant dimension. The sole exception was anecdotal evidence that the pushing effect was stronger when the irrelevant dimension changed in the Saturation Filtering condition.

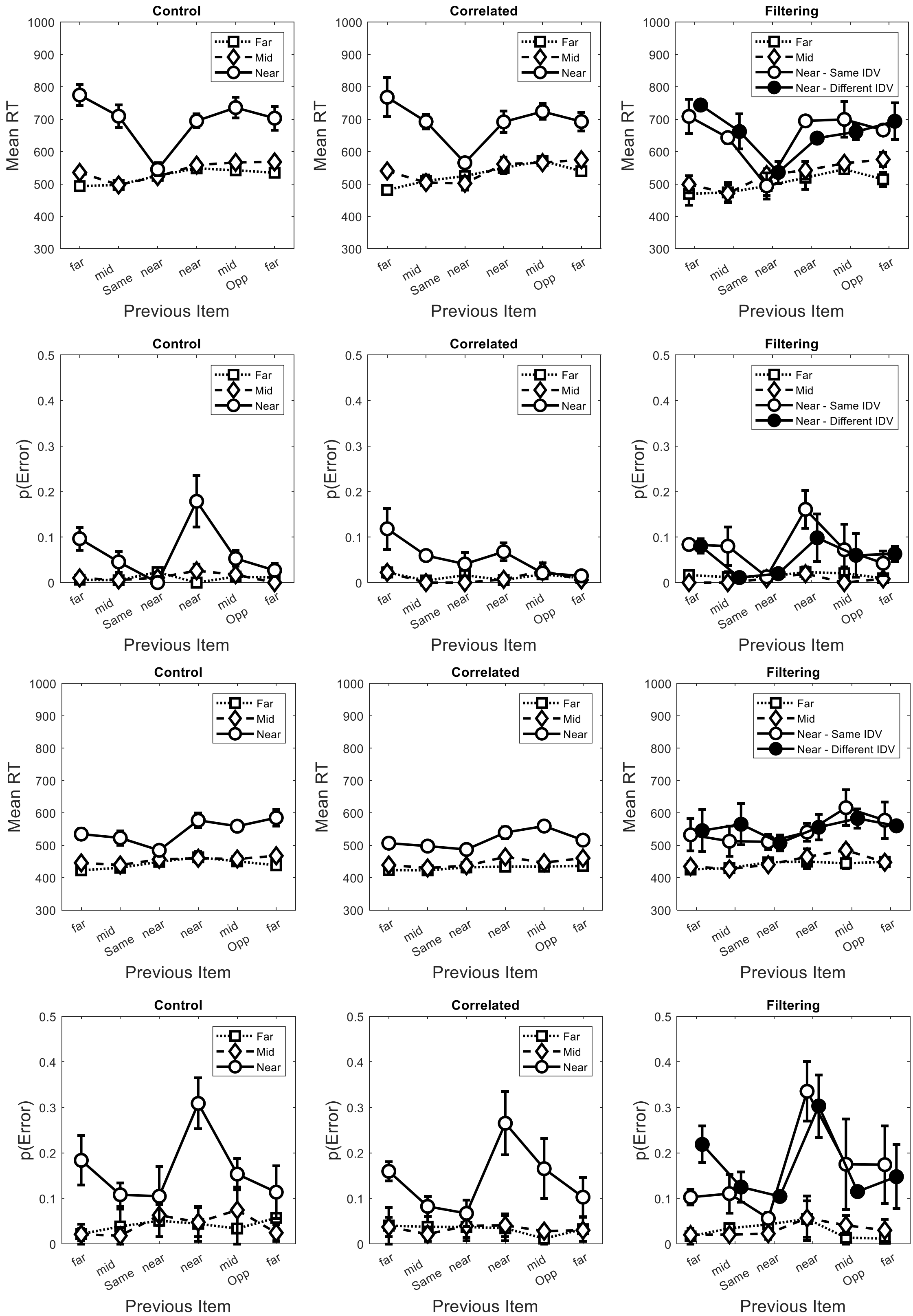

Figure 7. Experiment 3: Mean RT and error rates as function of the immediately preceding item in the (a) Attention Instructions Brightness Group (top two rows) and (b) Attention Instructions Saturation Group (bottom two rows). Error bars are one-standard error.

Sequential Effects - Experiment 4: Saturation and Line Position

The sequential effects in the mean RT and error rates for Experiment 4 are shown in Figure 8. In this experiment, participants were instructed to attend to the relevant dimension, either line position or saturation. For all conditions and tasks across both Mean RTs and error rates, there was moderate to extreme evidence in favour of the repetition, pushing, and pulling effects. Likewise for both mean RT and error rates, across both conditions, most effects had evidence in favour of a null difference when the irrelevant dimension varied in the filtering task. There was some evidence in favour of a weaker repetition effect and stronger pulling effect on accuracy when the irrelevant dimension changed in the Saturation Filtering condition; however, this evidence was only anecdotal.

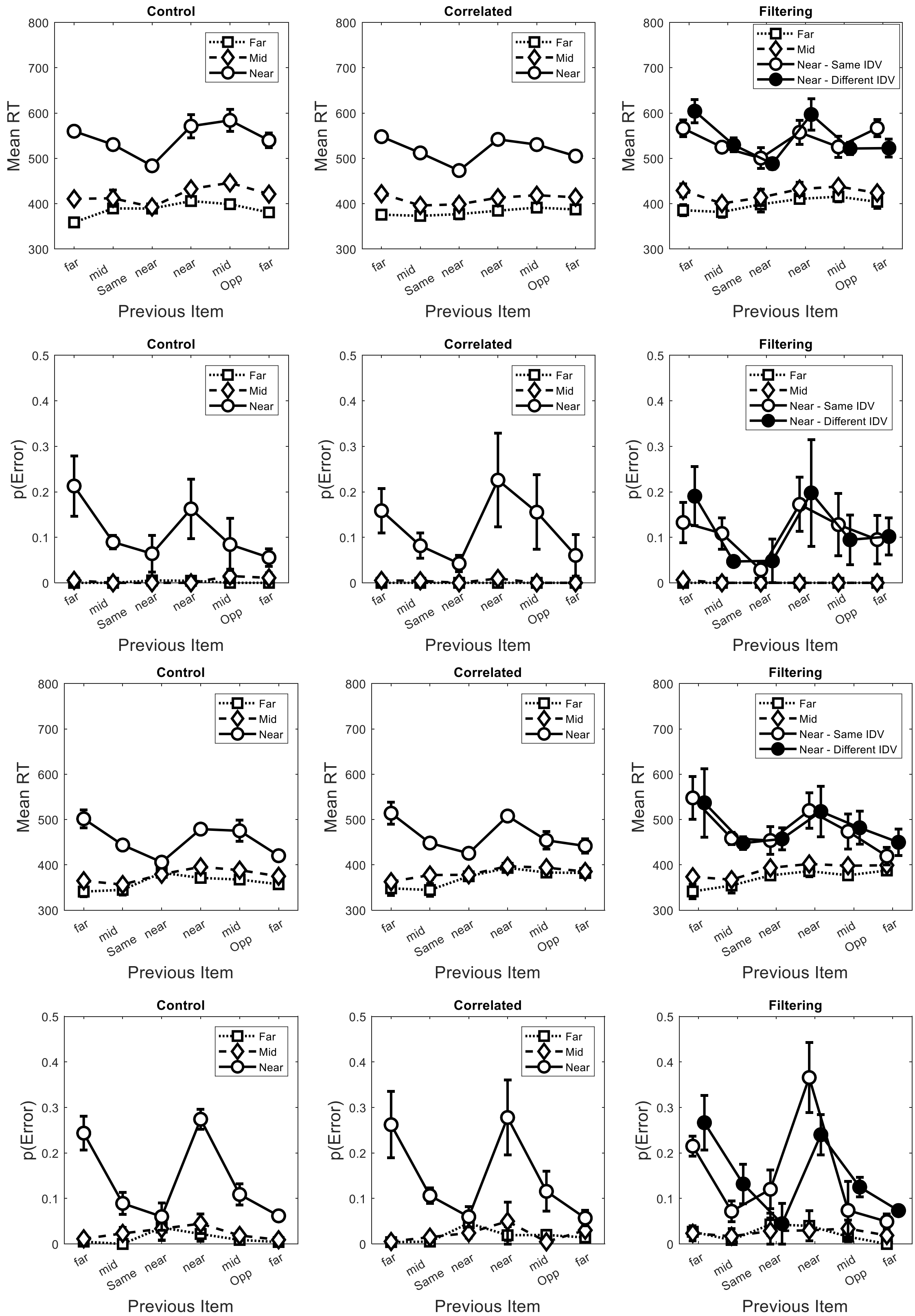

Figure 8. Experiment 4: Mean RT and error rates as function of the immediately preceding item in the (a) Attention Instructions Line Position Group (top two rows) and (b) Attention Instructions Saturation Group (bottom two rows). Error bars are one-standard error.

Interim Summary - Sequential Effects

The analysis of the sequential effects confirmed that, for the most part, we replicated the repetition, pushing, and pulling effects with two types of separable dimension stimuli both without (Experiments 1 and 2) and with (Experiments 3 and 4) instructions to attend to the relevant stimulus dimension. The same pattern of sequential effect results were previously observed in the modified Garner task using integral stimuli. Two important differences between these results are apparent: First, unlike for integral stimuli, there was little evidence for moderation of the sequential effects in the filtering task when the irrelevant dimension was varied. This supports the claim that participants are able to selectively ignore the irrelevant dimension. Second, unlike for integral dimensions, the magnitude of the sequential effects appears to be the same across the control, correlated, and filtering tasks. This supports the claim that selective attention makes all three tasks effectively the same - that is, all three tasks involve categorisation only along the relevant stimulus dimension without influence from the irrelevant dimension.

In the remainder of the paper, we turn to the computational modeling of the sequential effects in order to:

1. contrast the predictions of the sequential exemplar-based model with other theoretical explanations, and
2. compare the parameter estimates between the fits to the separable data, reported here, and the integral data reported in Little et al. (2016).

Computational Modeling

We fit the sequence-sensitive exemplar model introduced by Little et al. (2016). The model can be thought of as being a combination of sequence-insensitive and sequence-sensitive mechanisms. Following the Exemplar-Based Random Walk Model (Nosofsky & Palmeri, 1997b), the model assumes that when presented with an item to be categorized, exemplars race to be retrieved based on their similarity to the current item. Similarity between an exemplar j and the current item i is computed as a function of the distance, d_{ij} between the items:

$$s_{ij} = \exp\left(-cd_{ij}^{\rho}\right) \quad (1)$$

The sensitivity parameter, c, determines how much similarity decreases with distance while the exponent, ρ determines whether the decrease is exponential ($\rho = 1$) or Gaussian ($\rho = 2$; Nosofsky, 1986; Shepard, 1958a, 1958b). In Little et al. (2016), we found that the Gaussian distance function fit better for the integral dimension color stimuli, perhaps due to the high confusability of those stimuli even after extensive training (see Nosofsky, 1986; Shepard, 1986). Our initial model fits revealed that the exponential function fit better for the separable dimension stimuli. Consequently, we only report the exponential model for the separable dimensions and the Gaussian function for the integral dimensions.

Distance is computed based on the location of the items in a geometric, psychological representation of the stimulus space:

$$d_{ij} = \left[\sum_{m} w_m |x_{im} - x_{jm}|^r\right]^{\frac{1}{r}} \quad (2)$$

where the coordinate of items i and j on dimension m are given by x_{im} and x_{jm}, respectively. The parameter w_m reflects the attention weight applied to dimension m. The parameter r determines the distance metric, with typical values of $r = 2$ (Euclidean distance) being applied for integral dimensions and $r = 1$ (city-block distance) being applied for separable dimensions. We set $r = 1$ to fit our separable dimension stimuli data.

In the sequence-insensitive exemplar model, the retrieval of exemplars is modeled by assuming that the mean drift-rate towards category A, ν_A, in an evidence accumulation process, is computed by summing the similarity of the current item to all category A items divided by the summed similarity of the current item to items from both categories, as follows:

$$\nu_A = \frac{b_A \sum_{j \in A} M_j s_{ij}}{b_A \sum_{j \in A} M_j s_{ij} + (1 - b_A) \sum_{j \in B} M_j s_{ij}} \tag{3}$$

The parameter b_A (where $0 \leq b_A \leq 1$) represents the response bias toward category A, while the parameter M_j represents the memory strength of stored exemplar j. In the present application, memory strengths were set equal to 1 for all exemplars. The mean drift rate toward category B is $\nu_B = 1 - \nu_A$.

Two sequence-sensitive mechanisms are added to account for the repetition effect, the pushing effect, and the pulling effect. First, we assumed that there was a bias to change responses when the current item was different from the previous item. The change-response bias parameter for category A was defined as follows:

$$\beta_A = \begin{cases} \beta, & \text{if item } k-1 \in B \text{ and item } k \neq \text{to item } k-1 \\ 1, & \text{otherwise.} \end{cases} \tag{4}$$

The change-response bias parameter for category B was defined in the same manner.

The second mechanism adds a boost to the summed similarity of category A if the previous item belonged to A and a boost to category B if the previous item belonged to category B. The extent of this boost depends on the similarity of the current item (on trial k) and the previous item (on trial k-1).

$$\begin{aligned} BOOST_A &= \begin{cases} \alpha s_{k,k-1} & \text{if item } k-1 \in \text{category A} \\ 0 & \text{otherwise} \end{cases} \\ BOOST_B &= \begin{cases} \alpha s_{k,k-1} & \text{if item } k-1 \in \text{category B} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

Both of these mechanisms are used to generalize the computation of the drift-rate (see Equation 3) as follows:

$$\nu_A = \frac{\beta_A \left(ss_A + BOOST_A\right)}{\beta_A \left(ss_A + BOOST_A\right) + \beta_B \left(ss_B + BOOST_B\right)} \tag{6}$$

where $ss_A = b_A \sum_{j \in A} M_j s_{ij}$ and similarly for ss_B.

Further details of the model are provided in Little et al. (2016). We note here that to accommodate potential differences between the correlated and control condition, we allowed the attention weight to vary between the correlated condition and the control and filtering conditions. This would allow the correlated condition to show an overall facilitation effect as is typically observed with integral dimensions. This captures the fact that, in the correlated task, the irrelevant dimension provides additional information regarding the category membership. In the following, we report the attention to the *irrelevant* dimension, with attention to the relevant dimension given as one minus the reported value. With separable dimensions, we expect the attention weights for the irrelevant dimension to tend toward 0.

To allow for a potential filtering interference effect, we allowed the sensitivity parameter, c, to vary between the filtering condition and the control and correlated conditions. This captures the potential decrease in sensitivity resulting from increasing the number of stimuli compared to the control condition. We additionally estimated different biases, b_A, for each task (i.e., control-1, control-2, correlated-1, correlated-2, and filtering).

The mean drift rates from Equation 6 are computed for each trial and are used to drive a pair of linear ballistic accumulators (LBA; Brown & Heathcote, 2008), one for category A and another for category B. Each of these accumulators has a uniformly varying start point in the range $[0, A]$. The decision thresholds for each accumulator are T_A and T_B. We further assumed that drift rate varied between trials according to a normal distribution with a freely estimated standard deviation, s_ν. Finally, a freely estimated non-decision time constant, t_0, was added to the prediction decision times in order to capture non-decision components. Full details of the LBA are provided in Brown and Heathcote (2008).

In short, the model can be thought of as sequence-sensitive exemplar-based "front-end" that generates a drift rate on each trial based on the current item's relationship to the long-term knowledge of the members of both categories and the item from the immediately preceding trial. These drift rates are used to generate RT predictions using an LBA "back-end". We use the terms front-end and back-end here to indicate that, to some extent, these mechanisms may be replaced with other mechanisms to build new models. For instance, the back-end LBA model could be replaced by a diffusion decision model (Ratcliff, 1978) to predict the RTs given the drift rates computed by the front-end model. In our comparison models, we introduce new front end models but maintain the same back-end LBA structure.

Comparison models

As indicated in our introduction, there are at least two plausible comparison models.

Sequence-sensitive GRT-LBA. Recall that the sequence-sensitive GRT-LBA predicts that the pushing and pulling effects should be equivalent between previous items that are members of the same category as each other (e.g., the far and mid item pushing effects should be equivalent; see Figure 2). Full details of this model are provided in Little et al. (2016). In short, each item is represented by a bivariate normal distribution with a mean location and standard deviation. We assumed the locations were the ideal values for each feature (see Figure 1). We assumed a common standard deviation for all items in the control and correlated conditions but allowed the standard deviation to be greater in the filtering condition. We allowed different decision boundaries in the form, $m_1x + m_2y = m_0$, in all tasks. m_0 represents the bias toward either category region and is free to vary across all tasks. m_2 was fixed to equal 1. In the control and filtering tasks, m_1 was set to 0 to allow only boundaries orthogonal to the relevant dimensions, but m_1 was allowed to vary in the correlated conditions allowing for diagonal boundaries. There were additionally five parameters associated with the LBA back-end of the model (starting point, decision thresholds for the category A and B accumulators, drift rate variability, and non-decision time).

To model the sequential effects, the perceptual variability of an item was reduced by a multiplicative factor of κ when an item repeated. The decision boundary was also shifted by a factor of τ to increase or decrease the region of the category space occupied by the previous item's category. Hence, there were 16 parameters in total for the sequence-sensitive GRT-LBA model matching the 16 parameters of the sequence-sensitive EB-LBA model.

Although we later apply a full Bayesian comparison between our sequence sensitive exemplar and feature-based models, because there is a strong qualitative contrast in the predictions of the models, we compared the sequence-sensitive EB-LBA model and the sequence-sensitive GRT-LBA model by fitting the correct and error RT data from each trial across all four experiments using maximum likelihood (see Little et al., 2016). That is, given the predicted drift rates for each trial (computed from the EB or GRT front-end) along with the remaining LBA parameters, each model predicts a distribution of possible correct and error response times for each trial. This distribution was used to determine the likelihood of the observed RT for that trial. The logs of these likelihoods were then summed across trials to give the log likelihood for a participant. Log-likelihoods were summed across participants to give a log likelihood for the entire experiment.

To compare the models, we used the Bayesian Information Criteria (BIC; Schwarz et al., 1978), which penalizes the log-likelihood, $ln(L)$, on the basis of the number of free parameters, k, and the size of the sample, n; $BIC = -2ln(L) + kln(n)$. As shown in Table 3, the sequential EB-LBA model was preferred for both stimulus types with and without attentional instructions.

Table 3

Model comparison results for the sequence sensitive exemplar model (EB-LBA) and sequence sensitive rule-based model (GRT-LBA)

Experiment	Model	-lnL	BIC
1	EB-LBA	-14019	**-26726**
	GRT-DBT	-13974	-26637
2	EB-LBA	-4905	**-8499**
	GRT-DBT	-4831	-8353
3	EB-LBA	-9189	**-17422**
	GRT-DBT	-9167	-17377
4	EB-LBA	-16454	**-31952**
	GRT-DBT	-16304	-31652

As expected from examination of Figures 5-8, a key failing of the sequence sensitive GRT-LBA model was the invariant within-category RT prediction for the far and middle items. We do not pursue further analysis of the sequence sensitive GRT-LBA model here because for all of our conditions, the sequence sensitive GRT-LBA model is ruled out based on this strong qualitative misfit to the data.

Sequence-sensitive feature-based LBA. To test whether the irrelevant dimension is completely filtered out for the separable dimension category, we fit a model in which only the relevant dimension enters into the similarity comparison. This model is identical to the sequence-sensitive exemplar model, except that only one dimension enters into the computation of distance:

$$d_{ij} = |x_i - x_j|. \tag{7}$$

More specifically, the attention weights are no longer estimated for the sequence sensitive feature model. Additionally, the change-response bias for each category only occurs if there was a change on the attended dimension rather than a change on the entire stimulus.

Fitting procedure

We implemented the sequence-sensitive exemplar and feature models in a hierarchical Bayesian framework in which each subject's parameters were drawn from group-level distributions over those parameters. Full details of the fitting procedure are reported in Appendix D.

Model Comparison. To compare each of the models, we used the Deviance Information Criterion (DIC; Spiegelhalter, Best, Carlin, & van der Linde, 2002). Given a set of parameters, θ, the likelihood of the data, y, is defined as $p\left(y|\theta\right)$ and the posterior deviance is $D\left(\theta\right) = -2\ln p\left(y|\theta\right)$. DIC is computed as:

$$DIC = \bar{D}\left(\theta\right) + 2p_D \tag{8}$$

Table 4
DIC results for the sequence-sensitive l EB-LBA (Exemplar) model and sequence-sensitive feature model.

	Experiment				
	Integral	No Instructions		Attention Instructions	
Model	Bri/Sat	1: Bri/Sat	2: Line/Sat	3: Bri/Sat	4: Line/Sat
Exemplar	**-43756**	**-27678**	**-9510**	**-18141**	**-32559**
Feature	-43440	-24827	-2840.9	-18100	-32559

where $\bar{D}(\theta)$ is a measure of fit computed from the mean of the posterior deviances across samples of the posterior and p_D is penalty based on the complexity or flexibility of the model; $p_D = 2var[\ln p(y|\theta)]$ (Gelman, Hwang, & Vehtari, 2014). Here the variance is taken across samples of the posterior. Depending on the number of participants in the experiment and the number of parameters in the model, which determines the number of sampling chains (see Appendix D), we used between 100,000 and 270,000 posterior samples to estimate the DIC.

Modeling results

We estimated the posteriors separately for each experiment. Likewise, we re-fit the integral dimension data from Little et al. (2016) to compare the sequential EB-LBA model with the sequential feature-based LBA model. The model comparison results favored the sequence-sensitive exemplar model in all cases (see Table 4); in Experiment 3, the DICs were similar and in Experiment 4, the DICs were nearly identical between the models. Evidently, the implication is that with separable dimensions without instructions to attend to the relevant dimension (i.e., Experiments 1 and 2), the influence of the irrelevant dimension is not completely filtered out. By contrast, when participants are instructed to attend to the relevant dimension, we find that adding attention to the irrelevant dimension does little to improve the fit. Even so, we still find strong within-dimension sequential effects remain.

We proceed primarily interested in the parameters governing the exemplar-based representation component of the model. These parameters can be partitioned into parameters responsible for the difference in the representation between tasks (e.g., control, correlated

& filtering) and parameters responsible for generating the sequential effects. The former set of parameters include the attention weights (i.e., attention to the irrelevant dimension), and the sensitivity parameters, which vary between the control/correlated task and the filtering tasks, respectively. The sequence-sensitive mechanisms include parameters governing the strength of the boost to the previous item's category, termed α, and the bias toward changing responses when the item changes from the previous to the current trial, termed β. To the extent that α is greater than 0, the category of the previous item will be boosted in proportion to the similarity between the current and previous item. Likewise, to the extent that β is greater than 1, there will be an increased bias to shift responses following a non-repetition.

Figure 9 shows the posterior parameter distributions for both the attention and sensitivity parameters for each participant. Figure 10 shows the posterior parameter distributions for both the sequential parameters, α and β, for each participant. Group posterior estimates of the remaining parameters are reported in Appendix E.

Compared to the integral dimension condition (Little et al., 2016), the attention weights for the separable dimensions estimated much closer to zero indicating less of an influence of the irrelevant dimension in the separable dimension experiments. Conversely, the sensitivity estimates are larger for the separable dimensions than for the integral dimensions. (Note that the scales of Figure 9 have been adjusted to allow for comparison between the c and $c_{filtering}$ parameters.) For the integral dimension data, the sensitivity estimate is consistently lower in the filtering task than in the control or correlated tasks. For the separable dimension experiments, this difference is much less regular with some participants showing the same direction of difference as for the integral dimensions and others showing a difference in the opposite direction or no difference.

The sequential parameters are shown in Figure 10. There is variation in the value of α across participants. Note that the previous item exerts more of an influence on the current response with greater values of α. On the other hand, the estimates of β are almost always greater than 1 indicating a positive boost to the category of the previous item.

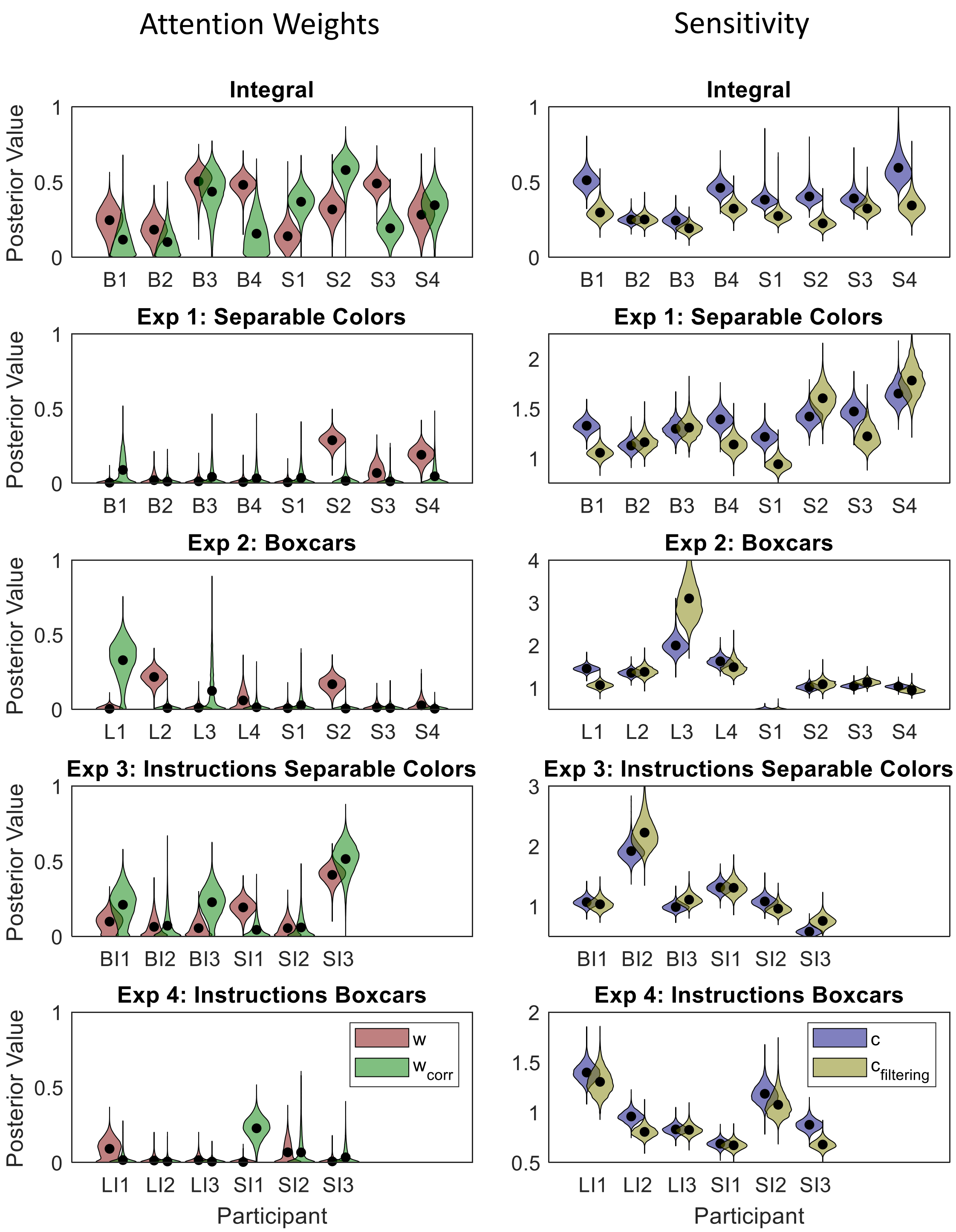

Figure 9. Left panels: Estimated posterior densities of the attention weights from the sequential EB model in each experiment. w indexes the attention weights in the control and filtering conditions; w_{corr} indexes the attention weights in the correlated condition. Right panels: Estimated posterior densities of the sensitivity parameters in each experiment. c indexes the sensitivity in the control and correlated conditions; $c_filtering$ indexes sensitivity in the filtering conditions.

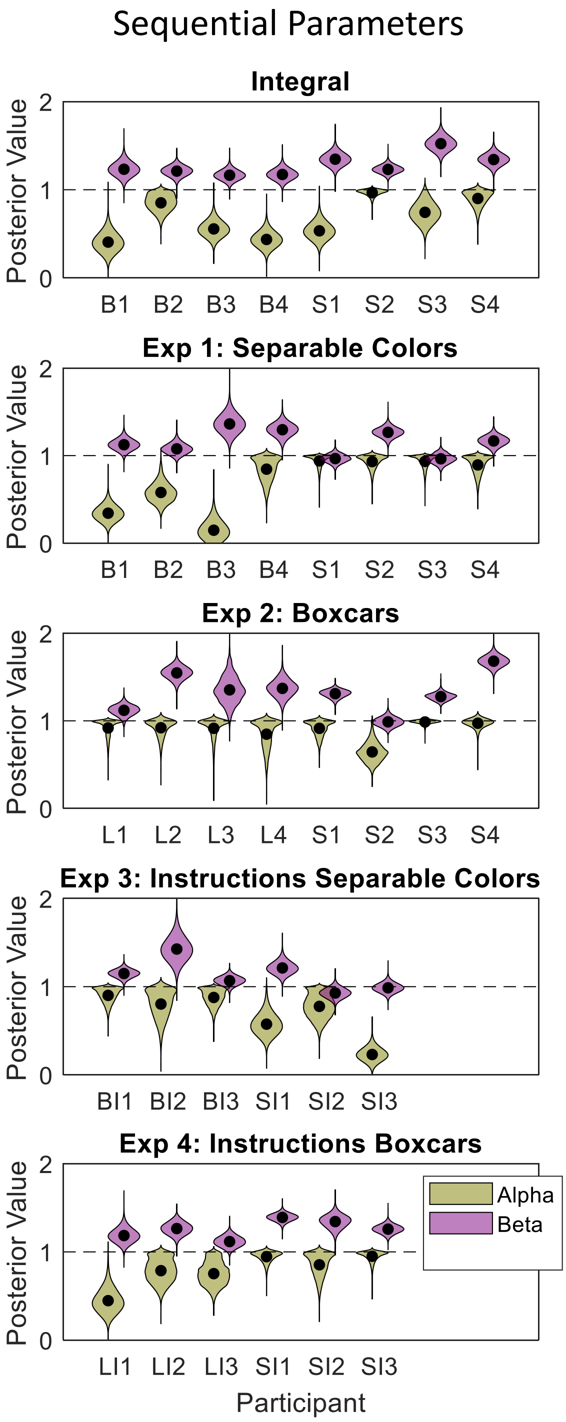

Figure 10. Estimated posterior densities of the sequential parameters from the sequential EB model in each experiment. Alpha indexes the effect of similarity of the previous item on the current item; Beta indexes the bias to shift responses when the item changes.

Finally, to demonstrate that the sequential feature-based model provides an adequate account of the data, we present the posterior predictive distributions for the best fitting model to the RTs from each experiment in Figures 11-14. The posterior predictions for the integral dimension data from Little et al. (2016) and the accuracies are also shown in Appendix F.

Discussion

In these four experiments, the analysis of trial-by-trial RTs and accuracy as a function of the preceding item for separable-dimension stimuli revealed distinct patterns of sequential effects for near-boundary current items similar to those found for integral-dimension stimuli in Little et al. (2016). Specifically, we found substantial support for near-boundary repetition effects, same-category far-item "pushing" effects, and opposite-category adjacent-item "pulling" effects in all tasks. However, unlike the integral dimension findings, there was evidence for no difference in the effect as a function of an irrelevant dimension change in the filtering task. This follows naturally from the fact that it is easier to selectively attend to the relevant dimension when categorizing separable stimuli (Garner, 1974; Shepard, 1964); the absence of a sequential irrelevant dimension change effect provides a novel characterization of performance using separable dimensions. Unlike Stewart and Brown (2004), we find clear evidence for an influence of the overall distance or magnitude of the difference between items. We characterized this result by fitting a sequence sensitive exemplar model which has separate mechanisms for representing differences in the distance between exemplars (using selective attention to highlight relevant dimensions) and mechanisms for allowing an influence of the previous item on the current categorisation response.

For most participants, the attention to the irrelevant dimension was estimated to be near zero. The fact that the additional dimension information does not affect the sequential effects seen with separable dimensions suggests that the sequential effects are the same as those found in related decision making tasks, such as unidimensional perceptual categorization, identification, and other simple RT tasks. The near-boundary repetition effect

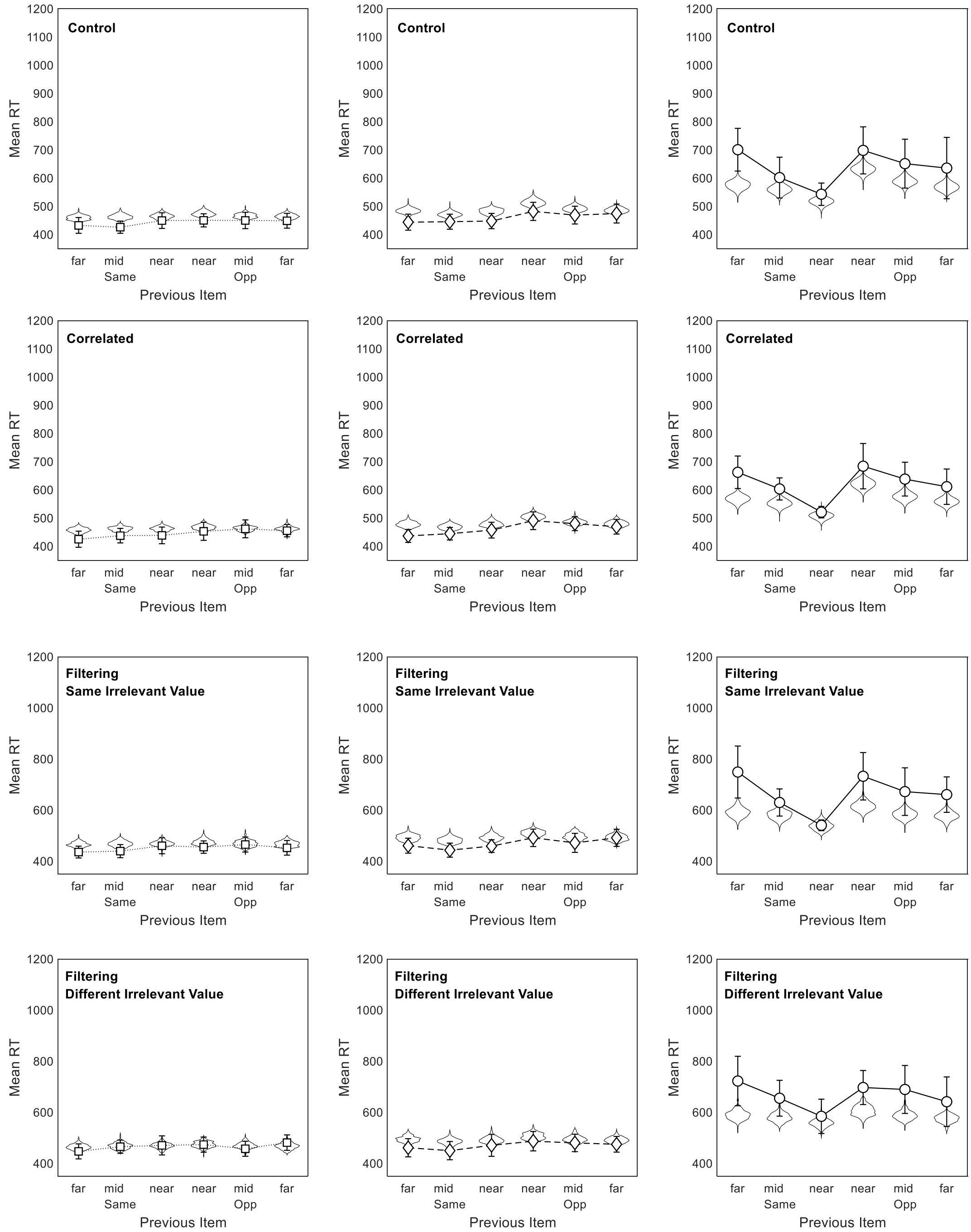

Figure 11. Separable Color data from Experiment 1. Posterior predictive distributions from the sequential feature-based model for RTs in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

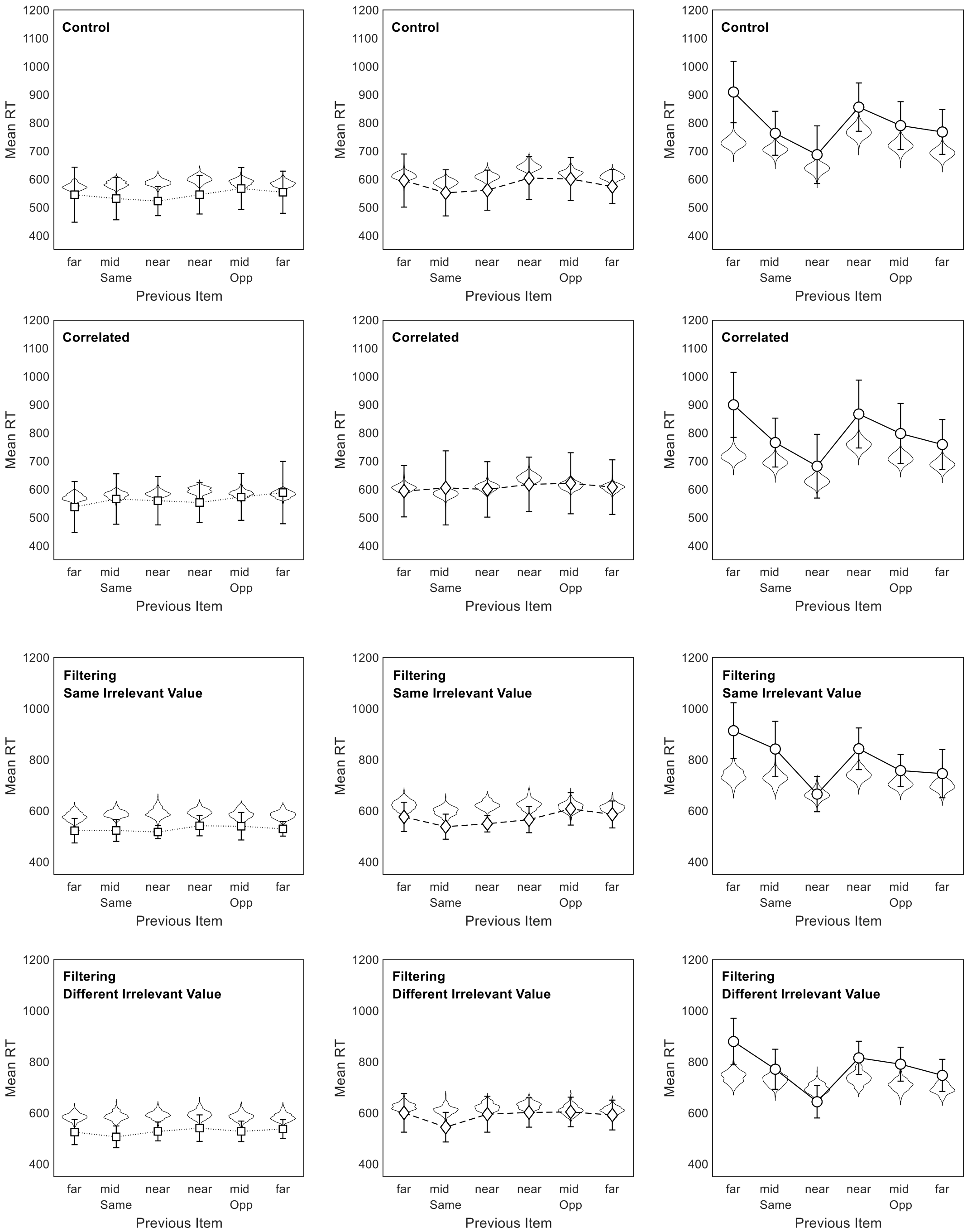

Figure 12. Separable Boxcar data from Experiment 2. Posterior predictive distributions from the sequential feature-based model for RTs in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

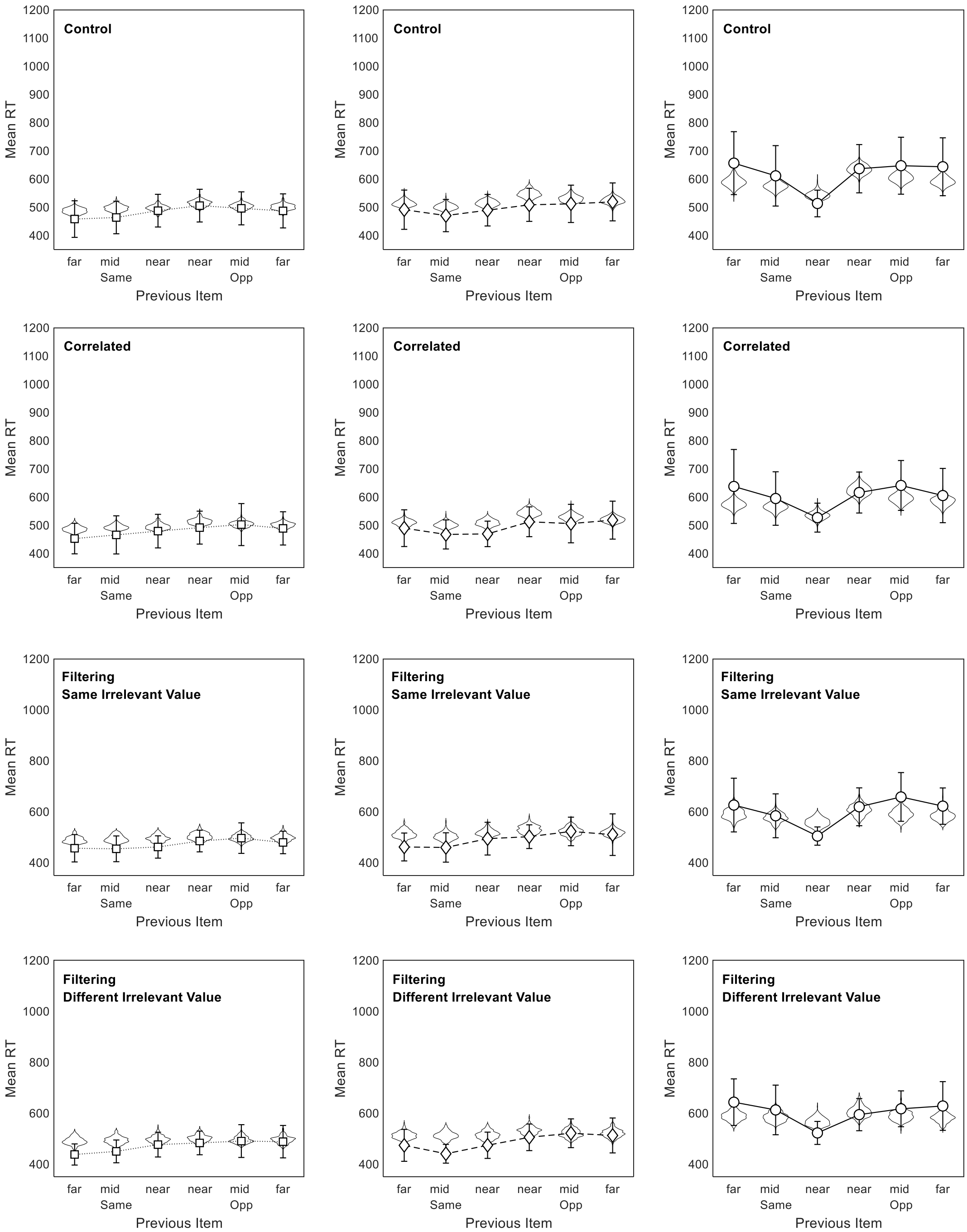

Figure 13. Separable Color data from Experiment 3. Posterior predictive distributions from the sequential feature-based model for RTs in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

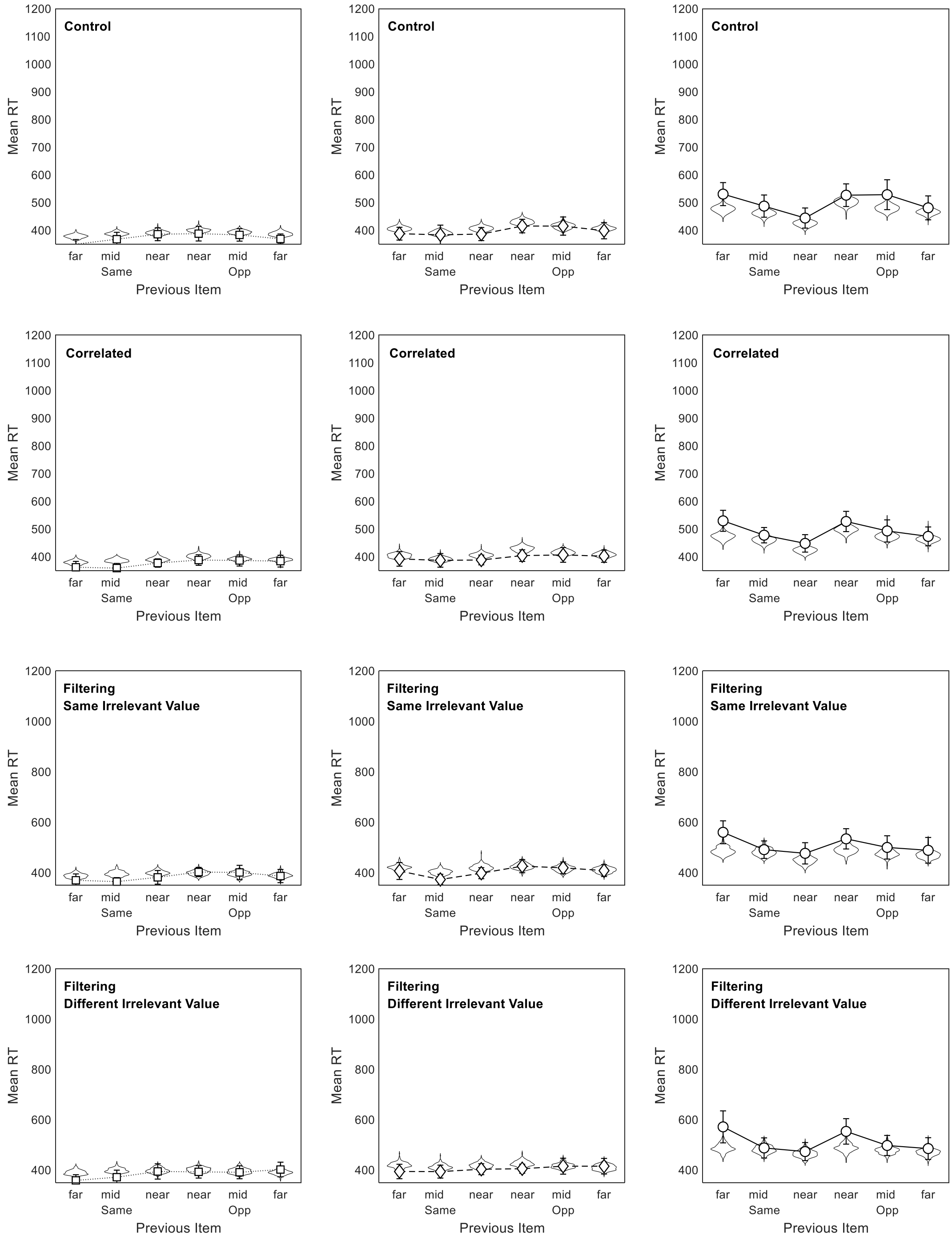

Figure 14. Separable Boxcar data from Experiment 4. Posterior predictive distributions from the sequential feature-based model for RTs in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

has previously been found for separable-dimension stimuli in Felfoldy (1974) and Lockhead, Gruenewald, and King (1978). The same-category far-item "pushing" effect has been observed as a "category contrast effect" in Stewart et al. (2002), who tested category learning using a 4-by-4 stimulus space using separable dimensions. A similar contrast effect has also been found in a categorization task with probabilistic rules (Jones, Love, & Maddox, 2006). The opposite-category adjacent-item "pulling" effect, while not directly observed within a standard Garner task, is largely similar to the assimilation effect observed in the categorization of facial expressions (Hsu & Lee, 2016; Hsu & Yang, 2013), as well as in unidimensional identification tasks (Lockhead & King, 1983; M. Treisman, 1985; Ward & Lockhead, 1970). By employing the modified Garner paradigm developed by Little et al. (2016), the present study has been able to capture a full set of sequential effects in perceptual categorization with separable dimensions.

The ubiquity of sequential effects across cognitive tasks involving repeated trials perhaps points towards a larger emphasis on "short-term" influences in addition to the reliance on "long-term" category knowledge. Although these are consistent with dual-store models of memory (Baddeley & Hitch, 1974; Cowan, 1995), our preference is to view the short-term influences as resulting from items which have recently occupied attention (McElree, 2006). Our sequence-sensitive model implements these two sources by combining exemplar-based representations of the stimulus space with similarity-based effects of the immediately preceding item and changes in response bias. The posterior predictive distributions show that the data was captured reasonably well by these models.

One important implication is that trials should not be assumed to be independent in any psychological task involving repeated trials, and an aggregation of trial data may result in the loss of critical information that could inform psychological models of perception and cognition. Our sequence-sensitive models offer a more complete picture of both the representation and the process underlying perceptual categorization. Building on the present study, one potential avenue of research could be to explore the extent of sequential effects in the standard Garner task and other similar two-choice paradigms that involve repeated trials, and further investigate how short-term and long-term representations as well as decisional mechanisms interact.

What are the implications for real-world categorisation?

There has been a recent emphasis in categorisation research in applying the models and concepts from experiments using abstract stimuli to real-world, naturalistic category decision tasks (Meagher, Cataldo, Douglas, McDaniel, & Nosofsky, 2018; Nosofsky & McDaniel, 2019; Nosofsky, Sanders, Gerdom, Douglas, & McDaniel, 2017; Nosofsky, Sanders, & McDaniel, 2018a, 2018b; Nosofsky, Sanders, Meagher, & Douglas, 2019). Our tasks are tightly controlled, using abstract stimuli, in order to test highly diagnostic predictions arising from several prominent theories of categorization. Nevertheless, it does not seem far-fetched to us that many repetitive real-world categorisation tasks do involve repetition of the same categories with short intervals between each decision. Tasks like medical diagnosis in radiography, auditing and accounting, passport control, baggage checking, photo editing, and so on, likely involve the repetitive and quick categorisation of stimuli. Consequently, the sequential effects identified here and in Little et al. (2016) are relevant to these real-world tasks. In a nutshell, we show that when making decisions, people base their decisions on their memory of category information and on recent memory of the previous item and the associated response.

In a recent "big data" analysis of over 135,000 sequential grocery purchase decisions, Hornsby and Love (2022) showed that the decision to select an item for purchase from an online store is cued by recently selected items. In their model, the similarity between potential choices and previously selected items is based on a comparison across different representations. That is, associations are computed across semantic similarities (e.g., conceptual overlap as measured by Latent Semantic Analysis; Landauer, McNamara, Dennis, and Kintsch (2013)), hierarchical relations (e.g., shampoo is more likely to be cued by hair care products than by cleaning products), and episodic relations (e.g., cereal and milk often co-occur within the same context). The largest influence seemed to be from the most recent items. Hence, in real world scenarios, once the computational representation of similarity is adapted to the stimulus and decision context, the influence of previous items may be similar to what we propose here. Though some of the specific effects we observe, such as the pulling or pushing effects, may not be particularly relevant or consequential for untimed, one-off, real-world decisions, such effects are important for differentiating competing theoretical accounts of categorisation.

What are the implications for the standard Garner task?

Garner (1974) posited sequential effects might provide an explanation for the filtering interference observed in the standard task. While we observe sequential effects in our modified task with separable stimuli, we crucially do not observe any effects analogous to filtering interference. There have a number of papers which have investigated sequential effects with integral stimuli in the standard Garner task in order to determine the sequential locus of filtering interference (Burns, 2016; Dyson & Quinlan, 2010; Felfoldy, 1974; Garner, 1974; Huettel & Lockhead, 1999). Presumably our results imply that similar sequential effects should be found with both integral and separable dimensions in the standard Garner task. So why do we not observe filtering interference in the standard task?

Our theoretical modeling suggests one plausible answer to this question. Our sequence-sensitive models assume that the one key contributor to the observed sequential effects is the similarity between the current and previous item. For integral dimension stimuli, the configuration of items in each of the tasks results in a different trial-by-trial ordering of previous-to-current item similarities. That is for the integral dimension stimuli, the filtering condition contains four distinct previous-to-current item comparisons: direct repetitions (both relevant and irrelevant dimensions repeat, RR), previous same category items (relevant dimension repeats, irrelevant dimension changes, RC), previous opposite category items varying only on the relevant dimension (CR), and previous opposite category items varying on both dimensions (CC; see Dyson and Quinlan, 2010, and Huettel and Lockhead, 1999).

By contrast, in the separable condition, where only the relevant dimension is represented, only two distinct previous-to-current item comparison arise: RR and RC are equivalent as are CR and CC. For separable dimensions, these sequential orderings are shared across the control and correlated tasks. The implication is that any sequential effects in the standard Garner task with separable dimensions should be the same across the control, correlated, and filtering tasks. We might expect these sequential effects to look like the patterns associated with repetitions and alternations in simple 2AFC tasks (Gökaydin et al., 2016; Jones et al., 2013).

Consequently, Garner interference with integral dimensions arises due to the contributions of previous item similarity and due to the different configurations of similarity across tasks. This suggests that one target for future research would be to manipulate the configurations of items in the standard Garner task and examine the pattern of sequential effects.

Sequential effects in more complex tasks

Recent work in categorisation using combinations of rule-based decisions have shown that many types of separable dimensions (including the stimuli used in Experiments 2 and 4) are processed independently either in serial or in parallel (see e.g., Little et al., 2011, 2013; Moneer et al., 2016). For example, to demonstrate that separable dimensions are processed independently, a variant of the SFT double factorial task is used (Blunden, Hammond, Howe, & Little, 2022; Blunden et al., 2020, 2015; Cheng et al., 2018; Fifić, Little, & Nosofsky, 2010; Fifić et al., 2008; Little et al., 2011, 2013; Moneer et al., 2016). In this task, one category of items is defined by a conjunctive decision rule, such that items belong to this category if they are, for instance, brighter *AND* more saturated than the values given by the rule. A second category is defined by a disjunctive rule that only requires one of the dimensions to be less than some threshold value (e.g., the brightness is less then x *OR* the saturation is less then y).

For the conjunctive category, values of both dimensions are combined orthogonally to create stimuli for which these decisions are either easy on both dimensions, difficult on both dimensions, or easy on one dimension but difficult on the other. By performing contrasts of the mean RTs (and the distributions), one can differentiate whether the stimulus dimensions are combined in serial, in which case the interaction will be additive, in parallel, in which case the interaction will be underadditive, or coactive and pooled into a single decision, in which case the interaction will be overadditive (Algom, Eidels, Hawkins, Jefferson, & Townsend, 2015; Altieri, Fifić, Little, & Yang, 2017; Little, Eidels, Houpt, Garrett, & Griffiths, 2019; Little, Yang, Eidels, & Townsend, 2022; Townsend & Nozawa, 1995). Hence, both dimensions of the stimulus are relevant and selective attention cannot be used to filter out either dimension unlike in the present task. However, it is possible that selective attention may still play a role, for serial processing models at least, by prioritizing some dimensions for processing before other dimensions (e.g., brightness is processed before saturation). For example, Lamberts (1995, 1998, 2000) showed that differences in dimensional salience can result in some dimensions being processed before others. Eye-tracking studies (Blair, Watson, Walshe, & Maj, 2009; Rehder & Hoffman, 2005a, 2005b) have shown that estimates of the selective attention weights correlate strongly with sequential fixations (e.g., higher weights indicate prioritization).

This provides a clue to how sequential effects might influence responding in more complex tasks involving conjunctive and disjunctive rules. Consider a novel analysis of generalization developed by Jones et al. (2006). Across learning, similarity-based generalization comes to be restricted to solely the relevant dimension (Jones, Maddox, & Love, 2005). Using their sequential analysis in a 4-category task, Jones et al. found that generalization was independent across the horizontal and vertical categories indicating that the influence of sequential items was independent along each dimension. Their experiment employed Gabor stimuli varying on frequency and orientation, which are likely to be separable dimensions (cf. Moneer et al., 2016). The implication is that when there are multiple, relevant separable dimensions, sequential effects should be evident from changes in both dimensions. However, an important consideration is whether the separable dimensions are processed in serial or in parallel (Griffiths et al., 2017). A target for future work will be to incorporate sequential effects into models that vary in their processing architecture (e.g., Fifić et al., 2010).

Conclusion

Garner's famous speeded classification tasks are among the most important results for differentiating integral and separable dimensions. A limitation of the standard Garner tasks is that a bypass strategy may be used in the control and correlated tasks. The modified task introduced by Little et al. (2016) circumvents the use of the bypass strategy but crucially preserves important features of the standard task such as typical pattern of mean RTs across the correlated, control, and filtering conditions. Little et al. also discovered complex but systematic sequential effects in the modified tasks and developed an extended exemplar-based random walk model to account for the full set of results. In the present paper, we observe the same systematic sequential effects with separable dimensions but restricted to the relevant dimension. The conclusion is that, for separable dimensions, attention is largely restricted to the relevant dimension, whereas attention encompasses all of the dimensions for integral dimensions. A further conclusion is that other cognitive processes are essentially the same. We show that our sequential EB-LBA model provides a good account of our separable dimension data demonstrating the generality of the sequential effects and the generality of the cognitive mechanisms formalised by the model.

References

Algom, D., Eidels, A., Hawkins, R. X., Jefferson, B., & Townsend, J. T. (2015). Features of response times: Identification of cognitive mechanisms through mathematical modeling. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *The oxford handbook of computational and mathematical psychology* (p. 63-98). Oxford.

Algom, D., & Fitousi, D. (2016). Half a century of research on garner interference and the separability-integrality distinction. *Psychological Bulletin*.

Altieri, N., Fifić, M., Little, D. R., & Yang, C.-T. (2017). A tutorial introduction and historical background to systems factorial technology. In D. R. Little, N. Aliteri, M. Fifić, & C.-T. Yang (Eds.), *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms.* Elsevier.

Ashby, F. G. (1989). Stochastic general recognition theory. *Human information processing: Measures, mechanisms, and models*, 435–457.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology*, *44*, 210-329.

Ashby, F. G., Boynton, G., & Lee, W. W. (1994). Categorization response times with multidimensional stimuli. *Perception & Psychophysics*, *55*, 11-27.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 33-53.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*(2), 154-179.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47–89.

Blair, M. R., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eyetracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1196.

Blunden, A. G., Hammond, D., Howe, P. D. L., & Little, D. R. (2022). Characterizing the time course of decision-making in change detection. *Psychological Review*. (Accepted 10-May-2021)

Blunden, A. G., Howe, P. D. L., & Little, D. R. (2020). Evidence that within-dimension features are generally processed coactively. *Attention, Perpection, & Psychophysics*, *82*(1), 193-227.

Blunden, A. G., Wang, T., Griffiths, D., & Little, D. (2015). Logical-rules and the classification of integral dimensions: Individual differences in the processing of arbitrary dimensions. *Frontiers in Psychology*, *5*, 1531.

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

Bürkner, P.-C. (2017). *BRMS: Bayesian regression models using Stan, version 1.10.2.* https://cran.r-project.org/package=brms.

Burns, D. M. (2016). Garner interference is not solely driven by stimulus uncertainty. *Psychonomic bulletin & review*, *23*(6), 1846–1853.

Cheng, X. J., McCarthy, C. J., Wang, T. S., Palmeri, T. J., & Little, D. R. (2018). Composite faces are not (necessarily) processed coactively: A test using systems factorial technology and logical-rule models. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *44*(6), 833-862.

Cho, R. Y., Nystrom, L. E., Brown, E. T., Jones, A. D., Braver, T. S., Holmes, P. J., & Cohen, J. D. (2002). Mechanisms underlying dependencies of performance on stimulus history in a two-alternative forced-choice task. *Cognitive, Affective, & Behavioral Neuroscience*, *2*, 283–299.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford: Oxford University Press.

Dyson, B. J., & Quinlan, P. T. (2010). Decomposing the garner interference paradigm: Evidence for dissociations between macrolevel and microlevel performance. *Attention, Perception, & Psychophysics*, *72*(6), 1676–1691.

Felfoldy, G. L. (1974). Repetition effects in choice reaction time to multidimensional stimuli. *Attention, Perception, & Psychophysics*, *15*(3), 453–459.

Fifić, M., Little, D. R., & Nosofsky, R. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review*, *117*, 309-348.

Fifić, M., Nosofsky, R. M., & Townsend, J. (2008). Information-processing architectures in multidimensional classification: A validation of test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 356-375.

Fletcher, B., & Rabbitt, P. M. A. (1978). The changing pattern of perceptual analytic strategies and response selection with practice in a two-choice reaction time task. *The Quarterly Journal* *of Experimental Psychology*, *30*, 417-427.

Garner, W. R. (1974). *The processing of information and structure.* New York: Psychology Press.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225-241.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models* [Book]. New York, NY: Cambridge University Press.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and Computing*, *24*(6), 997-1016.

Gökaydin, D., Navarro, D. J., Ma-Wyatt, A., & Perfors, A. (2016). The structure of sequential effects. *Journal of Experimental Psychology: General*, *145*, 110.

Gottwald, R. L., & Garner, W. (1972). Effects of focusing strategy on speeded classification with grouping, filtering, and condensation tasks [Journal Article]. *Perception & Psychophysics*, *11*(2), 179-182.

Grau, J. W., & Nelson, D. K. (2016). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General*, *117*, 347.

Green, D., & Swets, J. (1966). Signal detection theory and psychophysics. *New York.*

Griffiths, D. W., Blunden, A. G., & Little, D. R. (2017). Logical-rule based models of categorization: Using systems factorial technology to understand feature and dimensional processing. In D. R. Little, N. Aliteri, M. Fifić, & C.-T. Yang (Eds.), *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms* (p. 245-269). Academic Press.

Hornsby, A., & Love, B. C. (2022). Sequential consumer choice as multi-cued retrieval. *Science Advances*, *8*, eabl9754.

Hsu, S.-M., & Lee, J.-S. (2016). Relative judgment in facial identity perception as revealed by sequential effects. *Attention, Perception, & Psychophysics*, *78*(1), 264–277.

Hsu, S.-M., & Yang, L.-X. (2013). Sequential effects in facial expression categorization. *Emotion*, *13*(3), 573.

Huettel, S. A., & Lockhead, G. R. (1999). Range effects of an irrelevant dimension on classification. *Perception & Psychophysics*, *61*(8), 1624–1645.

Jeffreys, H. (1961). *Theory of probability.* Jeffreys: Clarendon Press.

Jentzsch, I., & Sommer, W. (2002). Functional localization and mechanisms of sequential effects in serial reaction time tasks. *Perception & Psychophysics*, *64*, 1169–1188.

Jones, M., Curran, T., Mozer, M. C., & Wilder, M. H. (2013). Sequential effects in response time reveal learning mechanisms and event representations. *Psychological review*, *120*(3), 628.

Jones, M., Love, B. C., & Maddox, W. T. (2006). The role of similarity in generalization. In *Proceedings of the cognitive science society* (Vol. 28, p. 405-410).

Jones, M., Maddox, W. T., & Love, B. C. (2005). Stimulus generalization in category learning. In *Proceedings of the cognitive science society* (Vol. 27, p. 1066-1071).

Krueger, L. E., & Shapiro, R. G. (1981). Intertrial effects of same-different judgements. *The Quarterly Journal of Experimental Psychology*, *33*, 241–265.

Kruschke, J. (2014). *Doing bayesian data analysis: A tutorial with r, jags, and stan.* Academic Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* Test. *Journal of Experimental Psychology: General*, *142*, 573-603.

Lamberts, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, *124*, 161-180.

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *24*, 695-711.

Lamberts, K. (2000). Information-accumulation theory of speeded classification. *Psychological Review*, *107*, 227-260.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2013). *Handbook of latent semantic analysis.* Psychology Press.

Li, X., Liang, Z., Kleiner, M., & Lu, Z.-L. (2010). Rtbox: A device for highly accurate response time measurements. *Behavior Research Methods*, *42*(1), 212-225.

Little, D. R., Altieri, N., Fifić, M., & Yang, C.-T. (2017). *Systems factorial technology: A theory driven methodology for the identification of perceptual and cognitive mechanisms.* New York: Academic Press.

Little, D. R., Eidels, A., Houpt, J. W., Garrett, P. M., & Griffiths, D. W. (2019). Systems factorial technology analysis of mixtures of processing architectures. *Journal of Mathematical Psychology*, *92*, 102229.

Little, D. R., Nosofsky, R., & Denton, S. E. (2011). Response time tests of logical rule-based models of categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*, 1-27.

Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical-rules and the classification of integral dimensioned stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *39*, 801-820.

Little, D. R., & Smith, P. L. (2018). Replication is already mainstream: Lessons from small-N designs. *Behavioral and Brain Sciences*, *41*, E141.

Little, D. R., Wang, T., & Nosofsky, R. M. (2016). Sequence-sensitive exemplar and decision-bound accounts of speeded-classification performance in a modified garner-tasks paradigm. *Cognitive Psychology*, *89*, 1-38.

Little, D. R., Yang, H., Eidels, A., & Townsend, J. T. (2022). Extending systems factorial technology to errored responses. *Psychological Review*. Retrieved from `psyarxiv.com/jtk6q` (Accepted 19-Dec-2021)

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. [Journal Article]. *Journal of Experimental Psychology*, *72*(1), 95.

Lockhead, G. R., Gruenewald, P., & King, M. (1978). Holistic vs. attribute repetition effects in classifying stimuli. *Memory & Cognition*, *6*(4), 438–445.

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 461.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (p. 147-180). Erlbaum.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification-speeded classification relationship. *Journal of Experimental Psychology: Human perception and performance*, *22*(4), 795.

McElree, B. (2006). Accessing recent events. *Psychology of learning and motivation*, *46*, 155–200.

Meagher, B. J., Cataldo, K., Douglas, B. J., McDaniel, M. A., & Nosofsky, R. M. (2018). Training of rock classifications: The use of computer images versus physical rock samples. *Journal of Geoscience Education*, *66*, 221–230.

Moneer, S., Wang, T., & Little, D. R. (2016). The processing architectures of whole-object features: A logical rules approach. *Journal of Experimental Psychology: Human Perception and Performance*, *43*, 1443-1465.

Nickerson, D. (1936). The specification of color tolerances. *Textile Research*, *6*, 505-514.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39-61.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *13*, 87-108.

Nosofsky, R. M., & McDaniel, M. A. (2019). Recommendations from cognitive psychology for enhancing the teaching of natural-science categories. *Policy Insights from the Behavioral and Brain Sciences*, *6*, 21–28.

Nosofsky, R. M., & Palmeri, T. (1997a). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Perception & Psychophysics*, *59*, 1027-1048.

Nosofsky, R. M., & Palmeri, T. (1997b). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266-300.

Nosofsky, R. M., Sanders, C. A., Gerdom, A., Douglas, B. J., & McDaniel, M. A. (2017). On learning natural-science categories that violate the family-resemblance principle. *Psychological Science*, *28*, 104–114.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018a). A formal psychological model of classification applied to natural-science category learning. *Current Directions in Psychological Science*, *27*, 129–135.

Nosofsky, R. M., Sanders, C. A., & McDaniel, M. A. (2018b). Tests of an exemplar-memory model of classification learning in a high-dimensional natural-science category domain. *Journal of Experimental Psychology: General*, *147*, 328.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2019). Search for the missing dimensions: Building a feature-space representation for a natural-science category domain. *Computational Brain & Behavior*, 1–21.

Plummer, M. (2003). *Jags: A program for analysis of bayesian graphical models using gibbs sampling.*

R Core Team. (2017). *R: A language and environment for statistical computing.* http://www.R-project.org/.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59-108.

Rehder, B., & Hoffman, A. B. (2005a). Eyetracking and selective attention in category learning [Journal Article]. *Cognitive Psychology*, *51*, 1-41.

Rehder, B., & Hoffman, A. B. (2005b). Thirty-something categorization results explained: Selective attention, eyetracking, and model of category learning [Journal Article]. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *31*, 811-829.

Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Shepard, R. N. (1958a). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, *65*(4), 242.

Shepard, R. N. (1958b). Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology*, *55*(6), 509.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*(1), 54-87.

Shepard, R. N. (1986). Discrimination and generalization in identification and classification: Comment on nosofsky. *Journal of Experimental Psychology: General*, *115*, 58-61.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shepp, B. E., & Swartz, K. B. (1976). Selective attention and the processing of integral and nonintegral dimensions: A developmental study. *Journal of Experimental Child Psychology*, *22*, 73-85.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis* (Vol. 26). CRC press.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defence of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083-2101.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 598.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *64*, 583-639.

Stan Development Team. (2017). *Stan: A c++ library for probability and sampling, version 2.16.2.* http://mc-stan.org/.

Stewart, N., Brown, G. D., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 3.

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 416-430.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, *39*, 321-340.

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*, 1003-1035.

Travis, D. (1991). *Effective color displays: Theory and practice.* Academic press.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97-136.

Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties of a model for absolute judgment. *Journal of Mathematical Psychology*, *29*(2), 175–230.

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

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*, 368-384.

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.

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

Appendix A

Appendix: Screenshot of Experiment 1 Instructions

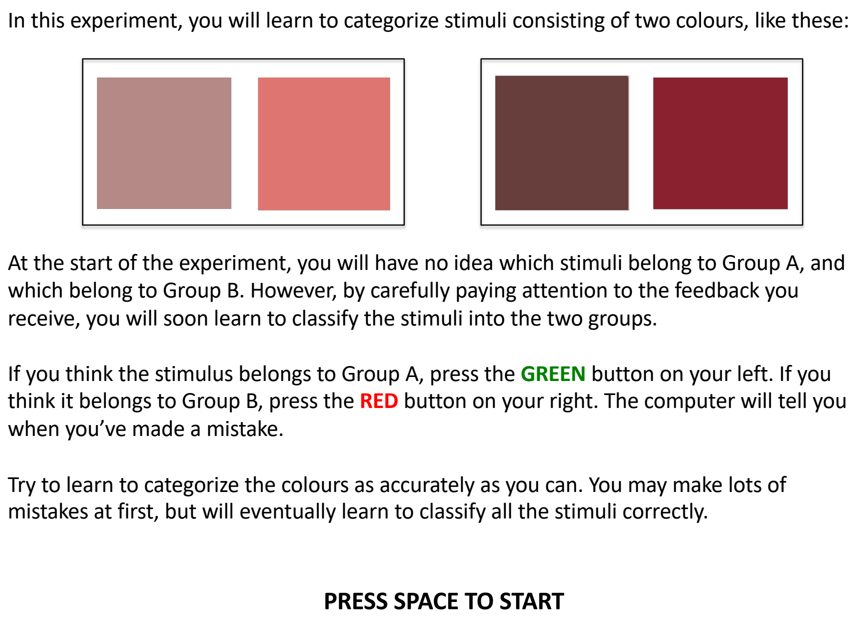

Figure A1. Experiment instructions presented to the participant on the monitor in Experiment 1.

Appendix B

Appendix: Hierarchical Bayesian Analysis of Overall Mean RT

Overall Average RT and Condition Average RT

We modeled the deviations using a hierarchical Bayesian model implemented in JAGS (Plummer, 2003). Each subject i's deviation score in each condition j was modeled as a draw from a normal distribution with a task specific mean, μ, and standard deviation, σ. A common prior was set over the task specific parameters as follows:

$$
\begin{aligned}
&\text{deviation}_{i,j} \sim \text{Normal}\left(\mu_j, \sigma_j\right) \\
&\mu_j \sim \text{Normal}\left(M, S\right) \\
&\sigma_j \sim \text{Uniform}\left(0.001, 10\right) \\
&M \sim \text{Normal}\left(0, 1000\right) \\
&\tfrac{1}{S} \sim \text{Gamma}\left(.001, .001\right)
\end{aligned}
\tag{9}
$$

We ran two chains with 500 burn-in samples, followed by 5000 samples keeping every tenth sample. All chains showed good convergence ($\hat{R} \leq 1.1$). Throughout, the bandwidth of the plots was determined by the minimum of .01 and the estimated bandwidth using Silverman's rule of thumb (Silverman, 1986).

Individual Average RTs

We estimated each individual subject's RT separately by assuming that the log RT on each trial k was normally distributed with a task i and item j specific mean and standard deviation. The full model is given as follows:

$$
\begin{aligned}
&\log\left(RT_{ijk}\right) \sim \text{Normal}\left(\mu_{ij}, \sigma_{ij}\right) \\
&\mu_{ij} \sim \text{Normal}\left(M, S\right) \\
&\sigma_{ij} \sim \text{Uniform}\left(0.001, 10\right) \\
&M \sim \text{Normal}\left(6, 10\right) \\
&\tfrac{1}{S} \sim \text{Gamma}\left(.001, .001\right)
\end{aligned}
\tag{10}
$$

We converted the posterior estimates of μ to the scale of the observed RTs by using the following transformation:

$$
m_{ij} = \exp\left(\frac{\sigma_{ij}^2}{2} + \mu_{ij}\right); \tag{11}
$$

We again ran two sampling chains for 500 burn-in samples followed by 5000 samples keeping every tenth sample. All chains showed good convergence ($\hat{R} \leq 1.1$).

Appendix C

Appendix: Hierarchical Bayesian analysis of the sequential effects

In our hierarchical Bayesian analysis, we estimated a group level effect for the relevant item in each specific sequential effect and allowed variation in the intercepts for each subject. For example, to analyze the repetition effect, the group effect was dummy coded to indicate whether the previous item was a repetition or not; analogous coding was used for the pushing and pulling effects in separate analyses. For the RT analyses, the models were estimated by log transforming the RTs and assuming that the log RTs were distributed as a t-distribution, which is more robust to outlying RTs than a Gaussian distribution (J. K. Kruschke, 2013). The log transformation of the RT data means that the estimated coefficient ($\times$ 100%) can be thought of as the percentage difference in the RTs as a function of the sequential effect (see e.g., Gelman and Hill, 2007). For the accuracy analyses, we used a logit link function and modeled the error frequency as a binomial distribution (Gelman & Hill, 2007).

The posterior estimate of the coefficient on the group effect can be used to determine the size of the sequential effect. We set priors over the coefficients to $Normal(0, 1)$. In Tables C1, C2, C3, and C4 we report the 95% highest posterior density intervals and the Bayes Factor (BF; computed using the Savage-Dickey method; Wagenmakers, Lodewyckx, Kuriyal, and Grasman, 2010) that the coefficient on the sequential effect is 0.

Using a Bayesian regression allows a good deal of flexibility in how hypotheses are evaluated. We can not only test the evidence that an effect is not likely to be zero (labeled BF_{10}) but we can also test specific directional effects (labeled BF_{dir}). For the repetition effect, we expect the value to be negative; consequently, we test whether the estimate of the coefficient for the repetition effect is less than zero. For the pushing and pulling effects, we test whether the effects are greater than 0. These tests can be thought of as direct tests of the directional hypotheses and are consequently more informative than the null hypothesis test for evaluating the sequence effect predictions.[6] We provide the Bayes Factors for both

[6]In some cases, we find a high value of BF_{dir} when the value of BF_{10} is less than one. The implication is that there is more evidence that the effect equals 0 than there is that the effect is different than 0, but that that there is more evidence that the effect is in the specified direction than in the opposite direction. While these tests are not completely independently, they are testing different regions of the posterior and

tests in Tables C1, C2, C3, and C4. However, we focus on reporting the directional tests since these follow directly from the predictions of our model (the BF_{dir} column).

Experiment 1

Table C1

Bayesian Linear Regression results for the tests of the sequential effects in Experiment 1.

		Mean RT						Error Rate					
		β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}	β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}
Brightness													
Control	Rep	-0.11	0.02	-0.15	-0.06	> 100	> 100	-0.70	0.34	-1.42	-0.08	3.23	65.67
	Far Same	0.13	0.02	0.08	0.17	> 100	> 100	0.53	0.24	0.06	0.99	2.94	63.52
	Adj Opp	0.04	0.02	-0.00089	0.09	0.14	36.74	-0.07	0.27	-0.61	0.44	0.26	0.70
Correlated	Rep	-0.09	0.02	-0.14	-0.05	> 100	> 100	-0.34	0.34	-1.03	0.28	0.51	5.07
	Far Same	0.04	0.02	-0.01	0.09	0.10	20.28	0.73	0.27	0.21	1.26	8.33	> 100
	Adj Opp	0.07	0.03	0.02	0.11	0.52	> 100	-0.06	0.31	-0.69	0.55	0.30	0.74
Filtering	Rep	-0.11	0.04	-0.18	-0.03	1.45	> 100	0.01	0.39	-0.79	0.73	0.38	0.94
	Far Same	0.07	0.03	-0.0013	0.13	0.21	> 250	0.87	0.32	0.24	1.50	12.50	> 100
	Adj Opp	0.01	0.03	-0.06	0.07	0.03	1.42	0.19	0.35	-0.55	0.85	0.42	2.53
Irrelevant	Rep Diff	0.02	0.05	-0.07	0.11	0.05		-0.72	0.56	-1.84	0.36	1.27	
Dimension	Far Same Diff	-0.02	0.05	-0.11	0.08	0.05		-0.33	0.39	-1.11	0.43	0.53	
Change	Adj Opp Diff	0.02	0.05	-0.07	0.11	0.05		0.22	0.43	-0.63	1.04	0.51	
Saturation													
Control	Rep	-0.21	0.03	-0.27	-0.15	> 250	> 100	-2.05	0.52	-3.13	-1.13	> 100	> 100
	Far Same	0.04	0.03	-0.02	0.10	0.08	10.70	-0.31	0.26	-0.84	0.18	0.51	0.13
	Adj Opp	0.13	0.03	0.07	0.19	> 100	> 100	1.11	0.21	0.69	1.51	> 100	> 100
Correlated	Rep	-0.25	0.03	-0.30	-0.19	> 100	> 100	-1.62	0.48	-2.60	-0.71	> 100	> 100
	Far Same	0.07	0.03	0.01	0.13	0.58	> 100	0.23	0.25	-0.27	0.71	0.39	4.57
	Adj Opp	0.11	0.03	0.05	0.18	7.69	> 100	1.26	0.23	0.81	1.72	> 100	> 100
Filtering	Rep	-0.29	0.04	-0.38	-0.21	> 100	> 100	-1.12	0.48	-2.13	-0.23	8.33	> 100
	Far Same	0.12	0.05	0.03	0.22	1.01	> 100	0.26	0.32	-0.40	0.88	0.44	3.76
	Adj Opp	0.16	0.05	0.07	0.26	20.00	> 100	0.79	0.29	0.22	1.34	12.50	> 100
Irrelevant	Rep Diff	0.09	0.05	-0.0084	0.18	0.26		0.19	0.63	-1.02	1.43	0.66	
Dimension	Far Same Diff	-0.05	0.06	-0.17	0.08	0.08		0.06	0.38	-0.71	0.81	0.39	
Change	Adj Opp Diff	-0.11	0.06	-0.23	0.01	0.30		0.29	0.33	-0.36	0.92	0.48	

Note: Rep - near-boundary repetition effect; Far Same - same-category far-item "pushing" effect; Adj Opp - opposite-category adjacent-item "pulling" effect; Rep Diff - difference in repetition effect across levels of the irrelevant dimension change; Far Same Diff - difference in the pushing effect across levels of the irrelevant dimension change; Adj Opp Diff - differences in the pulling effect across levels of the irrelevant dimension change.

thus can be interpreted independently. In a nutshell, while an effect might be more likely to be zero than either greater or less than zero, it is still possible for an effect to be, for instance, more likely to be negative than positive.

Experiment 2

Table C2

Bayesian Linear Regression results for the tests of the sequential effects in Experiment 2.

		Mean RT						Error Rate					
		β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}	β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}
Line Position													
Control	Rep	-0.13	0.03	-0.20	-0.07	> 100	> 100	-1.19	0.47	-2.18	-0.36	16.67	> 100
	Far Same	0.13	0.03	0.07	0.19	> 100	> 100	0.48	0.27	-0.07	1.01	1.39	23.69
	Adj Opp	0.08	0.03	0.01	0.14	0.40	70.43	0.54	0.28	-0.02	1.08	1.54	32.33
Correlated	Rep	-0.16	0.03	-0.22	-0.10	> 250	> 100	-1.20	0.43	-2.09	-0.42	50.00	> 100
	Far Same	0.10	0.03	0.04	0.17	2.44	> 100	0.61	0.25	0.11	1.07	3.70	> 100
	Adj Opp	0.11	0.04	0.04	0.18	2.33	> 100	0.64	0.26	0.10	1.12	3.85	92.02
Filtering	Rep	-0.20	0.05	-0.30	-0.10	> 250	> 100	-0.45	0.46	-1.39	0.41	0.70	5.08
	Far Same	0.09	0.04	0.01	0.18	0.36	53.05	0.75	0.34	0.08	1.41	3.45	66.80
	Adj Opp	0.08	0.04	-0.01	0.16	0.19	23.69	0.36	0.34	-0.32	1.03	0.64	5.76
Irrelevant	Rep Diff	-0.02	0.06	-0.13	0.10	0.05		-0.61	0.65	-1.88	0.65	0.92	
Dimension	Far Same Diff	0.05	0.05	-0.06	0.15	0.08		0.09	0.38	-0.64	0.83	0.39	
Change	Adj Opp Diff	-0.06	0.06	-0.17	0.05	0.10		-0.36	0.49	-1.36	0.58	0.59	
Saturation													
Control	Rep	-0.23	0.02	-0.27	-0.18	> 100	> 100	-1.92	0.37	-2.73	-1.23	> 100	> 100
	Far Same	0.15	0.03	0.10	0.21	> 100	> 100	0.04	0.20	-0.35	0.43	0.20	1.47
	Adj Opp	0.06	0.03	0.01	0.11	0.27	73.07	0.74	0.17	0.40	1.08	> 100	> 100
Correlated	Rep	-0.26	0.02	-0.30	-0.21	> 250	> 100	-2.02	0.41	-2.88	-1.28	> 100	> 100
	Far Same	0.11	0.03	0.06	0.17	20.00	> 100	0.79	0.17	0.45	1.13	> 100	> 100
	Adj Opp	0.10	0.03	0.04	0.15	5.56	> 100	0.97	0.17	0.62	1.30	> 100	> 100
Filtering	Rep	-0.22	0.03	-0.28	-0.15	> 250	> 100	-1.70	0.44	-2.64	-0.91	> 100	> 100
	Far Same	0.02	0.04	0.12	0.28	> 250	> 100	-0.09	0.31	-0.72	0.50	0.30	0.60
	Adj Opp	0.05	0.04	-0.03	0.12	0.09	8.95	0.70	0.24	0.22	1.17	12.50	> 100
Irrelevant	Rep Diff	-0.003	0.03	-0.07	0.06	0.03		-0.12	0.71	-1.54	1.27	0.70	
Dimension	Far Same Diff	-0.09	0.05	-0.19	0.01	0.20		0.78	0.35	0.09	1.48	4.00	
Change	Adj Opp Diff	0.02	0.04	-0.07	0.10	0.04		-0.17	0.31	-0.78	0.44	0.37	

Note: Rep - near-boundary repetition effect; Far Same - same-category far-item “pushing” effect; Adj Opp - opposite-category adjacent-item “pulling” effect; Rep Diff - difference in repetition effect across levels of the irrelevant dimension change; Far Same Diff - difference in the pushing effect across levels of the irrelevant dimension change; Adj Opp Diff - differences in the pulling effect across levels of the irrelevant dimension change.

Experiment 3

Table C3

Bayesian Linear Regression results for the tests of the sequential effects in Experiment 3.

		Mean RT						Error Rate					
		β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}	β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}
Brightness													
Control	Rep	-0.25	0.03	-0.31	-0.20	> 100	> 100	-2.02	0.54	-3.31	-0.99	>100	>100
	Far Same	0.10	0.03	0.05	0.15	> 100	> 100	0.25	0.28	-0.29	0.80	0.41	4.50
	Adj Opp	-0.02	0.03	-0.08	0.04	0.04	0.28	1.27	0.25	0.76	1.76	>100	>100
Correlated	Rep	-0.20	0.03	-0.26	-0.15	> 100	> 100	-0.32	0.36	-1.07	0.34	0.50	4.63
	Far Same	0.09	0.03	0.04	0.15	4.76	> 100	1.02	0.28	0.48	1.59	100	> 100
	Adj Opp	-0.02	0.03	-0.07	0.04	0.36	0.38	0.23	0.33	-0.42	0.88	0.43	3.07
Filtering	Rep	-0.27	0.03	-0.34	-0.20	> 100	> 100	-1.31	0.56	-2.47	-0.31	14.29	> 100
	Far Same	0.02	0.04	-0.06	0.09	0.47	2.25	-0.14	0.39	-0.93	0.59	0.38	0.57
	Adj Opp	0.05	0.04	-0.02	0.13	0.11	12.51	0.90	0.33	0.26	1.54	16.67	> 100
Irrelevant	Rep Diff	0.08	0.03	0.01	0.14	0.40		0.18	0.78	-1.29	1.75	0.80	
Dimension	Far Same Diff	0.07	0.05	-0.04	0.18	0.13		-0.01	0.47	-0.97	0.88	0.45	
Change	Adj Opp Diff	-0.08	0.04	-0.16	0.00	0.23		-0.53	0.39	-1.31	0.24	0.91	
Saturation													
Control	Rep	-0.12	0.03	-0.17	-0.07	> 100	> 100	-0.64	0.23	-1.12	-0.19	14.29	> 100
	Far Same	-0.05	0.03	-0.11	0.00	0.15	0.03	0.07	0.21	-0.35	0.48	0.20	1.68
	Adj Opp	0.06	0.03	0	0.12	0.22	42.96	1.03	0.19	0.65	1.4	0.71	> 100
Correlated	Rep	-0.06	0.03	-0.11	-0.02	0.56	> 100	-0.87	0.28	-1.42	-0.34	50	> 100
	Far Same	-0.05	0.03	-0.1	0	0.13	0.03	0.01	0.22	-0.42	0.43	0.22	1.09
	Adj Opp	0.05	0.03	-0.01	0.1	0.11	18.7	0.88	0.2	0.48	1.25	>100	>100
Filtering	Rep	-0.05	0.04	-0.13	0.02	0.10	10.24	-1.1	0.41	-1.95	-0.33	20	> 100
	Far Same	-0.06	0.04	-0.14	0.02	0.11	0.09	-0.68	0.32	-1.37	-0.09	3.33	0.01
	Adj Opp	-0.04	0.05	-0.13	0.05	0.07	0.27	1.11	0.26	0.62	1.63	> 100	> 100
Irrelevant	Rep Diff	-0.02	0.05	-0.12	0.07	0.06		0.46	0.50	-0.50	1.47	0.86	
Dimension	Far Same Diff	0.01	0.06	-0.1	0.12	0.06		0.68	0.38	-0.04	1.44	2	
Change	Adj Opp Diff	0.03	0.05	-0.07	0.14	0.07		-0.17	0.3	-0.76	0.41	0.35	

Note: Rep - near-boundary repetition effect; Far Same - same-category far-item "pushing" effect; Adj Opp - opposite-category adjacent-item "pulling" effect; Rep Diff - difference in repetition effect across levels of the irrelevant dimension change; Far Same Diff - difference in the pushing effect across levels of the irrelevant dimension change; Adj Opp Diff - differences in the pulling effect across levels of the irrelevant dimension change.

Experiment 4

Table C4

Bayesian Linear Regression results for the tests of the sequential effects in Experiment 4.

		Mean RT						Error Rate					
		β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}	β	*S.E.*	*L95%*	*U95%*	BF_{10}	BF_{dir}
Line Position													
Control	Rep	-0.13	0.03	-0.18	-0.07	> 100	> 100	-0.64	0.31	-1.28	-0.07	2.78	71.73
	Far Same	0.01	0.03	-0.05	0.07	0.03	2.07	0.93	0.22	0.51	1.35	> 100	> 100
	Adj Opp	0.04	0.03	-0.02	0.1	0.08	9.75	0.45	0.23	0	0.91	1.49	41.55
Correlated	Rep	-0.12	0.03	-0.17	-0.07	> 100	> 100	-1.23	0.34	-1.94	-0.6	> 100	> 100
	Far Same	0.04	0.03	-0.02	0.1	0.08	11.46	0.27	0.22	-0.17	0.7	0.46	7.42
	Adj Opp	0.05	0.03	-0.01	0.1	0.12	22.53	0.84	0.21	0.43	1.24	50	> 100
Filtering	Rep	-0.08	0.04	-0.16	0.01	0.19	26.97	-1.22	0.49	-2.24	-0.32	14.29	> 100
	Far Same	0.05	0.04	-0.04	0.14	0.09	7.21	-0.02	0.34	-0.73	0.61	0.33	0.98
	Adj Opp	0.01	0.04	-0.07	0.1	0.04	1.77	0.58	0.29	0.01	1.13	1.89	41.55
Irrelevant	Rep Diff	-0.01	0.05	-0.11	0.08	0.05		0.48	0.61	-0.67	1.69	1.42	
Dimension	Far Same Diff	0.03	0.07	-0.1	0.16	0.07		0.4	0.39	-0.36	1.19	0.65	
Change	Adj Opp Diff	0.05	0.04	-0.04	0.13	0.07		0.07	0.35	-0.62	0.75	0.36	
Saturation													
Control	Rep	-0.11	0.02	-0.15	-0.06	> 100	> 100	-0.96	0.30	-1.55	-0.39	100	> 100
	Far Same	0.10	0.03	0.05	0.15	14.29	> 100	0.73	0.19	0.36	1.1	> 100	> 100
	Adj Opp	0.06	0.03	0.02	0.11	0.76	> 100	0.91	0.19	0.53	1.26	> 100	> 100
Correlated	Rep	-0.09	0.02	-0.14	-0.05	> 100	> 100	-0.98	0.3	-1.59	-0.42	100	> 100
	Far Same	0.11	0.03	0.05	0.16	100	> 100	0.85	0.19	0.46	1.22	> 100	> 100
	Adj Opp	0.12	0.03	0.06	0.18	14.29	> 100	0.88	0.20	0.48	1.25	> 100	> 100
Filtering	Rep	-0.06	0.04	-0.13	0.01	0.13	22.12	-0.36	0.32	-1.03	0.25	0.68	6.59
	Far Same	0.13	0.04	0.05	0.20	11.11	> 100	0.40	0.28	-0.14	0.94	0.78	11.27
	Adj Opp	0.08	0.04	0.01	0.16	0.46	73.07	1.46	0.26	0.95	1.95	> 100	> 100
Irrelevant	Rep Diff	0.01	0.05	-0.09	0.10	0.05		-0.81	0.52	-1.88	0.16	1.64	
Dimension	Far Same Diff	0.00	0.05	-0.11	0.10	0.05		0.30	0.32	-0.31	0.94	0.48	
Change	Adj Opp Diff	-0.01	0.05	-0.11	0.10	0.05		-0.51	0.32	-1.12	0.12	1.23	

Note: Rep - near-boundary repetition effect; Far Same - same-category far-item “pushing” effect; Adj Opp - opposite-category adjacent-item “pulling” effect; Rep Diff - difference in repetition effect across levels of the irrelevant dimension change; Far Same Diff - difference in the pushing effect across levels of the irrelevant dimension change; Adj Opp Diff - differences in the pulling effect across levels of the irrelevant dimension change.

Appendix D

Appendix: Bayesian implementation of Sequence-sensitive Exemplar and Feature-based models

Fitting procedure

Before fitting the data, we first subtracted each individual's average RT and added in the average RT for the group. All RTs were fit in seconds; hence, the parameters can be interpreted on that scale. We set informative priors over these parameters based on (a) our expectations about the likely values of the parameters and (b) prior application of the model (e.g., Little et al., 2016). The prior parameter settings are shown in Table D1. Posterior parameter distributions were estimated using Differential Evolution-Markov Chain Monte Carlo (DE-MCMC; Turner, Sederberg, Brown, & Steyvers, 2013).

We estimated each experiment separately by combining the subjects from each condition into a single hierarchy. The estimates fit to each condition separately look similar; hence, we only report the combined estimates in the interest of brevity. We sampled 48 chains (3×16 parameters) for 3750 burn-in samples. From burn-in sample 2501 to 3250, we performed a migration step (Turner et al., 2013) every 20 trials. We then took an additional number of samples (between 50,000 and 80,000; with a migration step probability of .01) keeping every 100th sample (or 150th sample) until the $\hat{R}$-statistic was less than 1.1 for all chains.

Table D1
Prior parameter settings for the sequence-sensitive models

Subject Level	
Parameter	Prior Distribution
w, w_{cr}	Beta(a_w, b_w)
c, c_f	Normal(μ_c, σ_c)
b_1, b_2, b_3, b_4, b_5	Beta(a_b, b_b)
β	Normal$(\mu_\beta, \sigma_\beta)$
α	Beta(a_α, b_α)
A	Normal(μ_A, σ_A)
$T_A - A$, $T_B - A$	Normal$(\mu_{T-A}, \sigma_{T-A})$
s_ν	Normal$(\mu_{s_\nu}, \sigma_{s_\nu})$
t_0	Normal$(\mu_{t_0}, \sigma_{t_0})$
Group Level	
Parameter	Prior Distribution
a_w	Normal$(1, 1)$
b_w	Normal$(1, 1)$
μ_c	Normal$(1, 2)$
a_b	Normal$(1, 1)$
b_b	Normal$(1, 1)$
μ_β	Normal$(1.2, 2)$
a_α	Normal$(1, 1)$
b_α	Normal$(1, 1)$
μ_A	Normal$(.05, 1)$
μ_{T-A}	Normal$(.25, .5)$
μ_{s_ν}	Normal$(.22, 1)$
μ_{t_0}	Normal$(.1, 1)$

Note: Each subject level parameter was given its own group level prior; however, for brevity, when the priors have the same distributions and same parameters, we list them together omitting the parameter specific subscript. w and w_cr index attention to the irrelevant dimension. w_{cr} refers to the attention weight in the correlated condition. The a_w prior for the w_c parameter was set to Normal$(5, .5)$. c_f refers to the sensitivity parameter in the filtering condition. $b_1, b_2, b_3, b_4,$ andb_5 are the biases for category A for the control-1, control-2, correlated-1, correlated-2, and filtering tasks, respectively. Thresholds for categories A and B were parameterized as the distance between the threshold and the starting point. All standard deviation parameters, σ, were given Gamma$(1, 1)$ priors.

Appendix E

Appendix: Posterior Group Parameter estimates

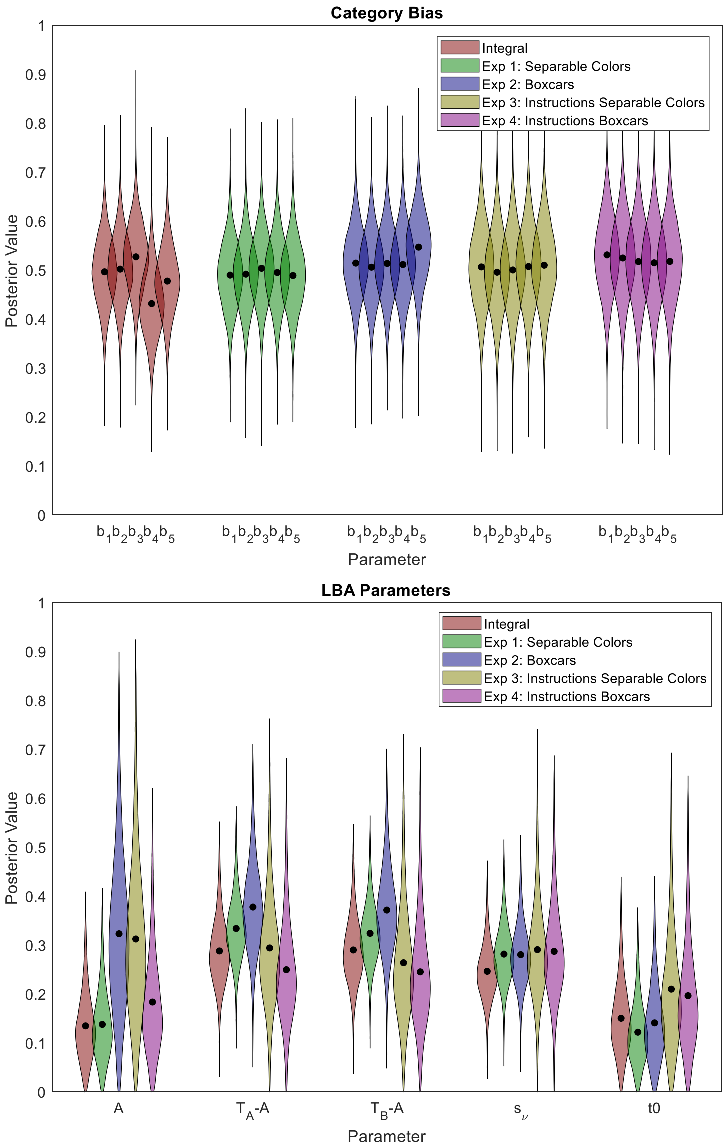

Figure E1. Top panel: Posterior densities for the group level mean bias parameters for the sequence-sensitive exemplar-based. Bottom panels: Group level mean LBA parameters for the sequence-sensitive exemplar-based model.

Appendix F

Appendix: Posterior Predictions

Integral dimension data from Little et al. (2016)

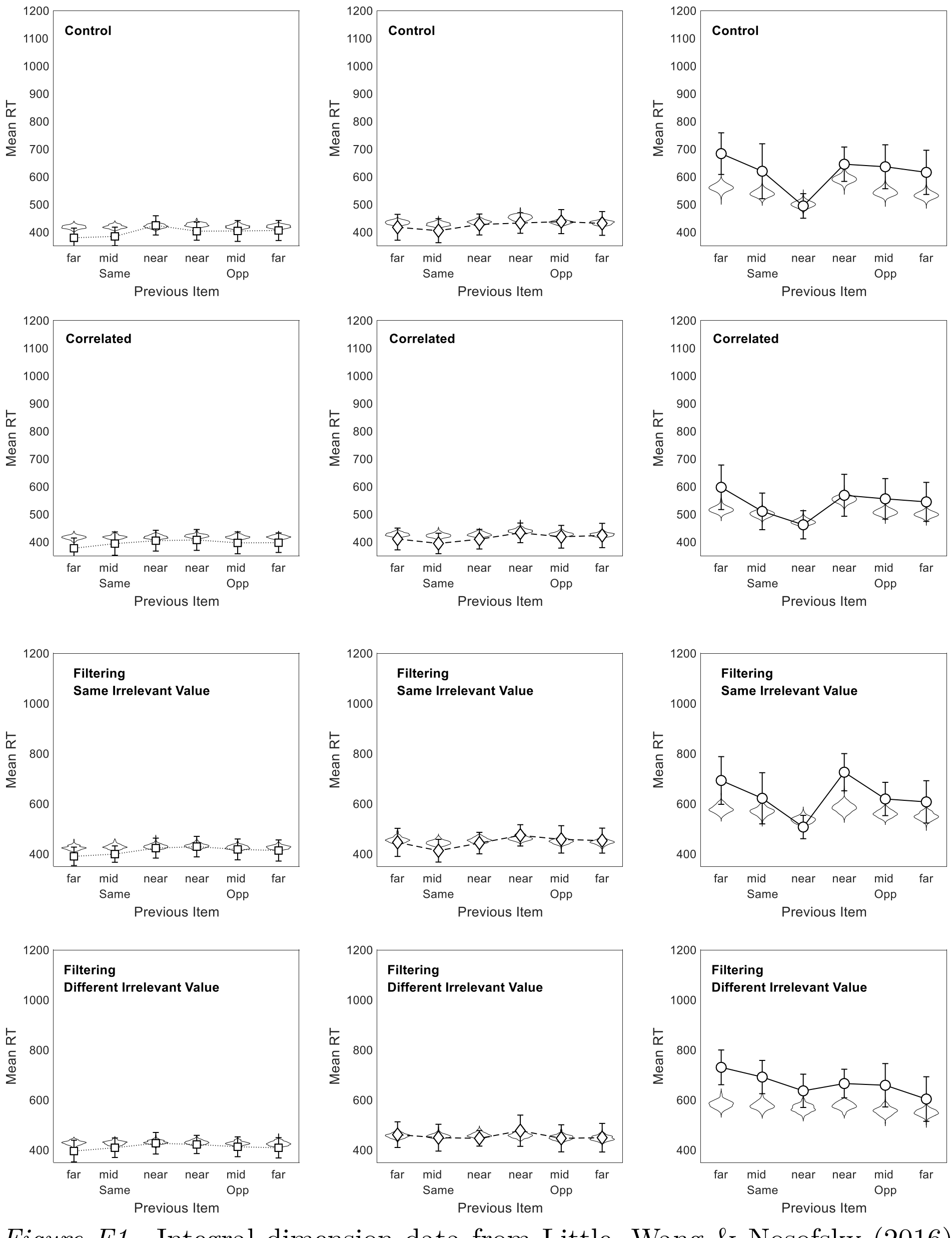

Figure F1. Integral dimension data from Little, Wang & Nosofsky (2016). Posterior predictive distributions for RTs in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

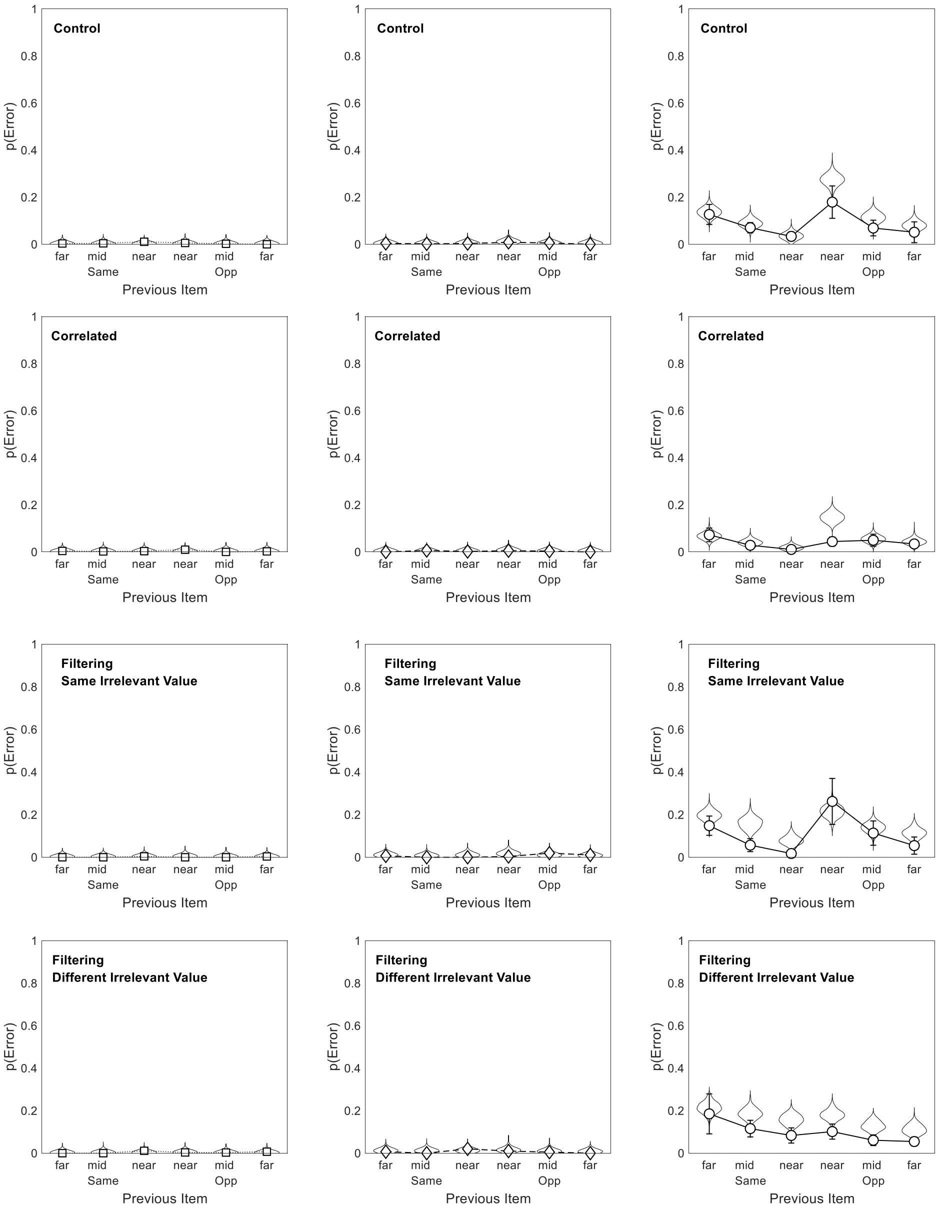

Figure F2. Integral dimension data from Little, Wang & Nosofsky (2016). Posterior predictive distributions for accuracy in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

Experiments 1 and 2: Separable dimension accuracy data

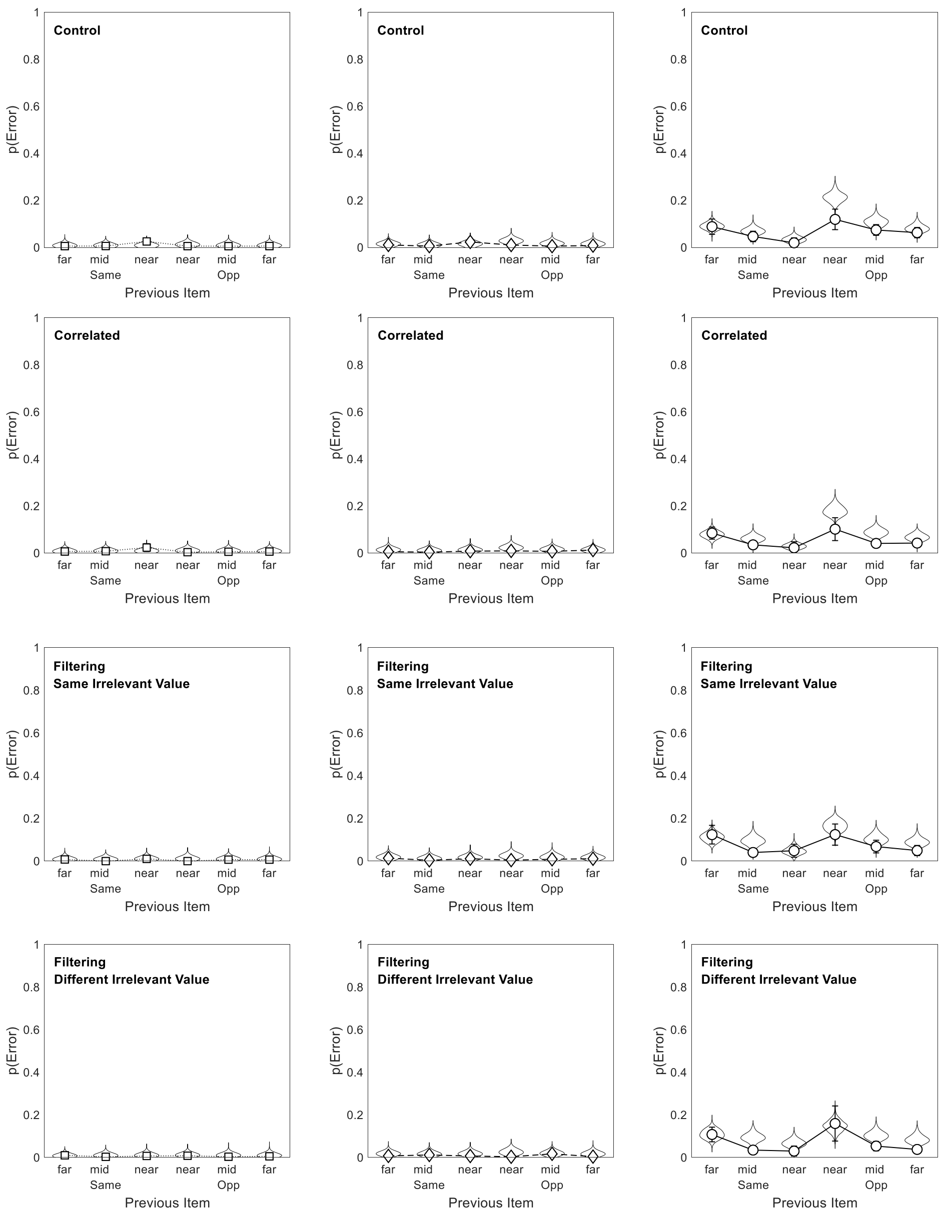

Figure F3. Separable Color data from Experiment 1. Posterior predictive distributions for accuracy in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

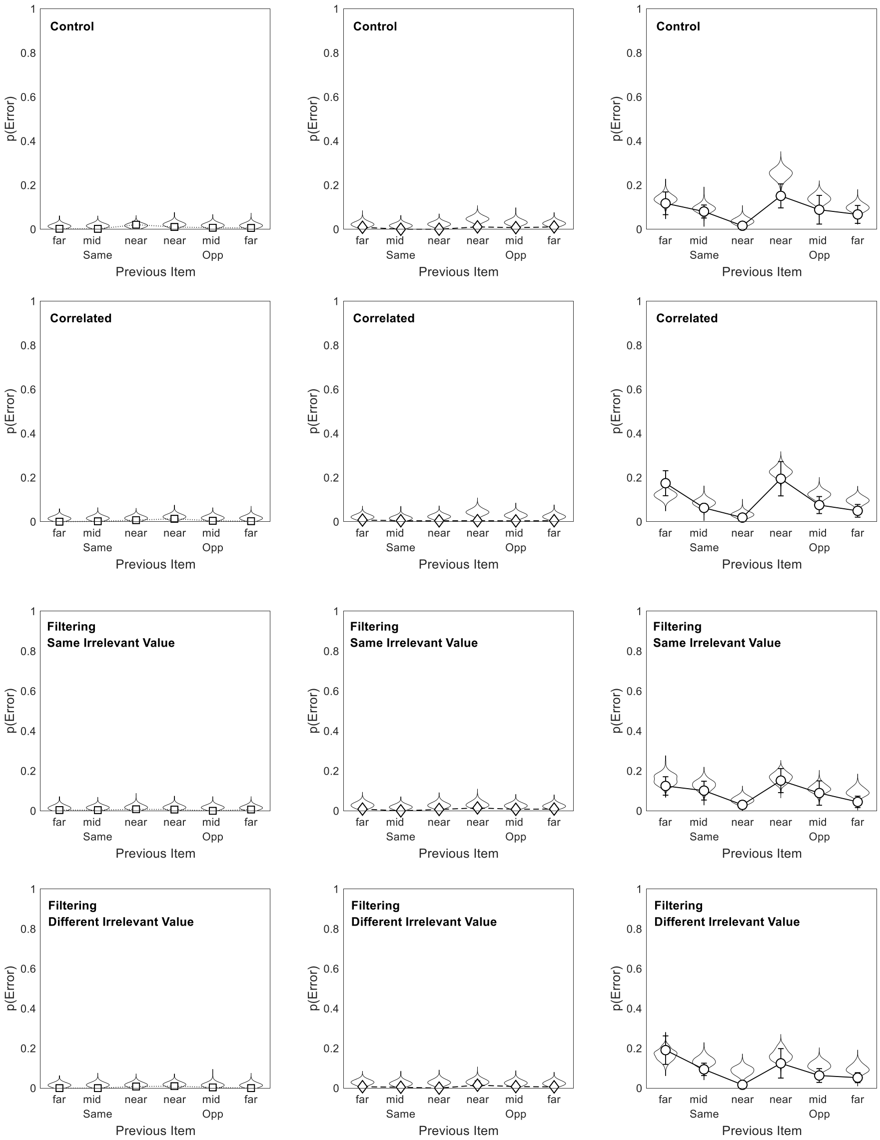

Figure F4. Separable Boxcar data from Experiment 2. Posterior predictive distributions for accuracy in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

Experiments 3 and 4: Attentional Instructions Separable dimension accuracy data

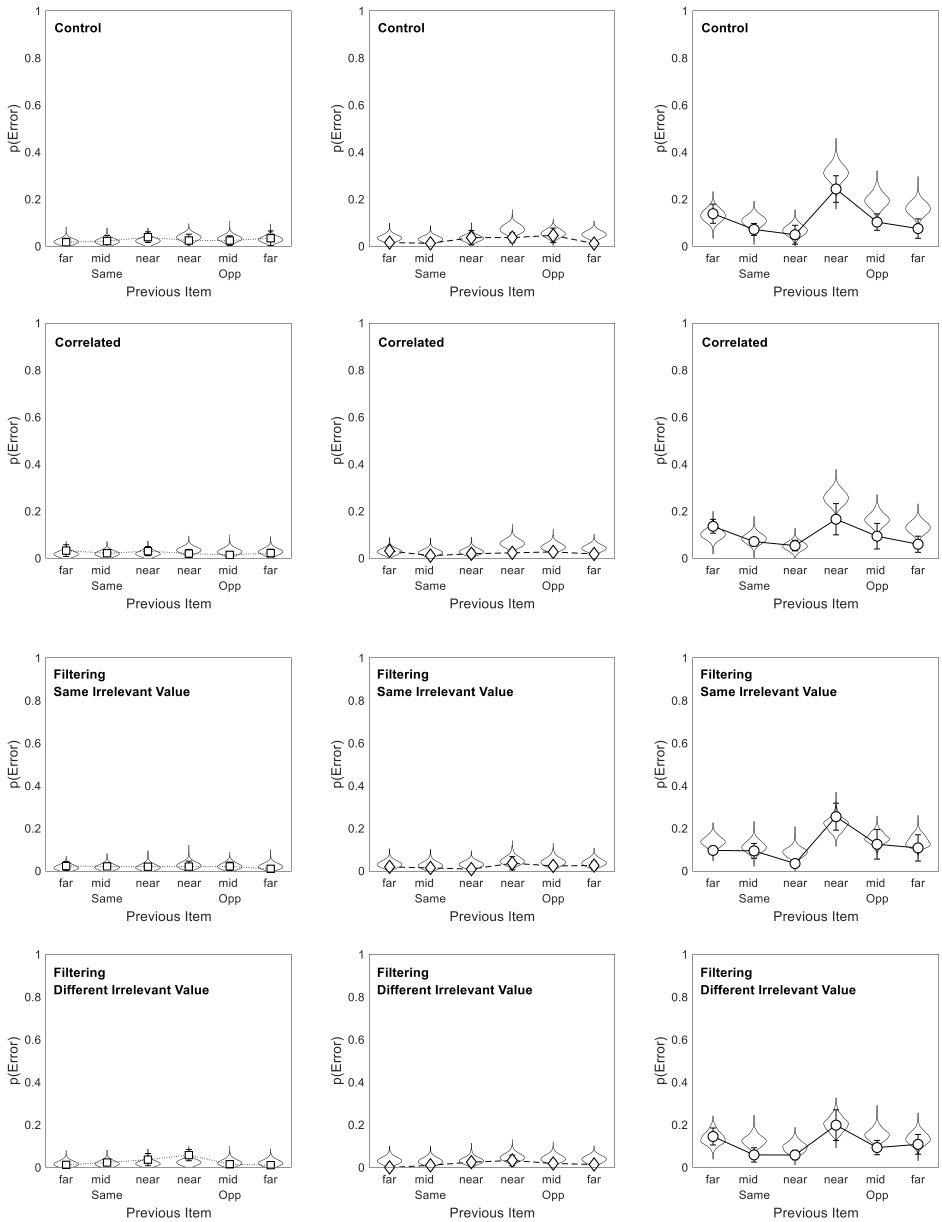

Figure F5. Separable Color data from Experiment 3. Posterior predictive distributions for accuracy in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.

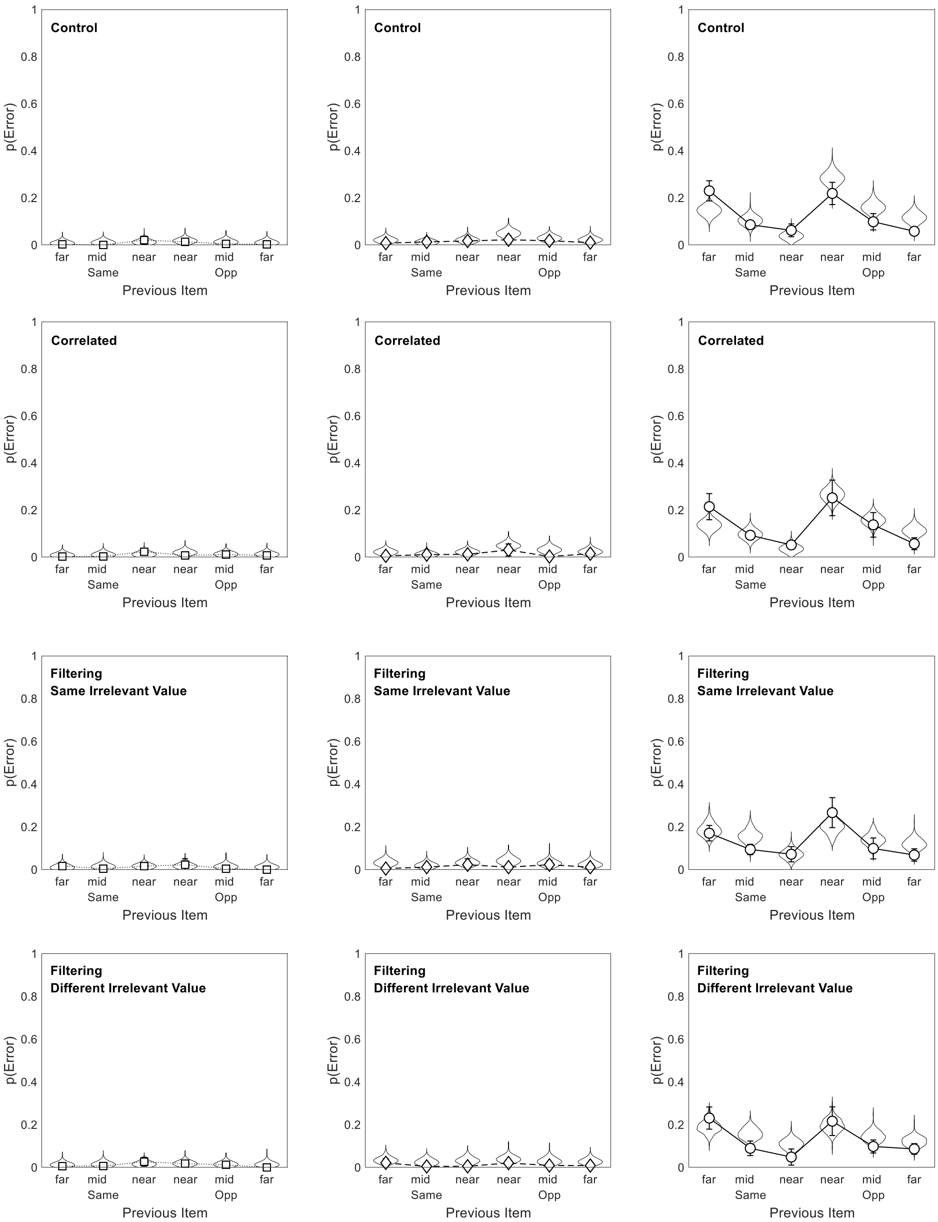

Figure F6. Separable Boxcar data from Experiment 4. Posterior predictive distributions for accuracy in the control (top row), correlated (second row), and filtering tasks (bottom two rows). The left hand column shows the far items, the middle column shows the middle item, and the right hand column shows the near boundary item.