

Contrast and assimilation in categorization and exemplar production

Vladimir Zotov · Michael N. Jones · D. J. K. Mewhort

Published online: 17 November 2010

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence 2010

Abstract We studied contrast and assimilation in three tasks: an exemplar-production task, a categorization task, and a combined categorization-then-production task. On each trial of the first task, subjects produced a circle when prompted with a category label. In the second task, they classified lines that differed in length into one of four categories. On each trial of the combined task, they classified two lines and then produced a line when prompted by a category label. All three tasks converged on the same conclusion: subjects' representation of the categories (measured in pixels in the production tasks and by the direction of errors in classification) shifted systematically from trial to trial. When successive stimuli were from the same category, the representation of that category was pulled toward the exemplar from the previous trial. When successive stimuli were from different categories, the representations of the

neighbouring categories were pushed from the category of the initial stimulus. We conclude that accounts of categorization and identification must accommodate both assimilation and contrast as a function of trial-to-trial shifts in representation.

Keywords Classification · Categorization · Assimilation · Contrast · Production · Representation · Practice

Introduction

In an absolute-identification task, subjects are asked to label stimuli that differ on at least one dimension. In a corresponding categorization task, the stimuli are grouped into at least two categories, and subjects are asked to label the category to which each stimulus belongs. Hence, identification requires a one-to-one mapping of a stimulus to its label, whereas classification requires a many-to-one mapping.

Systematic trial-to-trial errors have been well documented in the absolute-identification literature. Subjects err by responding as if the current stimulus is closer to the previous stimulus than it actually is—a bias called *assimilation* (Garner, 1953; Holland & Lockhead, 1968; Hu, 1997; Lacouture, 1997; Lockhead, 1984; Luce, Nosofsky, Green, & Smith, 1982; Purks, Callahan, Braid, & Durlach, 1980; Ward & Lockhead, 1970, 1971).

Standard accounts of absolute identification explain assimilation in terms of a shift in the subject's internal representation of each item on the defining dimension. Internal representations for the other items are shifted *away* from the stimulus that has just been identified. The shift is discussed in terms of prototypes (Petrov & Anderson, 2005), decision boundaries (Braid & Durlach, 1972; Purks et al., 1980; Treisman, 1985; Treisman & Williams, 1984), or the category boundaries that define the stimulus (Luce et

The Canadian Government grants Springer Science+Business Media non-exclusive permission to use the copyrighted article in any way desired, as long as its source is acknowledged. This includes giving to others permission to translate and to reproduce the article in any form, providing that its source, the author, and Defence R&D Canada are clearly indicated.

V. Zotov (✉)
Defence R&D Canada,
P.O. Box 2000, 1133 Sheppard Ave West,
Toronto, Ontario M3M 3B9, Canada
e-mail: Vladimir.Zotov@drdc-rddc.gc.ca

M. N. Jones
Indiana University,
Bloomington IN, USA
e-mail: jonesmn@indiana.edu

D. J. K. Mewhort
Queen's University at Kingston,
Kingston, ON, Canada
e-mail: mewhortd@queensu.ca

al., 1982). The working assumption is that assimilation reflects a *shift* in the subject's memory for the items. Our notion of shift in memory for items, or *shift in representation* of category is a general term that will be used to describe our performance findings; it is neutral in terms of endorsing any specific representation (e.g., based on prototypes, decision boundaries, response criteria, etc.) that might underlie empirical phenomena we present in this paper. Thus, our theoretical explanation should not be treated as a model, but rather as a general framework.

To illustrate how the shift idea accounts for assimilation, Fig. 1 (top panel) shows an initial state in which the physical and psychological scales are aligned. When a subject is shown Stimulus 1, his/her internal representation of the items is *pushed away* from Stimulus 1. The middle panel shows the consequences for the internal representation, and the bottom panel shows how a subsequent stimulus—Stimulus 5 in the example—could be misidentified. Because its position will be judged relative to the shifted psychological scale, Stimulus 5 appears to be closer to Position 4 than it really is. Regardless of the underlying assumption of shift mechanism, the responses should reflect the current represen-

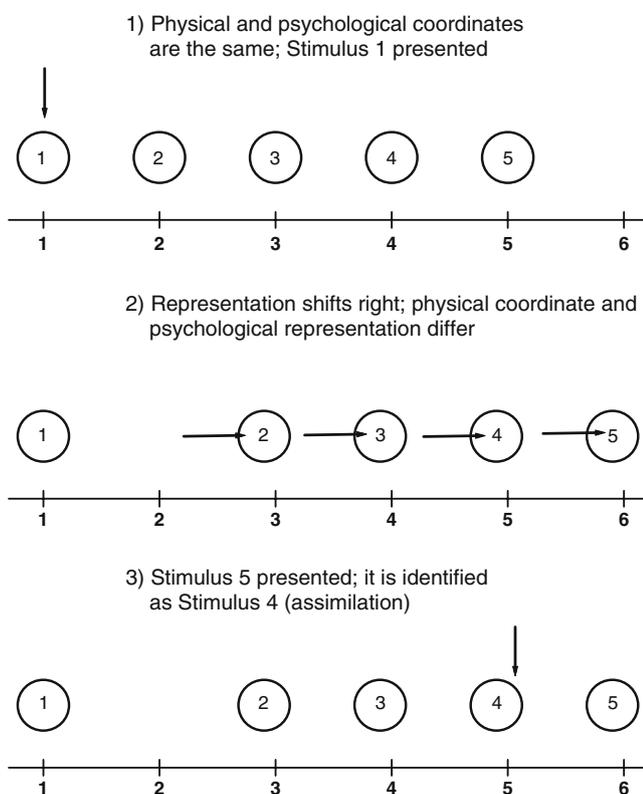


Fig. 1 An illustration of assimilation in identification. The *top panel* shows the physical (*abscissa*) and psychological scales (*circles*) aligned. The *middle panel* shows a shift in the psychological representation induced by the presentation of Stimulus 1. The *bottom panel* shows why Stimulus 5 is misidentified as Stimulus 4 when it is compared against the shifted representation of the scale

A mapping of categories in Stewart et al. (2002) experiments

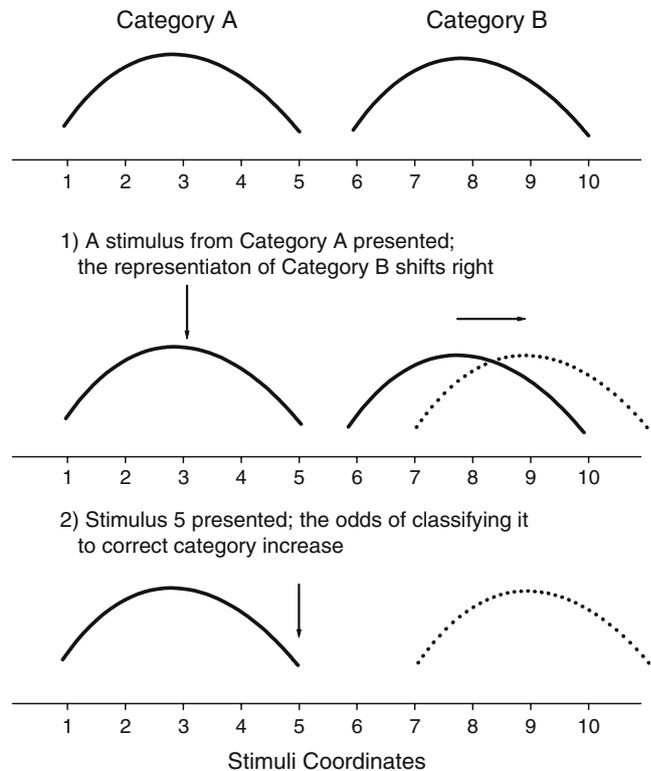


Fig. 2 The *top panel* diagrams the category structure used by Stewart et al. (2002). The *middle panel* shows a shift in representation parallel to the shift found in identification. The *bottom panel* illustrates the paradox raised by Stewart et al. by showing the counter-factual prediction derived from shift in representation associated with identification tasks

tation of items. Thus, when the representation is pushed away from the current exemplar, the next item will be judged to be closer than it actually is.

Although assimilation in absolute identification is well documented, systematic trial-to-trial errors have not received comparable attention in the categorization literature. Recent studies, however, have shown that subjects in a categorization task respond as if the current stimulus is further from the category of the previous stimulus than it actually is—a bias called *contrast* (Jones, Love, & Maddox, 2006; Stewart & Brown, 2004; Stewart, Brown, & Chater, 2002). Figure 2 (top panel) sketches Stewart et al.'s (2002) experiment. Ten tones, spaced equally in psychological space, were divided into two categories, A and B. When subjects classified a borderline stimulus from the same category as the previous trial, they often misclassified it as an exemplar from the wrong category; for example, they often miscategorized Tone 5 following Tone 1 as an example from Category B. By contrast, when subjects classified a borderline stimulus from a different category

(e.g., Tone 5 following Tone 10), the responses were significantly more accurate. Such errors indicate that the subjects perceived the second stimulus as if it were further from the previous stimulus than it actually was—a contrast effect.

Why should *assimilation* occur in identification whereas *contrast* occurs in categorization? Stewart et al. (2002) summarize the paradox nicely: “at present, [there is] no account of why contrast effects are observed in categorization but assimilation effects are observed in absolute identification” (p. 10). A main difference between the two tasks concerns the mapping of stimuli to responses, but it is hard to imagine how a change in the mapping could convert a systematic error in one direction into a systematic error in the other.

Nevertheless, the difference in mapping makes the two tasks differ in one important respect. While successive trials in classification and absolute identification tasks include both cross-category and same-category transitions (between-item and same-item in absolute identification), only classification includes successive trials in which same-category transitions can include different items. The same-item transitions in the identification task are stimulus repetitions. The difference in sequence effects might not reflect mapping per se, but rather the way the mapping affects different types of trial-to-trial transitions: a same-category transition and a different-category transition.

One way to interpret Stewart et al. (2002) findings of contrast is in terms of shifts in representation that we outlined above. If shifts of representation in classification are like those in identification, Stewart et al.’s results present an interesting dichotomy. The middle and bottom panels of Fig. 2 illustrate the dichotomy. The initial exemplar is from Category A (Tone 1). If category B shifts away from category A—as occurs in identification—when Tone 5 is presented, it should be judged against the shifted representation of the categories, and misclassifications to the wrong category should *decrease*. Stewart et al. reported an increase; hence, on their evidence, the shift should be in the opposite direction.

The example illustrated in Fig. 2 relies on the assumption that representational shifts in classification are of the same nature as they are in identification. Considering that there are two types of same-category trial-to-trial transitions in classification (same- vs different-category), it is reasonable to propose that classification has different shifts in representation than absolute identification. To investigate how the different types of category transitions are reflected in the subject’s representation of categories and how these differences might account for differences in sequence effects, we conducted a series of three experiments. The experiments studied changes in representation by using standard measures of performance as well as a production

technique designed to reveal each subject’s changes in representation.

The production task

Exemplar generation is a standard way to assess a subject’s representation of category structure with a long history in studies of semantic categorization (Rosch, 1973), and memory distortion (e.g., Zangwill, 1937).

More recently, Busemeyer and Myung (1988) used a production task to study trial-to-trial changes in prototype representation. Subjects were shown a sequence of exemplars generated from one or more prototypes. The exemplars consisted of random dot patterns used by Posner and Keele (1968, 1970). After observing each exemplar, subjects were asked to reproduce (graphically or numerically) their current estimate of a dot pattern of each prototype. Unlike a standard categorization task, in which learning must be inferred from the percentage of correct decisions pooled across blocks of trials, this production task provided a direct and immediate view of the evolution of a prototype over time. The stimulus-production task has also been used in magnitude-estimation. For example, subjects in DeCarlo and Cross’s (1990) cross-modality matching task used a digital stylus to adjust a line length so that it matched the loudness of an auditory stimulus.

The production studies have yielded good correspondence between production data and other measures. For example, in Zangwill’s (1937) experiments, both recognition and production induced shifts from the standard in the same direction. Busemeyer and Myung’s (1988) data showed close correspondence between the deviations from the prototype in a production task and the accuracy data obtained in a standard classification task (i.e., Posner & Keele, 1968, 1970); DeCarlo and Cross’s (1990) study also revealed good correspondence between production and estimation tasks.

Even though production tasks have been used to study trial-to-trial changes in representation, they were not designed to analyze how changes in production are related to contrast and assimilation. We adopted a production procedure to investigate shifts in representation. In our task, we asked subjects to produce exemplars of particular categories by clicking-and-dragging a mouse to create a stimulus that perceptually matched the size of the category they were asked to produce. Each exemplar produced is, in effect, a snapshot of the subject’s momentary understanding of the category structure. Evidence for a shifting representation that has been inferred from the production task’s data may or may not have a direct relation to subjects’ representation. Nevertheless, if trial-to-trial variations in production correspond to trial-to-trial data using other performance indices, the claim would become stronger.

An overview of the experiments

Experiment 1 used a production task to measure trial-to-trial shifts in the subject's representation of the category structure. In the production task, subjects were given a category label and were required to produce an exemplar of the category on a computer monitor by moving the computer's mouse. The categories were defined using circles of different sizes. We found that there were trial-to-trial shifts in the size of produced exemplars; the subject's representation of the categories was altered by the context of the immediately preceding trial. The second experiment confirmed that the pattern of errors in a standard classification task is consistent with the trial-to-trial shifts in the production task. The third experiment combined classification with production so that the exemplars produced by the subjects would provide a direct measure of any shift in the subject's representation induced by classifying a stimulus.

Experiment 1: using production to measure shifts of representation

Experiment 1 had four equally spaced categories bounded on both sides. On each trial, subjects were asked to produce an example of a named category. Suppose that the subject's internal representation does not shift systematically from trial to trial. On that assumption, successive exemplars should be close to the center of the required category, with random deviations (likely Gaussian) about the center of the category. Alternately, if systematic trial-to-trial shifts in their representation occur, successively produced exemplars should be biased from the center of the required category as a function of the preceding stimulus.

Because presentation of categories was randomized, the experiment included examples of both cross-category and same-category trial-to-trial transitions.

Method

Subjects Eight graduate students from Queen's University participated in the experiment. They were each paid \$10 for participation and all reported normal or corrected-to-normal vision.

Procedure Subjects produced circles by clicking-and-dragging a computer mouse on a digital canvas that measured 19 cm square. The circles appeared as black, filled objects on the white canvas. The canvas was presented at the center of a 19-inch (c.48.3-cm) CRT monitor with a resolution of 1,024 × 768 pixels. At a viewing distance of 60 cm, the digital canvas subtended a maximum visual angle of about 18°, horizontally and vertically. Circles could be drawn anywhere inside the digital canvas, and clicking-and-dragging the optical mouse always produced a perfect circle.

There were four categories of circles (A, B, C, D), each containing 51 possible exemplars defined by their diameter in pixels. The range of category space was from 100 to 300 pixels. We will refer to the central exemplar of a category as the category's prototype.

On each trial, the subject was prompted with a category label that was displayed on the screen (e.g., A) and was required to produce an exemplar that belonged to the category on the canvas. After producing the exemplar, the subject clicked on a rectangle labeled *classify*, and the computer provided the category label to which the produced exemplar belonged (e.g., B). An exemplar that was smaller than the lower boundary of the smallest category (A), or was larger than the upper boundary of the largest category (D), received an out-of-range label (?) as feedback. Once feedback had been given, a subject could not amend the produced exemplar. Subjects used the feedback to learn the categories. To start the next trial, the subject clicked on a rectangle labeled *next*. Each trial began with a clear canvas.

Subjects completed 12 blocks of 40 trials, 120 trials for each of the four categories. The order in which successive categories were requested was determined randomly for each subject.

Results

Cross-category transitions To analyze the effect of the preceding category on the production of the current category, we calculated mean size of the exemplar produced as a function of the preceding category. Table 1 shows the

Table 1 Mean production for each category in pixels as a function of the preceding category in **Experiment 1**

		Present Category				Mean	Slope
		A	B	C	D		
Previous Category	A	125.6	174.9	229.4	277.0	201.8	-1.93
	B	120.9	177.1	231.9	278.8	202.2	
	C	121.6	169.8	228.3	278.2	199.5	
	D	118.1	169.5	224.1	273.1	196.2	

means of productions of each category. Overall, there was a negative relation between mean production and the category presented on the previous trial; the linear contrast was significant, $F(1, 7) = 15.86, p < .01$.

Although subjects produced responses very close to the center of each category, successive responses were biased away from the category of the immediately preceding trial. To illustrate the bias, Fig. 3 shows smoothed distributions of exemplars for Category B (which is flanked by both a smaller category, A, and a larger category, C). The solid distribution (center) shows all responses to Category B. The dashed distribution (on the left) shows the subset of Category B responses preceded by the immediately larger category, C. The dotted distribution (on the right) shows the subset of Category B responses preceded by the immediately smaller category, A. As is evident in the figure, the mean of Category B responses was biased away from the preceding category.

Same-category transitions There are four cases in which subjects were asked to produce the same category twice: the AA, BB, CC, and DD transitions. When the same category was required on successive trials, the second *exemplar* of the pair was biased away from the center of the category in the same direction as the *exemplar* from the previous trial, and the deviation from the center of the category for the two productions was highly correlated, $r(30) = .87, p < .001$. We interpret the correlation as evidence that the internal representation used for the second production has shifted towards the first.

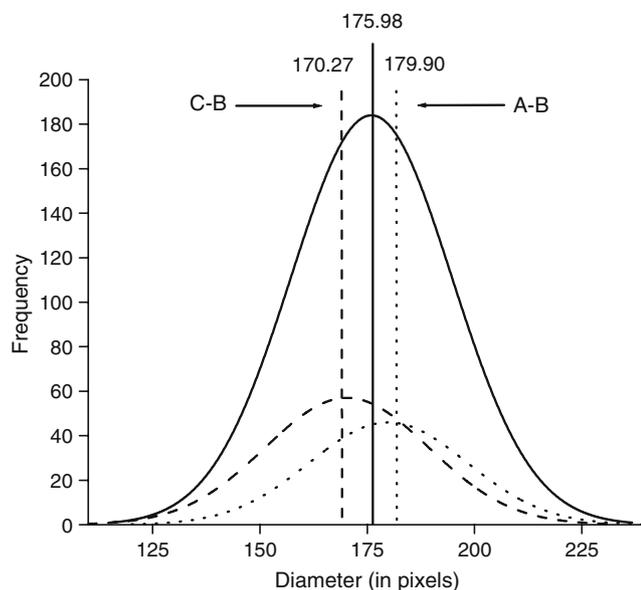


Fig. 3 Smoothed frequency distributions for the diameters of produced exemplars for trials preceded by a smaller or a larger category (Category B only)

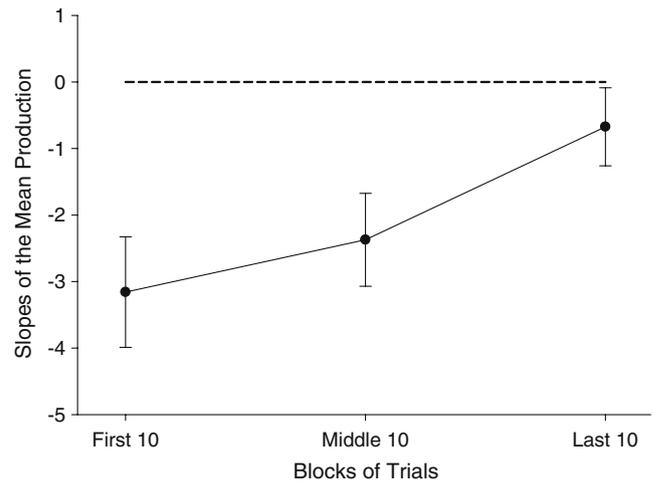


Fig. 4 The slopes of the mean production function across the whole Experiment 1

Sequence effects across the experiment Subjects learned the categories quickly. Mean proportion correct ranged from .79 in the first block to .91 in the final block; accuracy increased linearly across blocks, $F(1, 7) = 6.44, p < .05$. Figure 4 presents the slopes of the mean production function across the entire experiment.¹ The slopes show the relationship between mean production and the category presented on the previous trial as in the analysis of cross-category transitions. The absolute value of slopes of the function decreased linearly, indicating that the cross-category push in representation has been reduced with practice; the linear trend within the last third of trials was not significant, $F(1, 7) = 1.54, p > 0.05$.

Discussion

Experiment 1 demonstrated a complex picture of trial-to-trial transition effects. When asked to produce an example from the same category, responses were pulled towards the exemplar on the preceding trial. When asked to produce an example from a different category, responses were pushed away from the preceding category.

The biases in successive exemplars indicate two kinds of shift in each category’s representation: one is a *shift away* from a preceding *category* for different-category cases, and

¹ There were approximately 30 combinations of each of the possible $(n - 1) - n$ category pairs across the experiment; we used the first 10, next 10, and the last 10 of each pair for our analysis. Because the sequences of trials were randomly generated, the total number of each combination of pairs varied. In case there were less than 30 $(n - 1) - n$ pairs, there was a little bit of overlap between the second and the last 10 pairs. In case there were more than 30 $(n - 1) - n$ pairs, there was a small gap between the second and the last 10 pairs.

one is an adjustment in the direction of the preceding *exemplar* within the same category.

What do the shifts in representation imply for categorization and identification? The shifts in representation illustrated in the *between-category* transitions predict *assimilation* in a corresponding categorization task. In Fig. 1, we showed how assimilation in identification can be produced by a systematic shift in representation. The same logic applies to the category representation in the cross-category transitions: the systematic shift in cross-category responses implies that the subject's representation on the second of the two trials was shifted away from the category of the previous stimulus. If subjects were asked to categorize the stimuli, the shift should produce systematic errors of the sort illustrated in Fig. 1 for identification. The production data for cross-category transitions are consistent with the well-established pattern of assimilation found in identification experiments.

The same-category data, however, document a very different pattern: The second production of each pair was shifted away from the center of the category in the direction of the first production. Using the same logic that we used previously (Fig. 1), if subjects had been asked to categorize the stimuli, the shift should produce *contrast*. The production data for same-category transitions are consistent with Stewart et al. (2002) report of contrast in the two-category classification task. There is another possible interpretation for the same-category transitions, however. When required to produce successive exemplars from the same category, a subject may try to copy the first when making the second. That is, the correlation may reflect a response bias. Experiments 2 and 3 address this issue in detail; it is worth noting here that, in Experiments 2 and 3, the subjects' responses will be measured against a fixed set (3 stimuli per category) of the previous stimuli from the same category.

Experiment 2: does bias in production predict errors in classification?

Experiment 1 documented two systematic trial-to-trial shifts in representation. We take production responses to be direct measures of the subject's representation of the category structure; hence, the shifts in production indicate corresponding shifts in the representation. Shifts in representation, in turn, should produce *assimilation* for cross-category transitions and *contrast* for same-category transitions in a classification task. Experiment 2 was designed to assess whether the shifts of representation documented in Experiment 1 predict errors in classification.

Figure 5 illustrates the shifts in representation suggested by the production data from Experiment 1. The black arcs

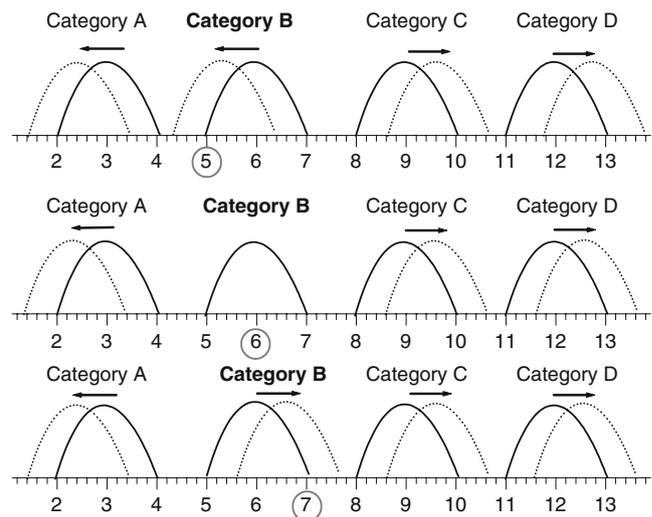


Fig. 5 A schematic diagram illustrating the shifts in representation found in Experiment 1. The black arcs show static representations of categories, and the dotted arcs show the shifted representation of categories

indicate the objective locations of the categories. The dotted arcs show the shifted representation after presenting Item 5 (top panel), Item 6 (middle panel), or Item 7 (bottom panel) of Category B. Figure 5 is not drawn to scale; it is intended as a schematic to illustrate predictions.

After presenting an item from one category, the other categories are pushed away. The fate of the initial category depends on the location of the stimulus within the initial category. Figure 5 shows the situation when the initial stimulus is in Category B. As is illustrated in the figure, when any of the three items in Category B are presented, the representation of Categories A, C, and D are pushed away from Category B—the between-category push documented in Experiment 1. Category B, itself, is pulled towards the initial item—the same-category pull documented in Experiment 1; hence, when Item 5 (the leftmost item in Category B) is the initial stimulus, Category B is shifted left, as shown in the top panel. When Item 6 (the middle item within Category B) is presented, Category B does not shift. Finally, when Item 7 (the rightmost item in Category B) is presented, Category B is shifted to the right, as shown in the bottom panel of Fig. 5.

What are the consequences of a between-category push on classification? When peripheral Items 4, 8, or 11 (of Categories A, C, and D, respectively) follow Item 5, their physical position should be closer to the shifted locations of the neighboring categories (B and C); as a result, errors are likely to be in the direction of assimilation. Classification of items central to the categories (Items 3, 9, and 12) and items at the periphery of the categories at the side opposite to the shift (Items 2, 10, and 13 of Categories A, C, and D, respectively), by contrast, should yield few errors because

the shifted representation of the categories remains consistent with the physical locations.

Next, consider the consequences of the same-category adjustment. Figure 5 shows shifts in the location of Category B following Items 5, 6, or 7. If Item 7 follows Item 5, its location should be almost equidistant to Categories B and C; as a result, subjects are likely to err by misclassifying Item 7 as a member of Category C. By contrast, if Items 5 or 6 follow Item 5, because both Items 5 and 6 fall within the shifted location of Category B, few errors should occur. Likewise, if Item 5 follows Item 7, subjects are likely to misclassify it because its shifted position is equidistant to Categories A and B.

In summary, we expect two independent transition effects for same- and different-category cases. For different-category transitions, the internal representation should shift *away* from the preceding category. As a result, subjects should make *assimilation* errors. For the same-category transitions, however, the internal representation of the category either should not change or should shift *toward* the preceding stimulus. When a shift occurs, the subjects should make *contrast* errors. The data from [Experiment 1](#) suggest that these two opposing forces are simultaneously operating on the memory representations for the categories.

Method

Subjects Nineteen undergraduate students from Queen's University participated in the experiment. They were each paid \$10 for participation, and all reported normal or corrected-to-normal vision. All were right-handed, and the experiment took about an hour to complete.

Apparatus and stimuli The stimuli were presented on a 14-inch (c.35.6-cm) SVGA monitor using a graphics system with resolution of 640 × 480 pixels. The stimuli were 14 lines spaced equally on a single physical dimension (line length). The shortest line was 40 pixels in length and the longest was 300 pixels. The intermediate lines were constructed at 20 pixel intervals lines presented in white on a dark background. At a viewing distance of about 70 cm, the visual angle subtended by the lines varied horizontally from approximately 5 to 16°.

The lines were divided into four categories labeled A to D; each category included three exemplars (A: lines 2–4; B: lines 5–7; C: lines 8–10; and D: lines 11–13). Each category had a central exemplar that was located at the middle of the category and that was surrounded by two peripheral exemplars. To ensure that all categories were bounded, as was a case in [Experiment 1](#), we included two additional lines (lines 1 and 14) that were peripheral to the

four categories and that served as anchors or peripheral category bounds.

Procedure At the start of each trial, a fixation cross (+) was presented at the center of the screen for 500 ms. The fixation point was followed immediately by a line to be classified. The line was presented for 250 ms after which the screen was blank until the subject responded. The subjects were instructed to respond as quickly as possible without sacrificing accuracy.

To classify the line as a member of one of the four categories, subjects used their right hand to identify the category by pressing one of four buttons on the computer's QWERTY keyboard; the buttons were from the list [m, . /] and corresponded to the categories A to D in order. Subjects used their left hand to identify a line that fell outside of the 4 categories (i.e., lines 1 and 14) by pressing the “z” key.

After the subject had responded, feedback was provided. A correct response was followed by a short high-pitched tone (800 Hz, 50 ms); an incorrect response was followed by a longer low-pitched tone (80 Hz, 200 ms). In both cases, the correct category label was displayed for 500 ms at the center of the screen. For lines 1 and 14, a question mark was presented instead of the category label. There was a 1-s inter-trial interval following the feedback during which the screen was blank.

The experiment was organized in 20 blocks of 49 trials each. Each group of four blocks (i.e., 196 trials) contained all possible combinations of sequence pairs.

Subjects were given a break after each block of trials. To signal the break, subjects were shown mean accuracy for all blocks completed so far. To end the break, subjects pressed a key on the keyboard; the duration of the break was under the subject's control.

The first block of 49 trials was treated as a practice session. Because we are concerned with sequences of two stimuli, we have not included the first trial of each block. After dropping the practice block and the first trial from each block, the remaining blocks did not contain all combinations. Hence, we report data from 20 blocks of 48 trials, each comprising approximately 65 presentations of each line and approximately 5 sequence pairs for each possible combination of stimuli.

Results

Cross-category transitions Analysis of the cross-category effects is complicated by chance considerations. When asked to classify an exemplar from Category B, for example, there are more response possibilities to the right of the correct category than to its left. Hence, chance favors a

response to the right over one to the left. Of course, chance favors a response in the other direction for cases from Category C. To take chance into account, we developed a scoring method that removes the chance component leaving a bias index that reflects the influence of the preceding category on the current one (see the Appendix). The index allows us to evaluate category-to-category biases using a single function; the function summarizes biases from all possible different-category pairs.

As is shown in the Appendix, we first tallied the responses for each subject to each category conditional on the category of the preceding trial. When we tallied responses to each category, we included responses to Stimuli 1 and 14 as separate categories. Hence, the basic data for each category are contained in a matrix of four rows (corresponding to the category presented on the preceding trial) and six columns (corresponding to the 6 possible responses).

Figure 6 (top panel) presents the bias index as a function of the category of the preceding trial. As is shown in the figure, bias increased linearly across the four categories as a function of the category of the preceding trial², $F(1, 18) = 42.29, p < .001$. As we illustrated in the Appendix, an increasing function implies assimilation; that is, the subjects responded to the exemplars as if they were closer to the preceding category than they actually were. The pattern of errors is consistent with the between-category push illustrated in Experiment 1.

Same-category transitions To assess same-category effects, we compared performance on sequence pairs from the same category as a function of the distance between them. In Experiment 1 with same-category pairs, produced exemplars shifted away from the center of the category in the direction of the preceding stimulus. Referring to Fig. 5 again, suppose Stimulus 7 is presented after Stimulus 5 (top panel), Stimulus 6 (middle panel), or Stimulus 7 (bottom panel). If the representation shifts in the direction of the stimulus from the preceding trial, the 7→7 pair (bottom panel) should yield best performance because Stimulus 7 is in the middle of Category B. Accuracy for the 6→7 pair (middle panel) should be lower because Stimulus 7 is on the border of the shifted position of Category B. Finally, the 5→7 pair (top panel) should yield lowest accuracy because Stimulus 7 is outside the shifted position of Category B.

² Taking into consideration that the bias data for individual categories showed the same assimilation trend as the average, we collapsed individual category data to avoid unnecessary repetitions. In this analysis, the linear trend was significant for all four categories ranging from $F(1, 18) = 15.3, p < .001$ for Category B to $F(1, 18) = 51.5, p < .0001$ for Category D. While quadratic and cubic trends were significant in some cases, a linear trend was a dominant trend in all four categories.

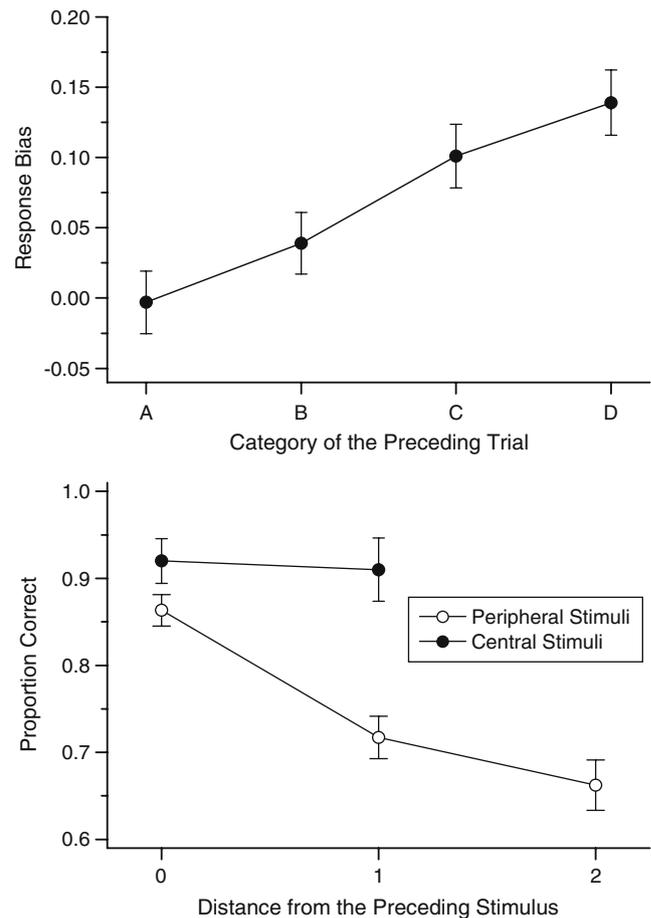


Fig. 6 Summary of Experiment 2. The top panel shows the mean bias score, as a function of the category on the preceding trial. The bottom panel shows accuracy for same-category trials as a function of the distance between successive stimuli. In both panels, error bars show the standard error of the mean

To assess the predictions for each category, we sorted successive lines into groups that varied in their distance from the preceding stimulus: a zero-step distance (i.e., a repetition), and a one-step distance, and a two-step distance. The three step sizes cannot occur equally often. For the peripheral stimuli within each category, all three sizes are possible, but for the central stimulus within each category, only zero- and one-step distances are possible. For Category A, for example, the zero-step pairs included 2→2, 3→3, and 4→4 pairs, the one-step pairs included 3→4, 3→2, 2→3, and 4→3, and the two-step pairs included 2→4 and 4→2 pairs.

Figure 6 (bottom panel) shows the proportion correct as a function of distance for successive same-category comparisons. As is shown in Fig. 6, accuracy for peripheral stimuli decreased linearly as the distance increased, $F(1, 18) = 232.76, p < .001$. For the central stimulus within each category, however, accuracy was at ceiling and was not affected by the preceding stimulus.

When subjects erred in judging pairs from the same category at the periphery of the category, they responded by classifying the stimulus as if it were from the adjacent category just over the boundary (other errors occurred in only 2% of cases). Hence, the increase in errors with step size is consistent with the same-category pull found in the production task.

Comparison to Stewart et al. (2002) Recall that Stewart et al. (2002) reported contrast in a two-category classification task. To consider whether or not the present data are comparable to their data, we selected the subset of our experiment that corresponded most closely to the conditions in their experiment. Specifically, we compared high-contrast pairs of stimuli from the same category against high-contrast pairs of stimuli from the adjacent category. We used peripheral lines (i.e., Stimuli 4, 5, 7, 8, 10, and 11) that were preceded by lines either from the same category (i.e., sequence pairs 2→4, 5→7, 7→5, 8→10, 10→8, 13→11), or from a different category (i.e., sequence pairs 6→4, 9→7, 3→5, 12→10, 6→8, 9→11).

Different-category pairs were classified more accurately (73.26%) than same-category pairs (62.80%), $t(18) = 4.07$, $p < 0.01$, the same relation reported by Stewart et al. (2002). They explained the difference in terms of a relative judgement strategy, but the relation is also consistent with the shift-in-representation that we have documented in both experiments.

Sequence effects across the experiment Recall that the bias index we used to analyse cross-category transitions in classification indicates the strength of assimilation. To evaluate the strength of cross-category assimilation across the experiment, we calculated the bias index for five groups: blocks 1–4, 5–8, 9–12, 13–16, and 17–20. Figure 7 (top panel) displays the bias indices as a function of block of trials: the cross-category assimilation decreased with practice, but still occurred late in the training; the linear trend for the last four blocks of trials was still significant, $F(1, 18) = 13.66$, $p < 0.05$.

To evaluate the effect of practice on the same-category transitions, we evaluated the same-category trial-to-trial transitions for blocks 1–4, 5–8, 9–12, 13–16, and 17–20. Figure 7 (bottom panel) shows the slopes of the same-category step-size functions (i.e., the function we used to analyze the same-category transitions in Experiment 2); as shown in the figure, performance was flat, the linear contrast was not significant, $F(1, 18) = 1.12$, $p > .05$.

Discussion

In a classification task, we have observed a systematic pattern of trial-to-trial errors predicted from the trial-to-trial

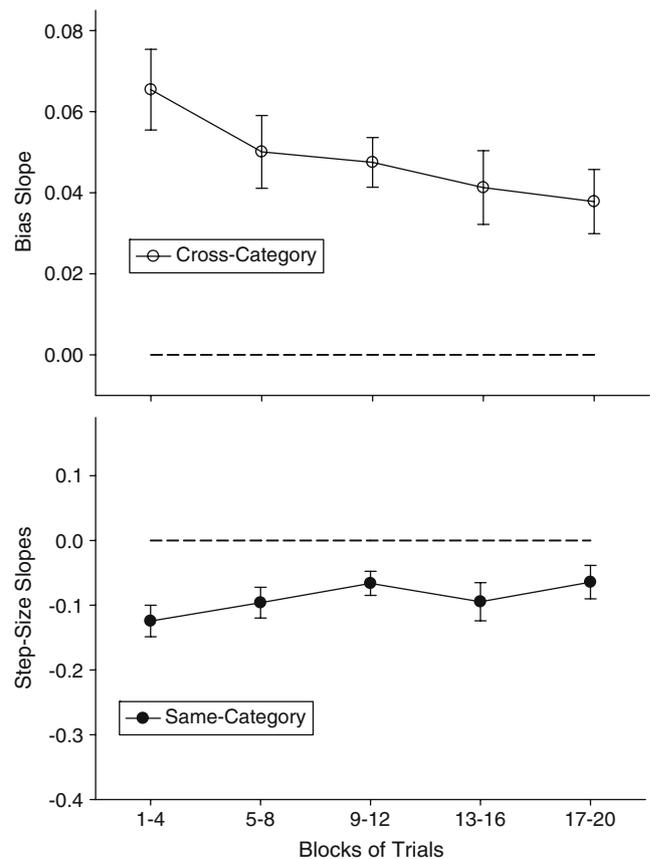


Fig. 7 Effect of practice on the cross- and the same-category trial-to-trial effect in Experiment 2. The top panel shows the mean bias score, as a function of the blocks of trials. The bottom panel shows the step-size slopes for same-category trials as a function of the blocks of trials. In both panels, the dotted horizontal line shows no sequence bias value

shifts of representation documented using a production task (Experiment 1). As anticipated, the between-category transitions illustrated assimilation consistent with the push observed in the production task. The same-category transitions illustrated contrast consistent with the trial-to-trial adjustments observed in the production task. For the high-contrast cases that corresponded most closely to the conditions explored by Stewart et al. (2002), we replicated their findings. We conclude that the classification data are consistent with an account based on shifts in representation.

Although the systematic trial-to-trial errors in the classification experiment are consistent with the representation shifts documented in Experiment 1, the data do not, by themselves, force the conclusion. The conclusion is based on a correlation between production and classification data: A test is needed that can connect the pattern of trial-to-trial errors in classification more directly to the trial-to-trial shifts in representation measured by production data. As Bussemeyer and Myung (1988) noted, “A more powerful test ... can be realized by combining both the

categorization paradigm and the prototype production paradigm into a single study”.

Experiment 3: measuring representation shifts after classification

Experiment 1, a production task, provided estimates of trial-to-trial changes in the state the subject’s internal category representation. **Experiment 2** showed that errors in classification are consistent with the shifting-representation idea, but, because the contingency is based on a correlation across experiments, the data do not force the conclusion that the errors in categorization reflect a shift in representation. To bring the shifts in representation under experimental control (i.e., to force the conclusion that the pattern of errors reflects a shift in representation), the third experiment combined the techniques in a series of three-trial tests. On each test, subjects classified two stimuli (as in the second experiment) and then produced a stimulus (as in the first experiment). The triplet-sequence design allows observing both classification-production and classification-classification sequence pairs in the same session. Importantly, both the stimuli to be classified and the exemplars produced by the subject were drawn on the same screen. If the pattern of errors reflects shifts in representation induced by classification, we should be able to measure the shift using the production data.

Method

Subjects Fifteen undergraduate students, from Queen’s University participated in the experiment. They received course credit for participation, and all reported normal or corrected-to-normal vision. All were right-handed, and the experiment took about an hour to complete.

Apparatus and stimuli The stimuli were presented on a 17-inch (c.43.2-cm) SVGA monitor using a graphics system with resolution of $1,280 \times 1,024$ pixels. The stimuli were 11 lines presented in black on a white background. The center of the screen was labeled pixel 0, and line lengths were calculated with the center of each line at pixel 0. The shortest line was 250 pixels and the longest was 500 pixels. The intermediate lines were constructed at 25-pixel intervals. At a viewing distance of about 60 cm, the visual angle subtended by the lines varied horizontally from approximately 9 to 18° .

The lines were divided into three categories labeled A, B, and C; each category included three exemplars (A: lines 2–4; B: lines 5–7; and C: lines 8–10). Each category had a central exemplar located at the middle of the category surrounded by two peripheral exemplars. To ensure that all categories were bounded, we included two additional lines

(lines 1 and 11) that were peripheral to the four categories and that served as anchors or peripheral category bounds. The category ranges were set at 262–337, 368–412, and 413–487 pixels for categories A, B, and C, respectively.

Procedure The experiment combined both classification and production trials organized in a series of triplets. Each triplet was composed of a two-classification trials followed by a production trial.

On each classification trial, 1 of the 11 stimulus lines appeared in the middle of the screen centered horizontally and vertically. To classify a line, subjects used a mouse to click on one of five buttons on the bottom of the computer’s screen. The buttons were labeled “-”, “A”, “B”, “C”, and “+” and corresponded to the left peripheral line (Stimulus 1), Categories A, B, C, and the right peripheral line (Stimulus 11), respectively. After the stimulus had been classified, it was replaced on the screen by feedback. Feedback for a correct response was the message “CORRECT” (in green) at the top of the screen and centered horizontally. Feedback for an incorrect response was the message “INCORRECT; the line was from category X” (where X corresponds to correct category) at the top-center corner in red. If a subject incorrectly classified Stimulus 1, feedback was “INCORRECT; the line was outside of range (TOO SHORT)”. If a subject incorrectly classified Stimulus 11, feedback was “INCORRECT; the line was outside of range (TOO LONG)”. The feedback messages were displayed for 500 ms, and the screen was blank for a 1-s inter-trial interval.

On each production trial, a category label, “A”, “B”, or “C” was presented. The label was centered horizontally 200 pixels above the vertical center of the screen. To produce an example of the category indicated, subjects used a mouse to enlarge a 3×3 -pixel black dot that appeared in the middle of the screen. To enlarge the dot, subjects clicked on one edge of the dot and dragged it to the right. Dragging the edge to the right extended the dot in both directions creating a horizontal line; the line was 3 pixels high.

When satisfied that the line was a satisfactory example of the category, the subject was prompted to click on a button labeled “Confirm” on the bottom of the computer’s screen. After they had confirmed that they were finished producing the response, feedback was provided.

For a correct response, feedback was “CORRECT” at the top-center corner of the screen (in a green font). For an incorrect response, it was “INCORRECT; you produced a line from X category” (where X is the name of the category produced) at the top-center corner in a red font. If the produced line was shorter than 262 pixels, feedback was “INCORRECT; you produced a line outside of range (TOO SHORT)” and if the produced line was longer than 487 pixels, it was “INCORRECT; you produced a line outside

of range (TOO LONG)”. The feedback was displayed for 500 ms. There was a 1-s inter-trial interval during which the screen was blank. A production was counted as correct when the line fell within the boundaries of the correct category.

The experiment was organized in 24 blocks of 45 trials each. Each block was organized in 15 triplets, each composed of a classification–classification–production sequence of trials. The order in which the stimuli were administered was randomized independently for each subject.

All possible combinations of 11 stimuli in the first position followed by 11 stimuli in the second position and 3 categories in the third position yield 363 combinations. With 24 blocks, the 363 combinations would require 1,089 trials. To keep the number of trials per block at 45, 9 trials (3 of the 11 × 11 × 3 combinations) were excluded. The cases excluded were selected at random for each subject.

Subjects were given a break after each block of trials. To signal the break, subjects were shown a graph with mean accuracy for classification and time (in s) for all trials completed so far. To end the break, subjects used the mouse to click on a “continue” button on the screen; the break’s duration was under the subject’s control.

Before the experiment started, subjects completed two sets of practice trials in which they learned to classify the stimuli and to use the mouse to draw lines. For the classification trials, the subjects were shown all 11 lines in random order. For the production practice, subjects were asked to draw a member of each category three times (in random order).

Results

To confirm that classifying a stimulus induces shifts in representation, we analyzed the lines produced as a function of the category presented on the previous trial.

Cross-category transitions For production trials, we calculated mean production size as a function of the preceding category (similar to Experiment 1). Table 2 shows the means of productions of each category. Overall, there was a negative relation between mean production and the distance to the category presented on the previous trial, the mean production decreased linearly, $F(1, 14) = 9.87, p < .01$. The pattern was equivalent to that observed in Experiment 1.

For classification trials, we calculated the same bias index that we used in Experiment 2. Figure 8 presents the bias index as a function of the category of the preceding trial. As is shown in the figure, bias increased linearly across the four categories as a function of the category of the preceding trial, $F(1, 14) = 68.92, p < .0001$. Recall that an increased bias implies assimilation; that is, the subjects

Table 2 Mean production for each category in pixels as a function of the preceding category in Experiment 3

		Present Category			Slope
		A	B	C	
Previous Category	-	321.2	389.8	487.4	399.5
	A	316.7	392.9	485.6	398.4
	B	312.6	392.1	487.0	397.2
	C	313.7	388.6	476.3	392.9
	+	313.9	389.7	468.9	390.8
					-2.28

responded to the exemplars as if they were closer to the preceding category than they actually were. The pattern of errors is identical to one we observed in Experiment 2.

Same-category transitions To consider the influence of classification on the representation for the same-category case, we expressed the deviation from the center of the category for the exemplar produced as a function of the position of the preceding exemplar within the category. There are three positions within each category, left, central, and right. Figure 9 shows the mean deviation from the center of the category as a function of the location of the exemplar within the category on a preceding classification trial; the data are plotted separately for each category.

As is shown in Fig. 9, the lines produced following classification were biased by the location within the category of the item classified; the deviations from the center of the category increased linearly as a function of the position of the item on the classification trial, $F(1, 14) = 60.38, p < .001$. As in Experiment 1, the same-category data document contrast: the representation on which the production was based shifted toward the stimulus that had been classified on the preceding trial.

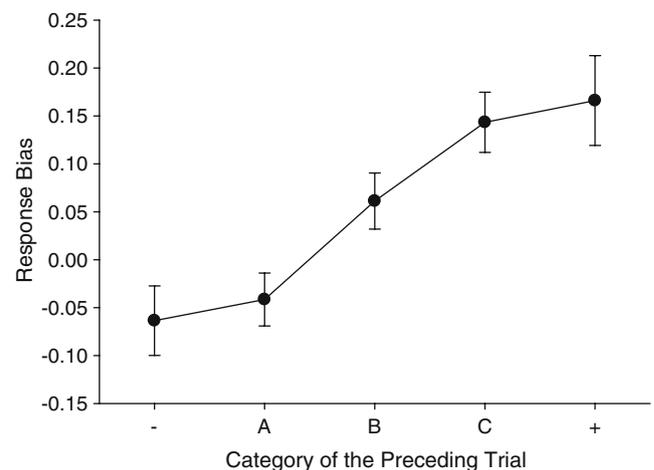


Fig. 8 Classification trials in Experiment 3. The mean bias score as a function of the category on the preceding trial

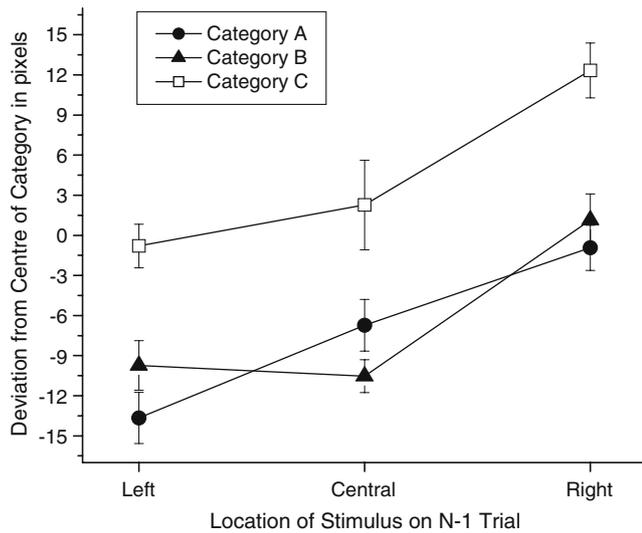


Fig. 9 The deviation (*in pixels*) from the center of the category produced on the current trial as a function of the category distance from the item classified on the preceding trial in Experiment 3

For classification trials, we compared performance on sequence pairs as a function of the distance between them in the same way we compared sequence pairs in Experiment 2. We sorted successive lines into groups that varied in their distance from the preceding stimulus: a zero-step distance, a one-step distance, and a two-step distance. As in Experiment 2, accuracy was at a ceiling for the central stimulus within each category and was not affected by the preceding stimulus. Thus, we report results for peripheral stimuli only. Figure 10 shows the proportion correct as a function of distance for successive same-category comparisons. As is shown in Fig. 10, accuracy for peripheral stimuli decreased linearly as the distance increased, $F(1, 14) = 8.61$, $p < .05$, analogous to the results of Experiment 2.

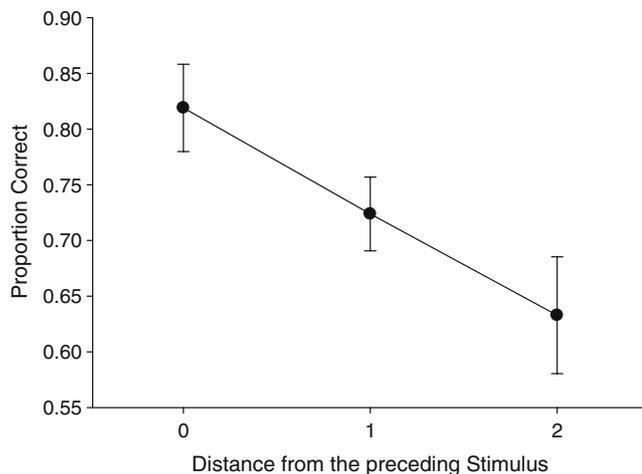


Fig. 10 Classification trials in Experiment 3. Accuracy for same-category trials as a function of the distance between successive stimuli

Sequence effects across the experiment Figure 11 presents the slopes of the mean production function for the both cross-category and same-category transitions for blocks 1–6, 7–12, 13–18, and 19–24. For the cross-category transitions, the slopes of the function decreased linearly, confirming the results we observed in Experiment 1: the cross-category push in representation has been reduced with practice; the linear trend for the last 12 blocks was not significant. For the same-category transitions, however, the same-category pull was not affected by practice.

Considering that the accuracy of subjects’ responses increased with practice and the classification–classification pairs of sequences constituted only third of all possible combinations in our triplets design (the two other pairs are classification–production and production–classification), we did not have sufficient data to perform and report any inferential statistics of the effect of practice across the experiment. Nevertheless, the patterns observed in practice effect (Fig. 12) was consistent with those observed in Experiment 2: the between-category assimilation decreases with practice, but still occurs late in training, and the same-category contrast stays strong through the experiment.

Effect of feedback

Our production data can help us to understand the role of feedback after correct and error responses. Specifically, do errors change the direction and magnitude of sequential effects? The effect of errors from the preceding trials has been studied extensively in the probabilistic categorization literature (Dorfman & Biderman, 1971; Kac, 1962; Kubovy & Healy, 1977; Thomas, 1973). While the settings of the

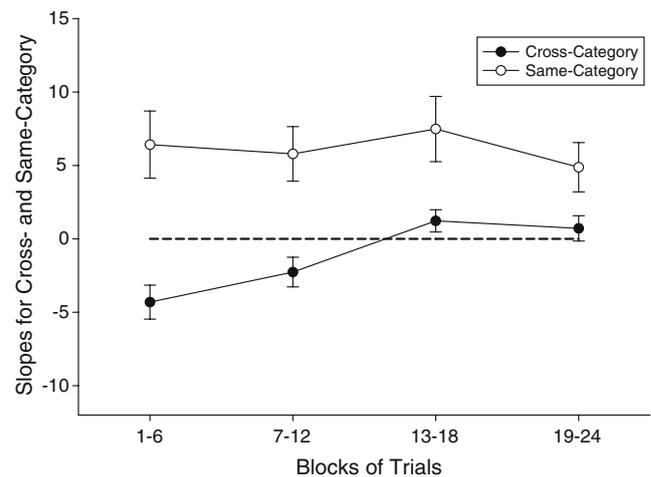


Fig. 11 Effect of practice for production trials in Experiment 3. The slopes of the mean production function are shown for both same-category and cross-category sequence pairs

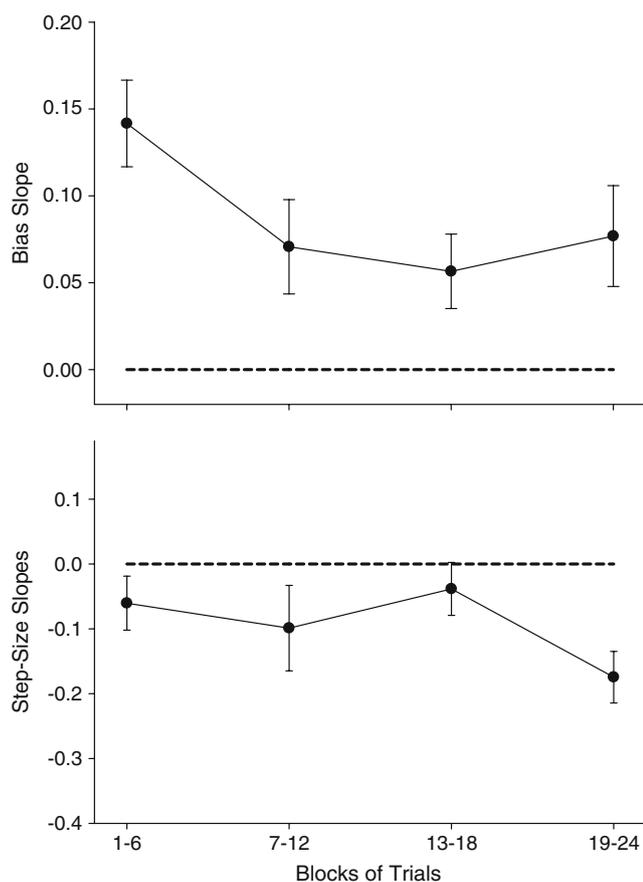


Fig. 12 Effect of practice for classification trials in Experiment 3. The slopes of the bias function are shown for cross-category (*top panel*) and same-category (*bottom panel*) sequence pairs

probabilistic classification task differ from ours in a number of important ways (e.g., the only two categories used bounded on one side, membership of exemplars is not deterministic), they provide important notion of response criteria that shifts systematically from trial to trial. For example, a number of accounts (commonly defined as the dynamic-cutoff models; see Dorfman & Biderman, 1971; Kac, 1962; Larkin, 1971) specify the direction of the criteria shift as a function of preceding trial. These papers provide solid empirical evidence that the error responses induce a shift of criteria away from the response category but that the correct responses induce smaller, non-systematic shift of response criteria. In the absolute identification task, Stewart, Brown, and Chater (2005 p. 57), investigated the carryover effect of errors and found that, when the preceding response was in error, the overall magnitude of assimilation increased slightly, but the effect was not significant.

Our framework of shifting representation has no mechanism to specify how representation shift would change after error responses. However, taking into account the findings from probabilistic classification and absolute

identification tasks, we might speculate that, for cross-category transitions, the magnitude of cross-category push might be larger after an error than after a correct response. It is harder to anticipate the direction of the same-category shift following an error response.

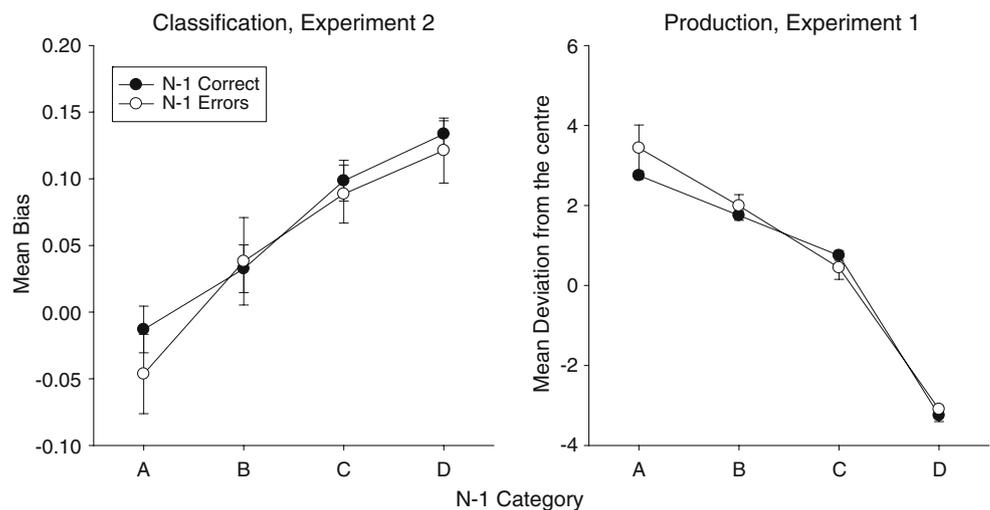
Recall that, in our shifting representation framework with same-category transitions, the production is pulled toward the last item, resulting in a pattern of diminishing accuracy as the distance between $n - 1$ and n items increases. If the first item from the same category is classified incorrectly, an additional corrective shift might be required to settle the category location. For example, after correct responses, the best performance for the same-category pairs was observed when the same peripheral items were repeated and the worst performance was observed when the items were drawn from the opposite ends of the same category. These patterns in performance were duplicated by corresponding trends in production data: the sizes of the produced lines were pulled toward the last item. We expect that errors will result in less pronounced same-category pull and reduced advantage of the repeated items. Due to a small number of errors that the subjects made in Experiment 3, we analyzed after-error data in production of Experiment 1 and classification of Experiment 2.

Cross-category analysis For production data, we found that the cross-category push increased (Experiment 1, right panel of Fig. 13) with a corresponding increase in assimilation for classification data (Experiment 2, left panel of Fig. 13). To evaluate if the increase was significantly larger for after-error pairs, we compared slopes of the functions for cross-category push and assimilation using a paired-sample t test; in both cases the gain was not significant, $t(11) = -0.21$, $p = .83$, and, $t(7) = -1.83$, $p = .11$, accordingly.

Same-category transitions To evaluate same-category transitions, we analyzed the accuracy data of Experiment 2 following both correct and error trials. We presented the same-category transitions by showing how the accuracy for peripheral items changes with a step-size between current and preceding item; thus, for step-size 0, it was a repetition of an item, for step-size 1, it was a central item followed a peripheral item, and for step-size 3, it was a peripheral item followed by another peripheral item from the opposite end of the same category. Considering that subjects made few errors to the central items, we report results for step-sizes 0 and 2 only. Thus, we compared performance of step-size 0 against step-size 2 for after-error responses.

Figure 14 shows the differences in step-size performances. Consistent with our predictions for the same-category pairs, after-error performance did not show an

Fig. 13 Cross-category sequence effect following error trials for classification and production tasks



advantage for step-size 0, $t(17) = 1.04$, $p = .31$. Recall that we did find the advantage for step-size 0 when first response consisted of mostly correct replies. The results suggest that errors on preceding trial reduced the magnitude of sequence effect for the same-category pairs.

If the same-category errors in the classification task reflect a shift in representation then we should find a similar pattern in production data. One complication of presenting parallel evidence is that the pure production task does not have a limited set of items that we can partition into step-sizes and present in the same graph with accuracy. Nevertheless, it is possible to evaluate the relations between two sequential productions to determine if an error on the preceding trial affects the production on the following trial. For the same-category transitions in the production data of [Experiment 1](#), we reported a high correlation between

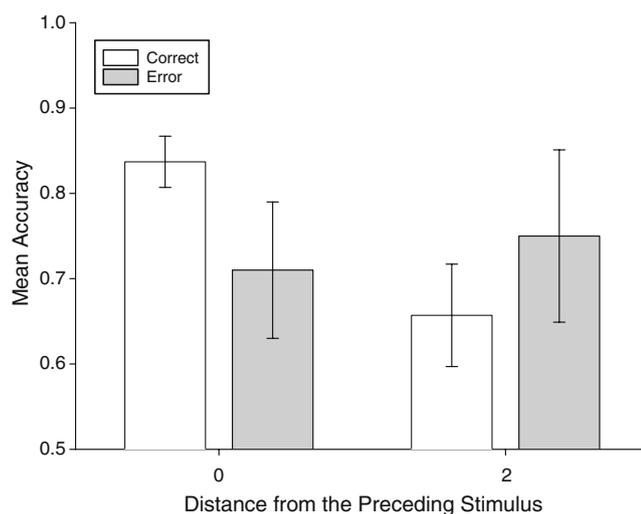


Fig. 14 Accuracy for same-category trials as a function of the distance between successive stimuli following correct and error responses

preceding and current production, which suggested that there was a pull toward the item on the last trial—the trend that was confirmed by our production trials of [Experiment 3](#). The error data suggest that errors interrupt the magnitude of the sequential effect; thus, it should be reflected in reduced correlation between production pairs. To analyze the correlations, we selected the same-category production pairs, in which the size of first production was outside of its category boundary (error response), and we computed correlation analyses separately for Categories 1 to 4. The correlation values were 0.23, 0.20, -0.32 , and 0.28 accordingly (only the last value was significant, $p < .05$).

To summarize, we failed to find increased assimilation for classification data and a stronger push for production data for the cross-category sequence pairs when an error was made on the preceding trial, as reported in probabilistic classification literature. At the same time, our results are consistent with those reported by [Stewart et al. \(2005\)](#). We assume that our multiple-category design is more consistent with absolute identification task than with probabilistic classification task, which employs two overlapping categories with unbounded peripheral space.

For the same-category sequence pairs, the performance data showed reduced contrast and corresponding reduction in the pull for production data when an error was made on the preceding trial. The results suggest that errors might interrupt the process of shifting a category representation toward the last presented item. Another question is why the positive feedback induced systematic shift in our data, but not in probabilistic classification for the same-category sequence pair. Once again, our simplified category structure of three clearly defined stimuli (unlike the unlimited number of stimuli in the probabilistic classification) provides precise anchoring references to the same-category shifts.

General discussion

We started with a conflict between the well-established finding of *assimilation* in identification and recent findings of *contrast* in categorization. To investigate the puzzle, we studied subjects' representation of the categories using a new production method to measure the subject's internal representation.

In [Experiment 1](#), we asked subjects to produce an item corresponding to one of four categories. We inferred the subject's understanding of the categories by measuring the size of the items produced. Production on each trial was strongly influenced by the category of the preceding trial. When both the preceding and current trials were from the same category, the current production was shifted from the center of the category in the direction of the preceding production. When the two trials were from different categories, however, the current production was shifted away from the preceding production. The data illustrate two shifts in representation: an adjustment towards the preceding exemplar when the current and the preceding trial were from the same category, and a push away from the preceding exemplar when the current and preceding trials were from the different categories.

[Experiment 2](#) asked whether errors in categorization are predicted by the trial-to-trial shifts in representation documented in the production task. In [Experiment 1](#), when subjects were shown an item at the left side of a category, the other categories were pushed away from the category and the category itself was adjusted left. For example, when Item 5 (at the left side of Category B) was shown, the representation of Category B was adjusted left, and the representations of Categories A, C, and D were pushed away from Category B. When subjects were then asked to categorize Item 7 (from Category B) immediately after they had classified Item 5, they were likely to misclassify Item 7 as an example from Category C. By contrast, if instead of Item 7, they were asked to classify Item 8 from Category C, they were likely to misclassify it as an example from Category B. The first case illustrates an error reflecting the adjustment of Category B to the left, and the second illustrates an error reflecting a shift of Categories C, and D away from Category B (the between-category push). Thus, the pattern of errors in [Experiment 2](#) paralleled the shifts in representation—the between-category push and the same-category adjustment—found in the production data of [Experiment 1](#).

[Experiment 3](#) tracked trial-to-trial representation shifts following categorization by placing a production trial immediately after a classification trial. When the category to be produced was the same as the category classified on the preceding trial, subjects shifted their production towards the previous item within the category. When the category to

be produced was from a different category, subjects produced an exemplar shifted away from the preceding category. Again, the pattern of productions following classification in [Experiment 3](#) paralleled the shifts in representation—the between-category push and the same-category adjustment—found in [Experiment 1](#).

In an identification task, each stimulus is, in effect, its own category; if we treat identification as classification with one exemplar per category, the between-category data in all three experiments are consistent with the identification literature.

The same-category data in [Experiments 2](#) and [3](#) illustrate the opposite bias. The representation of the current category was pulled away from the center of the category towards the preceding exemplar from the same category; as a result, the subjects misplaced the current stimulus to a category further from the previous category. The bias for same-category sequences is, of course, the same as the bias reported by Stewart et al. (2002).

To summarize, the production results document two shifts of representation: a between-category push and a same-category pull. The classification results show parallel assimilation and contrast. The effect of practice and the pattern of responses on post-error trials suggest that the two shifts in representation are independent from each other and might have different underlying mechanisms, but their nature, however, remains unclear.

Mechanisms underlying sequence effects

Stewart et al. (2002) argued that the same-category contrast in classification is a bi-product of an estimation process they call the Memory and Contrast (MAC) strategy. The idea is that subjects estimate the relative difference between successive items. If the current stimulus is not identical to the previous one and is in the direction of the adjacent category, the subject assesses the distance between the current and preceding stimuli. If the distance is too large, the subject decides for the adjacent category. In Stewart et al. (2002) words, if an item at the left periphery of Category A is followed by an item at the right periphery of the same category, “the large inter-trial difference will lead to an erroneous shift in response from Category A to [the adjacent] Category B. In other words, large within-category shifts will induce errors” (pp. 4–5). In their account, decision is based on the relative magnitude of the successive items; it does not rely on shifts in representation.

Our data point to shifts of representation, but we have no direct evidence to contradict Stewart et al. (2002) relative-judgment strategy in classification. It is quite possible that subjects use a fast-and-frugal strategy in simple cases, such

as the two-category example that Stewart et al. (2002) explored. It is also quite possible that, in early stages of classification, subjects rely on a relative judgment mechanism, but, with practice, rely more and more on to a long-term representation. While the MAC strategy offers a good explanation for the same-category contrast, it is not equipped to reconcile both the same-category contrast and the cross-category assimilation we reported in our experiments.

It is difficult to understand the cross-category push in such terms, however, as there are no immediate benefit to subjects to remember items as if they were located further away than they actually are. There are number of ways to explain why the representation is shifted *away* from the immediately preceding item. Purks et al. (1980) argued that the shift might reflect the process of broadening the response region around the last presented stimulus, pushing other criteria away. Criterion-setting theory (Treisman, 1985, p. 191) provides a similar explanation, suggesting that the tracking strategy that favors repetition of the same response on the next trial would expand the range of response criteria for the current item by shifting its left and right criteria downwards and upwards, whereas the criteria for neighboring items will be pushed away.

Jones et al. (2006) offer an attractive solution for the two opposite sequence effects we observed in our study. Their model of sequence effects in category learning (SECL) assumes that classification decisions are guided by both long-term knowledge of category structure and by immediately preceding stimuli. To evaluate the role of perceptual and decisional recency independently, Jones et al. (2006) used a probabilistic classification task. In the task, subjects classified uni-dimensional stimuli into two categories; unlike a deterministic two-category classification task (e.g., Stewart et al., 2002), the probability of membership of exemplars varied with their distance from the category's boundary. By controlling the location of the stimulus while manipulating its membership, the task allowed them to evaluate the effect of the previous response (decisional recency) separately from the effect of the previous stimulus (perceptual recency). The overall trends of the subject's responses were assimilation to the previous *response* and contrast to the previous *stimulus*.

The notion of perceptual and decisional recency as two independent effects maps directly to our findings of cross-category assimilation versus same-category contrast effects. The same-category adjustment can be viewed as a perceptual calibration process—in the absence of the stimulus' identification information in classification (beyond its category membership), subjects use the location of the item presented on the preceding trial to fine-tune the representation of the category. Since subjects tend to generalize their

responses to the category they classified on a preceding trial, the cross-category effects can be based on the decisional recency. Specifically, when a preceding stimulus belongs to a neighboring category, subjects tend to “assimilate” their current responses toward a previously classified category.

While Jones et al. (2006) explanation for decisional and perceptual recency fits our findings of assimilation and contrast, there is an unresolved issue with the SECL's ability to account for our data. Specifically, the decisional recency—which is essentially a response repetition bias—helps to explain the assimilation toward a category classified on the preceding trial. In our multiple-category design, however, we also observed assimilation of more distant categories. For example, if a Category A item is followed by a Category C item, the latter tends to be classified as a Category B item. In Jones et al.'s data, a high-distance item showed a slight negative recency, which would reverse the cross-category effect from assimilation to contrast. Considering that SECL was tested on a two-category task, it would be unfair to expect the model to account for our multiple-category design. Thus, the SECL's explanation of sequence effects might be limited to a two-category case and need not generalize to our multiple category design.

As we noted earlier, the empirical evidence of sequence effects in classifications has grown in the recent classification literature (Jones et al., 2006; Jones & Sieck, 2003; Stewart & Brown, 2004; Stewart et al., 2002). Nevertheless, none have investigated whether practice has any influence on the magnitude and the direction of sequence effects. While the investigation of sequence effects has been overwhelmingly documented in the absolute identification literature (Lockhead & King, 1983; Luce et al., 1982; Mori, 1998; Treisman, 1985; Ward & Lockhead, 1970), the role of practice has not been given the same rigor. Hartman (1954) tested subject ability to identify nine tones for eight test-week sessions. While Hartman did not report the properties of errors (e.g., away or toward the preceding item), he reported that the magnitude of errors was reduced; specifically, the reduction in error rate was significantly larger for distant tones than for near tones. Recently, Rouder, Morey, Cowan, and Pfaltz (2004) investigated the effect of practice in absolute identification of lines. While increased practice affected performance of subjects overall, it especially improved performance when the previous stimulus was far from the current one.

In our experiments, the magnitude of cross-category assimilation (and corresponding cross-category push) decreased with practice, and in some instances was eliminated completely. For the same-category transitions, the contrast (and corresponding pull in representation) remained strong throughout the entire experiment. The

findings reinforce the differences found in the absolute identification literature, where distant items were more affected than near items. The SECL's distinction of two independent sources for assimilation and contrast provides a good rationale why practice affects the same- and the cross-category transitions differently, but it is still not clear why practice should affect the perceptual and decisional recency differently. Regardless of the underlying cause of changes in magnitude of sequence effects with practice, the evidence imposes new requirements on existing classification accounts.

Future directions

We investigated subjects' performance by analyzing accuracy and response biases. One of the limitations imposed by accuracy data is that, as the subject errors decrease, the subset of useful data is reduced accordingly. A good solution to overcome this problem is to analyze response time (RT). The RT data provide a window into understanding the nature of cognitive representations and decision processes; the benefits of RT data have been discussed extensively (for review, see Luce, 1986). In the context of our work, RT data would allow us to obtain and analyze the same number of sequence pairs for both performance and production.

Our study has not analyzed the properties of sequence effects prior to the immediately preceding ($n - 1$) trial. It is well-established in the absolute identification literature that the effect of stimuli from $n - k$ trials ($k > 1$) reverses its direction from assimilation to contrast (Holland & Lockhead, 1968; Lacouture, 1997; Ward & Lockhead, 1970, 1971). In the absence of any empirical evidence, we can only speculate that, if cross-category assimilation changes to contrast, we might observe strong contrast effects on trials prior to $n - 1$.

Considering that the sequence effects are carried from one trial to another in a fixed time interval, it is reasonable to assume that changes in ISI timing might influence the properties of sequence effects. Recently, Matthews and Stewart (2009) manipulated ISI in a task of identifying 10 tones of varying frequency. By extending the ISI from 500 to 10,000 ms, the magnitude of assimilation from a preceding trial diminished as a function of increased ISI, while the contrast from the $n - k$ trials was enhanced. A potentially fruitful area of research for the future is to investigate whether a similar trend is observed in classification task. A production task can be used to evaluate how ISI manipulation affects the corresponding magnitude of shifts in representation of categories.

Our interactive production technique extends existing methods by accessing subjects' representation both directly

and with high resolution. Inasmuch as the pattern of errors in classification corresponded closely to changes in the subject's production, the interactive production method has proven to provide reliable measurements of the subject's category representation.

Finally, Lockhead (2004) has recently reminded us of the theoretical implications of shifts in internal representation of the sort documented here. He argues that classical psychophysics has erred by applying a Gaussian model of neural noise to judgment while ignoring the trial-to-trial variability introduced by assimilation and contrast. The present data show that contrast and assimilation are ubiquitous, but whether all noise can be attributed to sequence effects remains an open question.

Acknowledgements The research was supported by a grant to D.J. K.M. from the Natural Sciences and Engineering Research Council of Canada (NSERC) and by an AEG grant from Sun Microsystems of Canada. V.Z. was supported by a post-graduate research scholarship from OGS. Anthony A. J. Marley provided extensive commentary and suggestions. We would like to thank Elizabeth Johns, Don Franklin, and Randall Jamieson for their criticisms and comments on an earlier draft of this article.

Appendix

The bias-response index was calculated to evaluate the magnitude and direction of potential sequence effects. First, we tallied the number of responses to each category conditional on the category of the preceding trial. When we tallied responses to each category, we included responses 1 and 14 as separate categories. Hence, the basic data for each subject responding to stimuli from each category were contained in a matrix of four rows (corresponding to the category presented on the preceding trial) and six columns (corresponding to the 6 possible responses). To calculate bias associated with each preceding category (i.e., for each row of the matrix), we weighted the frequency by its ordinal position in the matrix (1 to 6), summed the weighted frequency values, and divided the sum by the total number of responses.

To illustrate the how the bias index works, we will examine two hypothetical cases: an example illustrating assimilation and an example illustrating contrast.

Assimilation Table A1 shows a frequency matrix of responses partitioned by the category of the preceding trial (each row), the category of the current trial (each block of four rows), and the responses 1, A, B, C, D, and 14 (each column). The rightmost columns show the

calculated bias index, and the bias index adjusted for the stimulus category.

Table A1 An example of assimilation: frequency of responses to each stimulus as a function of the category of the preceding trial and the associated bias indices

Category		Responses						Bias	Bias Adjusted
<i>n</i>	<i>n</i> - 1	1	A	B	C	D	14		
A	A	2	33	7	0	0	0	2.12	0.12
	B	1	32	9	0	0	0	2.19	0.19
	C	0	29	12	1	0	0	2.33	0.33
	D	0	27	11	4	0	0	2.45	0.45
B	A	1	10	25	6	0	0	2.86	-0.14
	B	0	5	31	6	0	0	3.02	0.02
	C	0	2	29	10	1	0	3.24	0.24
	D	0	1	22	18	1	0	3.45	0.45
C	A	0	2	13	24	4	0	3.70	-0.30
	B	0	0	9	30	2	0	3.83	-0.17
	C	0	0	7	27	8	0	4.02	0.02
	D	0	0	2	22	18	0	4.38	0.38
D	A	0	0	7	10	24	1	4.45	-0.55
	B	0	0	4	6	30	2	4.71	-0.29
	C	0	0	2	10	26	4	4.76	-0.24
	D	0	0	2	18	22	9	4.75	-0.25
Weights		1	2	3	4	5	6		

The bias index is calculated as,

$$BIAS_{N,N-1} = \frac{\sum_{i=1}^6 f_i \times w_i}{\sum_{i=1}^6 f_i}$$

For the first row of the matrix, bias equals,

$$BIAS_{A,A} = \frac{2 \times 1 + 33 \times 2 + 7 \times 3 + 0 \times 4 + 0 \times 5 + 0 \times 6}{2 + 33 + 7 + 0 + 0 + 0} = 2.12$$

To adjust the bias for each stimulus category, the weight associated with the category of the correct response was subtracted from the bias value:

$$BIAS_{AdjustedN,N-1} = Bias_{N,N-1} - weight_N$$

The bias score reflects the direction of the error responses. If the errors were distributed symmetrically around the correct answer, the adjusted bias score would be zero.

In the top panel of Table A1, there are more errors to the right of the correct response than to the left. The reason for the shift is complex: Part of it indicates assimilation, but part may reflect chance: When the stimulus is from Category A, there are more possible responses (B, C, D)

to the right of the correct answer (A) than to the left of it. However, chance is independent of the preceding category and is symmetrical across all the stimuli in the experiment. That is, whereas chance favors a response to the right of the correct response for stimuli from Category A, it favors a response to the left for stimuli from Category D.

To obtain a measure of bias uncontaminated by chance, we averaged the adjusted bias across stimuli from each category. Using the example in Table A1, to obtain a bias measure for trials preceded by Category A, we averaged values of 0.12, -0.14, -0.30, -0.55, obtaining the value of -0.22. The corresponding values for trials preceded by Categories B to D are -0.06, 0.09, and 0.26, respectively. The resulting function relating bias with the category of the preceding trial increases monotonically: (A, -0.22), (B, -0.06), (C, 0.09), (D, 0.26). Because chance has been taken into account by averaging across all stimulus categories, the increasing function indicates assimilation.

Contrast Table A2 summarizes the calculations for an example showing contrast. The function relating bias to the category of the preceding trial takes values (A, 0.18), (B, 0.04), (C, -0.09), and (D, -0.34), a monotonically decreasing function illustrating contrast.

Table A2 An example of contrast: frequency of responses to each stimulus as a function of the category of the preceding trial and the associated bias indices

Category		Responses						Bias	Bias Adjusted
<i>n</i>	<i>n</i> - 1	1	A	B	C	D	14		
A	A	3	33	4	0	0	0	2.03	0.02
	B	9	32	1	0	0	0	1.81	-0.19
	C	12	29	1	1	0	0	1.79	-0.21
	D	11	27	1	0	0	0	1.74	-0.26
B	A	0	6	25	10	1	0	3.14	0.14
	B	0	5	31	6	0	0	3.02	0.02
	C	1	10	29	2	0	0	2.76	-0.24
	D	1	18	22	1	0	0	2.55	-0.45
C	A	0	0	4	24	13	2	4.30	0.30
	B	0	0	2	30	9	0	4.17	0.17
	C	0	0	7	27	8	0	4.02	0.02
	D	0	0	18	22	2	0	3.62	-0.38
D	A	0	0	0	1	24	10	5.26	0.26
	B	0	0	0	3	30	9	5.14	0.14
	C	0	0	2	4	26	10	5.05	0.05
	D	0	0	2	18	22	9	4.75	-0.25
Weights		1	2	3	4	5	6		

References

- Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *The Journal of the Acoustical Society of America*, *51*, 483–502.
- Busemeyer, J. R., & Myung, I. J. (1988). A new method for investigating prototype learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 3–11.
- DeCarlo, L. T., & Cross, D. V. (1990). Sequential effects in magnitude scaling: Models and theory. *Journal of Experimental Psychology: General*, *119*, 375–396.
- Dorfman, D. D., & Biderman, M. (1971). A learning model for a continuum of sensory states. *Journal of Mathematical Psychology*, *8*, 264–284.
- Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, *46*, 373–380.
- Hartman, E. B. (1954). The influence of practice and pitch-distance between tones on the absolute identification of pitch. *The American Journal of Psychology*, *67*, 1–14.
- Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics*, *3*, 409–414.
- Hu, G. (1997). Why is it difficult to learn absolute judgment tasks? *Perceptual and Motor Skills*, *84*, 323–335.
- Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: Separating decisional and perceptual sequential effects in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 316–332. doi:10.1037/0278-7393.32.3.316
- Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 626–640. doi:10.1037/0278-7393.29.4.626
- Kac, M. (1962). *A note on learning signal detection*. IRE Transactions on Information Theory, IT-8, 126–128.
- Kubovy, M., & Healy, A. F. (1977). The decision rule in probabilistic categorization: What it is and how it is learned. *Journal of Experimental Psychology: General*, *106*, 427–446.
- Lacouture, Y. (1997). Bow, range, and sequential effects in absolute identification: A response-time analysis. *Psychological Research*, *60*, 121–133.
- Larkin, W. (1971). Response mechanisms in detection experiments. *Journal of Experimental Psychology*, *91*, 140–153.
- Lockhead, G. R. (1984). Sequential predictors of choice in psychophysical tasks. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 27–47). Hillsdale, NJ: Erlbaum.
- Lockhead, G. R. (2004). Absolute judgments are relative: A reinterpretation of some psychophysical ideas. *Review of General Psychology*, *8*, 265–272. doi:10.1037/1089-2680.8.4.265
- 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*, 461–473.
- Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.
- Luce, R. D., Nosofsky, R. M., Green, D. M., & Smith, A. F. (1982). The bow and sequential effects in absolute identification. *Perception & Psychophysics*, *32*, 397–408.
- Matthews, W. M., & Stewart, N. (2009). The effect of interstimulus interval on sequential effects in absolute identification. *The Quarterly Journal of Experimental Psychology*, *62*, 2014–2029. doi:10.1080/17470210802649285
- Mori, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology*, *52*, 72–83.
- Petrov, A., & Anderson, J. R. (2005). The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, *112*, 383–416. doi:10.1037/0033-295X.112.2.383
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*, 353–363.
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.
- Purks, S. R., Callahan, D. J., Braida, L. D., & Durlach, N. I. (1980). Intensity perception. X. Effect of preceding stimulus on identification performance. *The Journal of the Acoustical Society of America*, *67*, 634–637.
- Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*, 328–350.
- Rouder, J. N., Morey, R. D., Cowan, N., & Pfaltz, M. (2004). Learning in a unidimensional absolute identification task. *Psychonomic Bulletin & Review*, *11*, 938–944.
- 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. doi:10.1037/0278-7393.30.2.416
- Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 3–11. doi:10.1037/0278-7393.28.1.3
- Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*, 881–911. doi:10.1037/0033-295X.112.4.881
- Thomas, E. A. C. (1973). On a class of additive learning models: Error-correcting and probability matching. *Journal of Mathematical Psychology*, *10*, 241–264.
- Treisman, M. (1985). The magical number seven and some other features of category scaling: Properties for a model of absolute judgment. *Journal of Mathematical Psychology*, *29*, 175–230.
- Treisman, M., & Williams, T. C. (1984). A theory of criterion setting with an application to sequential dependencies. *Psychological Review*, *91*, 68–111.
- Ward, L. M., & Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, *84*, 27–34.
- Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics*, *9*, 73–78.
- Zangwill, O. L. (1937). An investigation of the relation between the processes of reproducing and recognizing simple figures, with special reference to Koffka's trace theory. *British Journal of Psychology*, *27*, 250–276.