******* *in press in Acta Psychologica* ******

CAN CATEGORICAL KNOWLEDGE BE USED IN VISUAL SEARCH?

Sébastien Hélie

Purdue University

Benjamin O. Turner

Nanyang Technological University

Denis Cousineau

University of Ottawa

Running head: Categorical knowledge

For correspondence,

Sébastien Hélie
Department of Psychological Sciences
Purdue University
703 Third Street
West Lafayette, IN 47907

Phone: (765) 496-2692
E-mail: shelie@purdue.edu

Abstract

Smith et al. (2005) have proposed a new categorization paradigm called the visual-search categorization task to study how display size affects categorization performance. Their results show that, in a wide range of conditions, category knowledge collapses as soon as multiple stimuli are simultaneously displayed in a scene. This result is surprising and important considering that humans parse and categorize objects from complex scenes on a daily basis. However, Smith et al. only studied one kind of category structure. This article presents the results of three experiments exploring the effect of display size on perceptual categorization as a function of category structure. We show that rule-based and information-integration categories are differently affected by display size in the visual search categorization task. For rule-based structures, target-present and target-absent trials are not much affected by display size. However, the effect of display size is bigger for information-integration category structures, and much more pronounced for target-absent trials than for target-present trials. A follow-up experiment shows that target redundancy (i.e., having more than one target in the display) does not improve performance with information-integration category structures. These results suggest that categories may be learned differently depending on their underlying structure, and that the resulting category representation may influence performance in the visual search categorization task.

Keywords: categorization, visual search, generalization, transfer.

Introduction

Learning about categories is critically important in everyday life. Categorization reduces the complexity of the environment by allowing for interaction with a smaller (but still extensive) set of categories instead of treating each new object as unique. In particular, categorizing a new object allows for a series of inferences about the object's properties and possible uses. For example, categorizing a new object as a 'chair' suggests that it is probably sufficiently robust to carry one's weight and that it is possible to sit on it. Given the importance of categorizing objects in large visual scenes, it should be possible to perform a visual search for a given category. Yet, this intuitive hypothesis was challenged by Smith, Redford, Gent & Washburn (2005). Herein, we revisit the question and examine visual search for two types of category structures.

Visual search and categorization

The above example points to at least two different components that are rarely a focus of categorization research. First, the object ('chair' in this example) needs to be segregated from the remainder of the scene in order to be identified as something that can be categorized (Hélie, 2017). Second, the categorization judgment is typically a means to an end, and is important to the extent that it provides information about additional properties of the object (Hélie & Ashby, 2012).

In a clever series of experiments, Smith and his colleagues (2005) tested the effect of display size on categorization judgment using the visual-search and categorization (VSC) task. In the VSC task, participants are first trained to categorize individual stimuli as members of one of a number of contrasting categories. After the categories have been learned, the participants transfer to a visual-search task where a number of stimuli are simultaneously presented and the participant's task is to find members of a target category in the display. This kind of transfer task is very important because the VSC task allows for studying how participants can parse a relatively complex scene in order to achieve a categorization decision. The results in Smith et al. (2005) were compelling: Categorization accuracy *collapsed* to (almost) chance performance as soon as more than one stimulus was presented simultaneously. Further exploration with the VSC paradigm suggests that difficulties in categorization persisted with a reduced number of categories, additional training with distractors, changes in the visual aspect of the stimuli, and changes in stimulus overlap in the display (Smith et al., 2005). The only factor that reduced the effect of display size in the VSC task was an increased similarity between members of the same category.

Results in the VSC task are counterintuitive considering that humans regularly parse scenes in everyday life to identify and categorize objects. One possibility is that the category representations learned during the category training portion of the VSC task were not sufficient to support visual search (Ell, Smith, Peralta, & Hélie, 2017; Hélie, Ell, & Shamloo, 2017). Categorization and visual search are two different tasks, and Hélie and colleagues have shown that the structures of the learned categories affect the generality and transferability of the category representations. For example, Hélie and Ashby (2012) trained participants using either rule-based (RB) or information-integration (II) category structures and then had participants transfer to a "same"-"different" categorization (SDC) task. In the SDC task, the participants see two stimuli simultaneously and are asked whether the two stimuli belong to the same category or not. The results show that participants can transfer their category knowledge to the SDC task only when the category structures are rule-based (e.g., verbalizable). When the category structures require the integration of information from more than one stimulus dimension at a pre-decisional stage (e.g., not verbalizable), there was little transfer from the categorization task to the SDC task. Hélie and Ashby argued that learning different category structures leads to different kinds of category representations, and that different category representations allow for different generalizability and transfer performance.

The nature of the categorical representations was further explored in Hélie and Cousineau (2015). In a series of experiments, Hélie and Cousineau tested the effects of backward masking and integration masking on RB and II categorization performance. The results show that RB categorization is more robust than II categorization to short delays between the stimulus and mask in backward masking, and to more opaque masks in integration masking. Hélie and Cousineau suggested that both backward masking and integration masking reduce the signal-to-noise ratio of the stimulus. RB category representations may be feature-based and more digital, which would explain their increased robustness to noise, whereas II category representations may be more holistic and analog, which would explain their increased sensitivity to noise (Hélie & Cousineau, 2015). These results are consistent with Hélie and Ashby (2012) in that the difference between two category members may be considered “decision noise”, and RB categories would be more robust than II categories to these differences in the SDC task.

Given Hélie and Ashby’s (2012) results, one might expect that the category structures may have a critical effect on performance in the VSC task. For example, Smith et al. (2005) used the well-known dot-distortion paradigm of category learning. This task was a natural choice given the amount of research that has been done with these stimuli. However, it is unclear whether these stimuli produce representations similar to RB or II category structures. Given the difficulty in verbalizing a rule to categorize dot-distortions, it is possible that the category structures in Smith et al. required the integration of several stimulus dimensions at a pre-decisional stage. If this is the case, then the category representations would be less general (e.g., more task-specific) and the results would be consistent with Hélie and Ashby's explanation of the generality of category representations. Moreover, this would suggest that category knowledge may not collapse (i.e., revert back to near-chance performance) in the VSC task if RB category structures are used instead. This is because the category representations learned with RB categories seem to be more general and transferable to new tasks than the category representations learned with II categories (Hélie & Ashby, 2012). One goal of this article is to directly manipulate category structures (Experiment 2) in order to explore how it affects performance in the VSC task.

Another possible explanation for Smith et al.'s (2005) results is that the strategy used by participants in the visual search task may interact with the category structures. One possible strategy for searching for a member of category "A" in a complex display is to sequentially look at each stimulus and (mentally) ask the question "Is this an A?". Maddox and colleagues (2004) used this task (hereafter referred as "YES/NO") and participants were unable to learn II category structures when queried about category membership. Ell et al. (2017) showed learning of II categorization with participants performing the YES/NO task, but the amount of learning was modest. In contrast, Hélie et al. (2017) found unimpaired II category learning with the YES/NO task. However, while there are some inconsistencies with regards to YES/NO performance with II category learning, all three studies found that participants were unimpaired in learning RB category structures in the

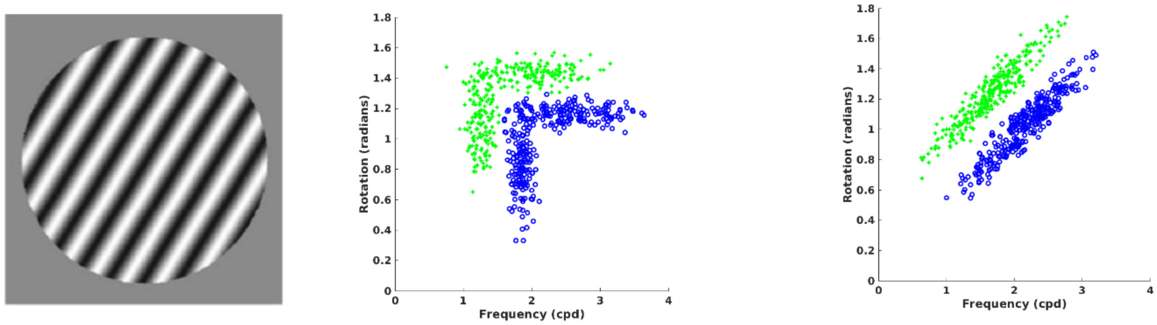
YES/NO task.

While categorization instructions may interact with category structures in early learning, it is unclear whether participants can shift from one set of instructions to another after the categories have been learned. If the interaction persists, then it is possible that categorization knowledge collapsed in the VSC task because the strategy adopted in the visual-search task (YES/NO) was at least partly incompatible with the category structures. This possibility needs to be addressed before we can explore the interaction between category structures and performance in the VSC task (Experiment 1).

Finally, if participants are using a strategy in which they are categorizing each stimulus one by one, then adding redundancy, e.g. by having more than one target present in the display, could reduce the collapse in the VSC task. Experiment 3 tests for this possibility by replicating Experiment 2 with the exception that more than one target can be present in the VSC task.

Overview of the experiments

The main goal of this research was to explore how performance in the VSC task interacts with category structures. This was done by performing three different experiments. All experiments used sine-wave gratings of constant contrast and size. Figure 1a shows an example stimulus, Figure 1b shows RB category structures, and Figure 1c shows II category structures. Note that these are the same stimuli and category structures used in Hélie & Ashby (2012).



(a) (b) (c)

Figure 1. Stimuli and category structures used in the experiments. (a) Example stimulus. (b) Rule-based categories. (c) Information-integration categories. Each stimulus is a sine-wave grating. The horizontal axis in each panel represents the frequency (inversely proportional to bar width) and the vertical axis in each panel represents the bar orientation (counterclockwise rotation from horizontal).

Experiment 1 tested whether participants could transfer knowledge when shifting from a regular category training condition (i.e., press “A” if the stimulus is an “A” and “B” if the stimulus is a “B”; henceforth referred to as A/B training) to YES/NO categorization (“Is this an ‘A’?”). Successful transfer is critical because changing from category training to visual search in the VSC task may involve a similar change in decisional process. Earlier studies have shown that participants can be impaired when learning II category structures (e.g., Figure 1c) with the YES/NO task (Ell et al., 2017; Maddox et al., 2004). Yet, it is unclear whether participants can perform the YES/NO task with II categories after having been trained with an A/B paradigm. To test for this possibility, Experiment 1 trained participants in A/B categorization before transferring to a YES/NO task that used the same category structures. Accuracy in the last training block of classification was compared with accuracy in the YES/NO transfer block. If accuracy in the transfer block is similar to accuracy in the last block of training, then the possible change in strategy in the VSC task is not responsible for the collapse in performance. However, if the transfer block shows a cost in accuracy, then it is possible that the collapse observed in the VSC task by Smith et al. (2005) was caused by a lack of proper training. To anticipate, Experiment 1 shows perfect transfer accuracy from A/B training to the YES/NO categorization paradigms with both RB and II category structures. Hence, this possible change in decisional process cannot account for Smith et al.'s results in the VSC task.

Having established that participants can transfer from an A/B to a YES/NO paradigm with both category structures, Experiment 2 directly addressed the main goal of the research: Is performance in the VSC task affected by category structures? To explore this possibility, two groups of participants were trained using either RB or II category structures and then were transferred to the visual search task. The effect of display on accuracy was separately calculated for target-present and target-absent trials. Obtaining a similar collapse in VSC accuracy in both conditions would suggest a general effect of the VSC task, whereas differential changes in performance for each condition would suggest that category structures modulate performance in the VSC task. The results with II categories reproduced Smith et al.'s (2005) results as categorization performance declined with display size, and this was caused by low accuracy in target-absent trials. However, in line with Hélie & Ashby (2012), RB categories were mostly spared and performance did not collapse with display size. Experiment 3 tested the effect of target redundancy by reproducing Experiment 2 but this time allowing for more than one target to be present on target present trials. As in Experiment 2, the effect of display on accuracy was separately calculated for target-present and target-absent trials. If accuracy in the VSC task is higher in Experiment 3 then in Experiment 2, then it is possible that the redundancy present in the natural world prevents collapse of categorical knowledge in visual search outside the laboratory. The results in Experiment 3 were very similar to those obtained in Experiment 2, in that performance collapsed for II stimuli but not for RB stimuli. Furthermore, the collapse with II stimuli was again seen only on target-absent trials. Hence, target redundancy did not have much of an effect in the VSC task.

Experiment 1

To successfully achieve the VSC task, each stimulus in the display needs to be individually categorized. Assuming the search is unrestricted, one possible process to achieve this goal is to consider each stimulus individually and ask the question 'Is this a *target*?' where *target* stands for the category that is currently being searched for. In Experiment 1, we explored whether using this strategy in the visual search portion of the VSC task could affect the ability of participants to successfully perform the task using RB and II category structures (as in Figure 1). Participants in past studies have sometimes showed impaired learning of II categories when asked categorization questions such as 'Is this an A?' or 'Is this a B?' (i.e., YES/NO training) (Ell et al., 2017; Maddox et al., 2004). This result could account for some of the difficulties that participants experienced in the VSC task when no verbal rule is available to describe the categories (as in Smith et al., 2005). However, it is unclear whether participants are only impaired in nonverbal learning in the YES/NO paradigm or if they are also impaired in using nonverbal knowledge once it has been learned. Experiment 1 addressed this question by training participants using a standard A/B training protocol and then transferring to a YES/NO condition. Establishing that participants can use their nonverbal knowledge to answer categorical questions is essential in interpreting the results of the VSC task.

Method

Participants

Forty-three undergraduate students at the University of California Santa Barbara were recruited to participate in Experiment 1. Twenty-one participants were trained using the RB category structures from Figure 1b, and the remaining 22 participants were trained using the II category structures from Figure 1c. Sample size was determined by using G*Power 3.1.9.2 (Faul, Erdfelder, Lang, & Buchner, 2007). With an effect size of 0.25 and $\alpha = 0.05$, 38 participants are sufficient to achieve a power of 0.85. Each participant was given credits for participation as partial completion of a course requirement.

Apparatus and stimuli

The stimuli were sine-wave gratings of constant contrast and size presented on a 21-inch LCD monitor (1280 × 1024 resolution). Each stimulus was defined by a pair (x_1, x_2) sampled from an arbitrary 100 × 100 stimulus space and converted to a disk using the following equations: $\textit{frequency} = x_1 / 30 + 0.25$ cpd, and $\textit{orientation} = 9x_2 / 10 + 20$ degrees. This yielded stimuli that varied in orientation from 20 to 110 degrees (counterclockwise from horizontal) and in frequency between 0.25 and 3.58 cpd. The stimuli were generated with Matlab using Brainard's (1997) Psychophysics Toolbox and occupied an approximate visual angle of 5 degrees. They were shown on a grey background (RGB of 128, 128, 128).

For the RB condition (Figure 1b), category "A" stimuli were generated using two multivariate normal distributions with the following parameters (Ashby & Gott, 1988): $\mu_{a1} = \{30, 50\}$; $\Sigma_{a1} = \{10, 0; 0, 150\}$ and $\mu_{a2} = \{50, 70\}$; $\Sigma_{a2} = \{150, 0; 0, 10\}$. A similar sampling method was used to generate category "B" stimuli: $\mu_{b1} = \{50, 30\}$; $\mu_{b2} = \{70,$

50}; $\Sigma_{b1} = \Sigma_{a1}$; and $\Sigma_{b2} = \Sigma_{a2}$. For the II condition (Figure 1c), category "A" stimuli were generated using a multivariate normal distribution with the following parameters: μ_a = {40, 50}; Σ_a = {10, 0; 0, 280}. The same sampling method was used to generate category "B" stimuli: μ_b = {60, 50}; $\Sigma_b = \Sigma_a$. The resulting stimuli were then rotated 45° counterclockwise around the center point of stimulus space. Note that perfect accuracy was possible in both conditions.

Stimulus presentation, feedback, response recording, and response time (RT) measurement were acquired and controlled using Matlab. The stimuli in both phases were centered both vertically and horizontally and occupied about five degrees of visual angle. The participants responded by using the 'd' and the 'k' keys on a standard keyboard (identified with blank stickers). The key labels were displayed on the screen. During the training phase (A/B training; Blocks 1 to 5), the 'A' label appeared at the bottom-left of the screen (associated with the 'd' key) and the 'B' label appeared at the bottom-right of the screen (associated with the 'k' key). At test (YES/NO; Block 6), the 'YES' label appeared at the bottom-left of the screen (associated with the 'd' key) and the 'NO' label appeared at the bottom-right of the screen (associated with the 'k' key). In addition, the question "Is this an 'A'?" was displayed in the middle top of the screen on half of the test trials. During the remaining test trials, the question "Is this a 'B'?" was displayed instead. No question was displayed during the training phase.

Importantly, the category structures were the same in both the training and test phases, and the participants were told to use the knowledge acquired during the training phase in the test phase. Correct responses were followed by the word 'Correct' in green font in the middle of the screen and incorrect responses were followed by the word

‘Incorrect’ in red font in the middle of the screen. If a response was too late (more than 5 seconds), participants saw the words “Too slow!” in black font. If a participant hit a wrong key, s/he saw the words “Wrong key!” in black font. Half of the stimuli in the training phase were ‘A’s and the remainder were ‘B’s. Similarly, half the stimuli were ‘A’s and the remainder were ‘B’s during the test phase, and category membership was counterbalanced with the category inclusion questions.

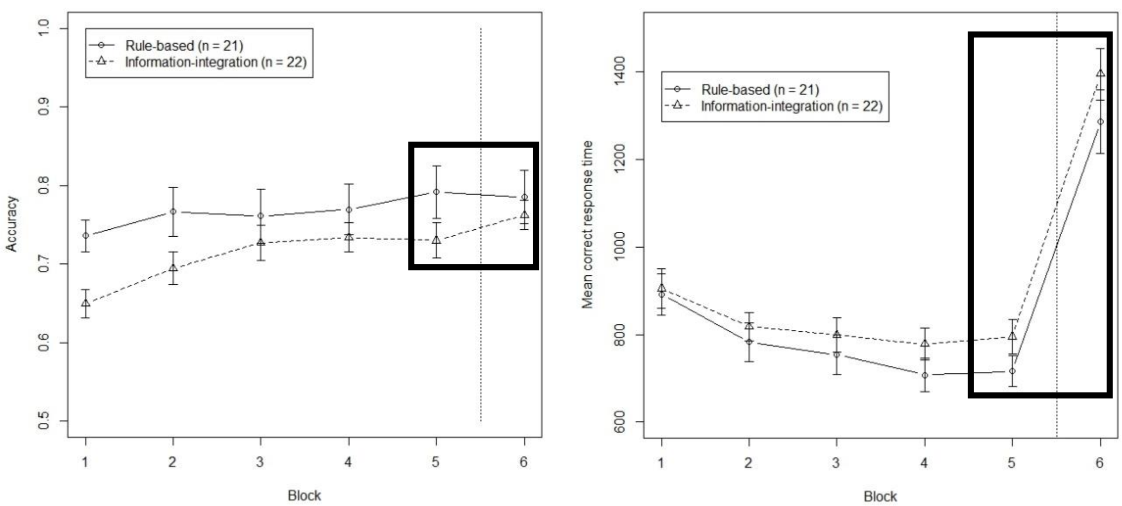
Procedure

Each experimental session was composed of 6 blocks of 100 trials (for a total of 600 trials). During the training phase (Blocks 1 to 5), participants partook in a regular perceptual categorization task (A/B training). Their task was to assign each stimulus to the ‘A’ or ‘B’ category by pressing the left or right buttons (respectively) as labeled on the screen. During the test phase (Block 6), the participants were shown a stimulus along with a categorical inclusion question (e.g., “Is this an ‘A’?”) (YES/NO paradigm). Participants responded ‘yes’ or ‘no’ by pressing the left or right button (respectively) as labeled on the screen. In both the training and test phases, a trial proceeded as follows: a fixation point (crosshair) appeared on the screen for 1,500 ms and was followed by the stimulus and response button labels (with a simultaneous category inclusion question during the test phase), which remained on the screen until the participant made a response. After a response was made, the stimulus, response button labels (and question) disappeared and correct or incorrect feedback was given for 750 ms. The participants were allowed to take a break between blocks if they wished.

Results

<u>Accuracy</u>

The mean accuracy per block for each condition is shown in Figure 2a. As can be seen, participants in both conditions improved with practice. This was confirmed by a 2 (RB vs. II) × 5 (Training Block) ANOVA. The effect of Block was significant ($F(4, 164) = 7.09$, $p < .001$), with mean accuracies increasing from 69.2% (Block 1) to 76.0% (Block 5). The effect of Condition ($F(1, 41) = 3.30$, *n.s.*) and its interaction with Block failed to reach statistical significance ($F(4, 164) = 1.29$, *n.s.*).



(a) (b)

<u>Figure 2</u>. Results of Experiment 1. (a) Mean accuracy as a function of Block for both groups. (b) Mean correct response time. In both panels, the vertical dashed line separates the training phase from the test phase. The rectangle highlights the relevant comparison to test for transfer effects. Error bars represent one standard error of the mean.

A more important question is whether there is a difference in performance between the last training block and the test block (the black rectangle in Figure 2a). A 2 (RB vs. II) × 2 (Block 5 vs. Block 6) ANOVA was performed on accuracies. As suggested by the

Figure, there was no effect of Block, Condition, or interaction (all *F*s < 1.92, *n.s.*). Hence, there was no evidence of a transfer cost when changing from an A/B categorization task to a YES/NO task after category training had already occurred.

<u>Response times</u>

The mean correct RTs are shown in Figure 2b. As can be seen, participants became faster during the training phase in both conditions. However, response times were much slower during the test phase (in both conditions). These observations were confirmed by a 2 (RB vs. II) × 5 (Training Block) ANOVA. The effect of Block was again significant ($F(4, 164) = 12.63$, $p < .001$), with correct response times decreasing from 891 ms (Block 1) to 758 ms (Block 5). The effect of Condition ($F(1,41) = 1.03$, *n.s.*) and its interaction with Block failed to reach statistical significance ($F(4, 164) < 1$, *n.s.*).

Similar to accuracies, we also performed a 2 (RB vs. II) × 2 (Block 5 vs. Block 6) ANOVA on mean correct RTs (the black rectangle in Figure 2b). Not surprisingly, there was a large effect of Block ($F(1, 41) = 296.60$, $p < .001$). Participants were much slower in the test phase (a difference of 599 ms), which could be caused by reading the question at the top of the screen. However, this slowdown was similar in both conditions, as suggested by the absence of statistically significant effect of Condition ($F(1, 41) = 1.89$, *n.s.*) and its interaction with Block ($F(1, 41) < 1$, *n.s.*).

Discussion

The goal of Experiment 1 was to test whether participants could transfer from an A/B category learning task to a YES/NO categorization task with both RB and II category structures. Addressing this question is critical to interpreting results from the VSC task, because this switch from A/B to YES/NO may mirror a switch in decisional process between the categorization and visual-search phases in the VSC task. The results from Experiment 1 show that participants can make the switch from A/B to YES/NO instructions without any interference on accuracy with both RB and II category structures. This result is important considering that learning II categories with the YES/NO task has produced unreliable results in the past (Ell et al., 2017; Hélie et al., 2017; Maddox et al., 2004). However, the present experiment suggests that using knowledge acquired from II categorization in a YES/NO tasks does not pose problem. Hence, it is unlikely that the collapse of category knowledge observed in Smith et al. (2005) was caused by such a change. Having ruled out this possibility, we can now proceed to exploring whether category structures affect performance in the VSC task.

Experiment 2

Experiment 1 showed that participants can do a YES/NO task after having learned the categories using A/B training with both RB and II category structures. We are now in a position to test whether performance in the VSC task is affected by the category structures. In Experiment 2, participants performed the VSC task using stimuli drawn from RB or II category structures. Because Hélie and Ashby (2012) showed that participants can transfer RB, but not II, category knowledge to the SDC task, we hypothesized that II category knowledge should collapse more markedly in the VSC task than RB category knowledge.

Method

Participants

Fifty-eight undergraduate students at Purdue University were recruited to participate in Experiment 2. Thirty participants were trained using the RB category structures from Figure 1b, and the remaining 28 participants were trained using the II category structures from Figure 1c. Sample size was determined by using G*Power 3.1.9.2 (Faul et al., 2007). With an effect size of 0.25 and $\alpha = 0.05$, 54 participants are sufficient to achieve a power of 0.95. Each participant was given credits for participation as partial completion of a course requirement. None of the participants had participated in Experiment 1.

Apparatus and stimuli

The stimuli and category structures were the same as in Experiment 1. The sine-wave gratings were presented on a 21-inch LCD monitor (1920 × 1080 resolution) and occupied 5 degrees of visual angle (as in Experiment 1). The response keys and the response labels displayed at the bottom of the screen were also the same as in Experiment 1. In the category learning phase, ‘A’ was displayed in the bottom-left of the screen and ‘B’ was displayed at the bottom-right (corresponding to the ‘d’ and ‘k’ keys, respectively). In each trial, one stimulus was displayed in the center of the screen (as in Experiment 1).

In the visual search phase, ‘YES’ was displayed in the bottom left and ‘NO’ was displayed in the bottom right (corresponding to the ‘d’ and ‘k’ keys, respectively, as in the test phase of Experiment 1). The screen was partitioned horizontally into 4 non-overlapping regions of equal sizes. There was no marking on the screen to indicate the partitions. Between 1 and 4 stimuli appeared simultaneously on the screen, with at most one stimulus per region (so that each region contained either one stimulus or no stimulus). The stimuli in the visual search task were centered vertically in each region. On each visual search trial, a question was displayed in the top-center screen asking ‘Is there an A?’ or ‘Is there a B?’. An example trial from the visual phase is shown in Figure 3.

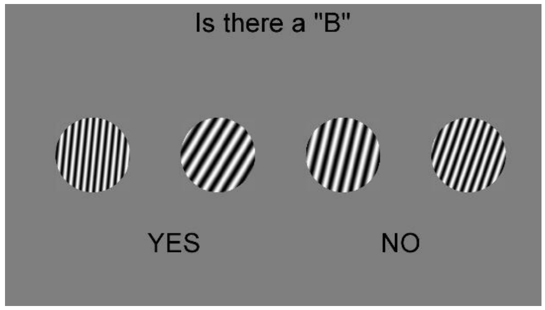


Figure 3. An example trial in the VSC task with a display complexity of four.

Procedure

The experiment was composed of two sessions scheduled during the same work week. In Session 1, all participants were trained for six blocks of 100 trials using an A/B paradigm identical to the training phase of Experiment 1 (using either RB or II category structures). In Session 2, participants were trained with the same category structures as in Session 1. Session 2 went as follows: First, participants were trained for one block of 100 trials in A/B categorization. This was a refresher of Session 1. Next, participants were trained for 6 blocks of 80 trials in the visual search task.

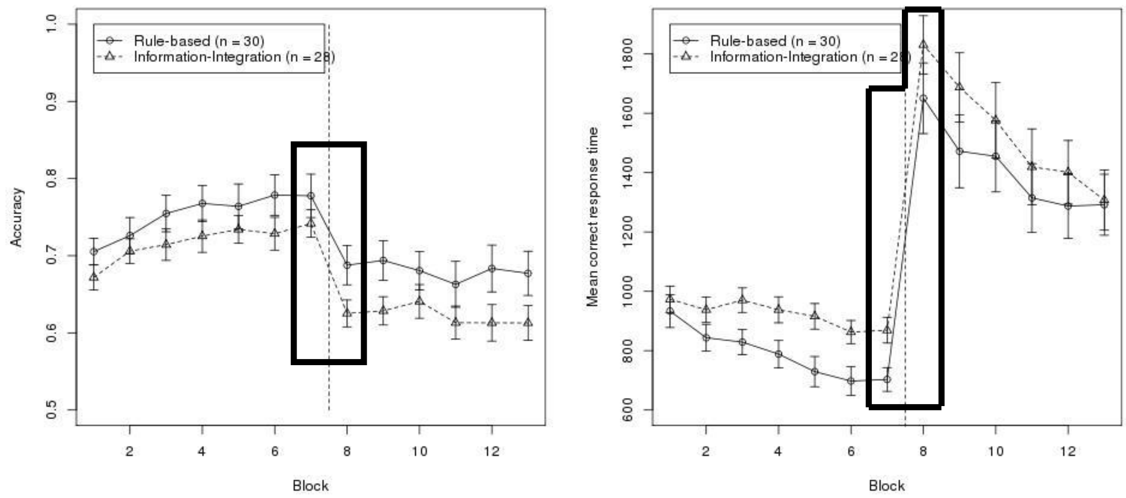
The procedure in the A/B category learning phase was the same as in the training phase of Experiment 1. In the visual search phase, participants saw between one and four stimuli on the screen. The participants were then asked to press ‘YES’ if they could find a target on the screen (as defined by the question on the screen) and ‘NO’ otherwise. Participants then received the same feedback as in the categorization experiment.

Participants were told that the categories were the same as in the categorization phase and that they should therefore use their category knowledge. The target was a member of the ‘A’ category in half of the trials and of the ‘B’ category in the other half. When participants were looking for an ‘A’, a single ‘A’ stimulus was present in half of these trials (a “target-present” trial), with the other half containing no ‘A’ stimulus (a “target-absent” trial). The same applied to trials where participants were looking for a ‘B’, yielding four types of trials. Each type of trials had the same number of trials of each display size (1 to 4). In all trials, distractors were randomly sampled from the non-target category. As in Experiment 1, participants could take a break between blocks.

Results

Accuracy

The mean accuracy per block for each condition is shown in Figure 4a. Blocks 1 to 7 are from the categorization phase while Blocks 8 to 13 are in the visual search phase. As can be seen, participants' accuracy in both conditions improved with practice in the categorization phase, but not much in the visual search phase. We performed separate ANOVAs for each phase. For the categorization phase, a 2 (RB vs. II) × 7 (Block) ANOVA confirmed the previous observations. The effect of Block was statistically significant ($F(6, 336) = 7.40$, $p < .001$), with mean accuracies increasing from 68.9% (Block 1) to 76.0% (Block 7). Similar to Experiment 1, the effect of Condition ($F(1, 56) = 1.88$, *n.s.*) and its interaction with Block failed to reach statistical significance ($F(6, 336) = 0.27$, *n.s.*). A similar 2 (RB vs. II) × 6 (Block) ANOVA was computed for the visual search task. None of the effects was statistically significant: Condition ($F(1, 56) = 3.38$, $p = .071$); Block ($F(5, 280) = 1.42$, *n.s.*); Condition × Block ($F(5, 280) = 0.54$, *n.s.*). The mean accuracy in visual search was 65.3%.



(a) (b)

Figure 4. Results of Experiment 2. (a) Mean accuracy as a function of Block for both groups. (b) Mean correct response time. In both panels, the vertical dashed line separates the categorization phase from the visual search phase. The rectangle highlights the relevant comparison to test for transfer. Error bars represent one standard error of the mean.

Similar to Experiment 1, we also tested transfer performance from the last block of categorization to the first block of visual search (the black rectangle in Figure 4a). A 2 (RB vs. II) × 2 (Block 7 vs. Block 8) ANOVA was performed on accuracies. As suggested by the Figure, accuracies decreased in both condition when transferring from the categorization phase to the visual search phase ($F(1, 56) = 96.76$, $p < .001$). The mean accuracy in Block 7 was 76.0%, which decreased to 65.8% in Block 8. The effect of Condition ($F(1, 56) = 2.55$, *n.s.*) and its interaction with Block ($F(1, 56) = 1.61$, *n.s.*) both failed to reach statistical significance.

Response times

The mean correct RTs are shown in Figure 4b. As can be seen, participants became faster with training in both phases and in both conditions. Response times in the RB condition appeared to be slightly faster than in the II condition in the categorization phase. Similar to accuracies, we computed separate ANOVAs for the categorization and visual search phases. For the categorization task, a 2 (RB vs. II) × 7 (Block) ANOVA confirmed the previous observations. The effect of Block was statistically significant ($F(6, 336) = 10.92$, $p < .001$), with mean correct RTs decreasing from 953 ms (Block 1) to 783 ms (Block 7). However, unlike accuracies, the effect of Condition also reached statistical significance ($F(1, 56) = 6.49$, $p < .05$). The mean correct RT for RB trials was 789 ms, whereas II trials were slower with a mean correct RT of 924 ms. Finally, the Condition × Block interaction failed to reach statistical significance ($F(6, 336) = 1.73$, *n.s.*).

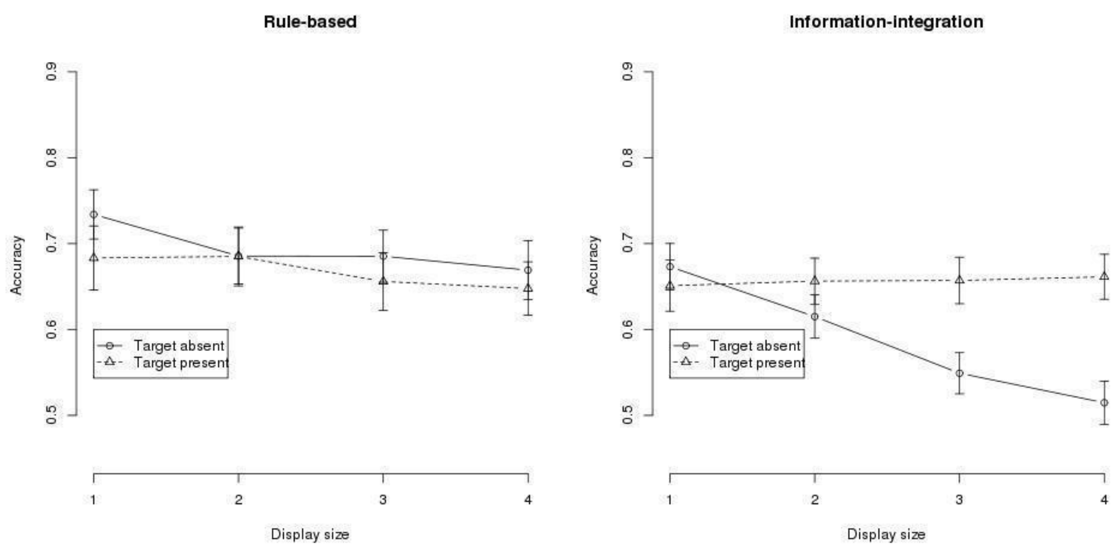
A similar 2 (RB vs. II) × 6 (Block) ANOVA was computed for the visual search phase. Mean correct RTs decreased with practice in both condition ($F(5, 280) = 18.9$, $p < .001$). Mean correct RTs decreased from 1,738 ms (Block 8) to 1,299 ms (Block 13). However, unlike what was seen in the categorization phase, the effect of Condition ($F(1, 56) = 0.74$, *n.s.*) failed to reach significance. The interaction with Block ($F(5, 280) = 0.79$, *n.s.*) was also not significant. As hinted by the standard error bars, variability in RT was larger in the visual search phase than in the categorization phase. This was likely caused by the fact that multiple locations must be examined in the latter.

Finally, we also tested transfer performance from the last block of categorization to the first block of visual search (the black rectangle in Figure 4b). A 2 (RB vs. II) × 2 (Block 7 vs. Block 8) ANOVA was performed on correct RTs. As suggested by the Figure, RTs increased in both condition when transferring from the categorization phase to the visual search phase ($F(1, 56) = 148.16$, $p < .001$). However, the effect of Condition ($F(1, 56) =$

2.95, *n.s.*), and its interaction with Block ($F(1, 56) = 0.01$, *n.s.*), both failed to reach statistical significance.

Display size effect in the VSC task

Smith et al. (2005) found an important effect of display size in the visual search phase of the VSC task. One of the goals of the current experiment was to test whether the collapse of category knowledge would interact with category structures. Figure 5 shows visual-search accuracies averaged across all blocks for target-present and target-absent trials as a function of display size for each category structure. As can be seen, accuracies for RB trials (both target-present and target-absent) and for target-present II trials seemed to be relatively stable across display size. However, performance in *target-absent* II trials substantially decreased with display size.



(a) (b)

Figure 5. The effect of display size on (a) rule-based and (b) information-integration accuracies in Experiment 2. Error bars represent one standard error of the mean.

To verify these observations, we performed a separate 2 (target-present vs. target-absent) × 4 (Display Size) ANOVA for each category structure. For the RB category structures, the effect of Display Size reached statistical significance ($F(3, 87) = 7.59$, $p < .001$). However, the effect of target presence/absence ($F(1, 29) = 0.63$, *n.s.*) and the Target × Display Size interaction ($F(3, 87) = 1.02$, *n.s.*) both failed to reach statistical significance, showing no evidence of differential difficulty for target-present and target-absent trials.

The slope (estimated from a linear regression) is -1.7% per additional item (±0.7%, where the sign ± is used to denote the standard error of the estimate). It is not different from the slopes calculated separately for target-absent (-2.0% per item) and target-present (-1.4% per item) trials, an unusual finding in visual search studies that will be explained when the results of Experiment 3 are examined. This shallow slope finding is smaller than the results from Smith et al. (2005) (although they did not report results separately for target-present and target-absent trials). It suggests that accuracies with RB stimuli decrease slightly, but do not collapse, when display size increases.

For the II category structures, there was also an effect of Display Size ($F(3, 81) = 14.06$, $p < .001$). However, unlike in the RB condition, the effect of target presence or absence ($F(1, 27) = 4.28$, $p < .05$), as well as the interaction between the factors ($F(3, 81) = 20.62$, $p < .001$), were also statistically significant. We proceeded to decompose the interaction by computing the effect of Display Size in each level of Target. For target-present trials, the effect of Display Size failed to reach statistical significance ($F(3, 81) = 0.19$, *n.s.*). The slope was +0.1% per additional item ± 0.5%. However, the effect of Display Size reached statistical significance for target-absent trials ($F(3, 81) = 25.30$, $p < .001$). The slope was -5.0% per additional item ± 0.8%. Hence, for II trials, categorization accuracy collapsed with increased display size and returned to chance performance *only when the* *target was absent*. This is different from RB trials, where the effect of display size was quite small (less than 2% / item) but affected both target-present and target-absent trials similarly.

Discussion

The goal of Experiment 2 was to test whether performance in the VSC task was affected by the structures of the categories. Participants were trained with either RB or II category structures and then transferred to a visual search phase. The results show that display size affects accuracies with both RB and II category structures. However, the most notable effect of display size is for target-absent trials with II category structures, where accuracy decreased by about 5% for every additional item. In contrast, the presence or absence of a target does not influence the effect of display size with RB structures. The decrease in target-absent trials is 3 times less pronounced in RB relative to II, hardly a collapse in performance. This selective effect of display size on target absent II trials was not anticipated or predicted by earlier results.

These findings suggest that only performance with II stimuli when the target is absent collapse with larger display sizes. Smith et al. (2005) did not analyze target-present and target-absent trials separately. Thus, the collapse in categorization performance they observed may have been caused by a collapse *mainly* in the target-absent trials under the assumption that the complex polygons stimuli that they used were processed similarly to II stimuli (i.e., requiring pre-attentional feature integration). If each stimulus is judged independently, the probability of producing a false alarm (making an error in target-absent trials) increases with display size. In contrast, the probability of missing a target (making an error in target-present trial) should not be affected by display size, at least if the search is exhaustive (which we can reasonably assume it to be at these display sizes). As a result, performance with the II stimuli in the VSC task is not surprising. What is more unexpected is the small effect of display size observed for absent trials with RB stimuli. This result—and the fact that the same effect is found for target-present trials—is consistent with a model of premature search termination, where each additional item only adds a fraction of the increased likelihood of false alarm (Cousineau & Shiffrin, 2004). For RB participants, this may be a small trade-off strategy that results in barely any errors and so may have been judged acceptable by the participants. In II, the collapse may have been perceived by the participants themselves who therefore kept an exhaustive termination rule as their strategy.

Experiment 3

The goal of Experiment 3 was to test whether target redundancy could be beneficial to II categorization in complex environments. Hélie and Ashby (2012) showed that the representation built with II learning is less generalizable and more difficult to use in new contexts. Adding target redundancy in the display, which is more ecologically valid in the highly redundant visual world we live in, might help reduce the collapse in performance in the VSC task.

Method

Participants

Forty-six undergraduate students from Purdue University were recruited to participate in Experiment 3. Twenty-three participants were trained using the RB category structures from Figure 1b, and the remaining 23 participants were trained using the II category structures from Figure 1c. Sample size was determined by using G*Power 3.1.9.2 (Faul et al., 2007). With an effect size of 0.25 and $\alpha = 0.05$, 46 participants are sufficient to achieve a power of 0.9. Each participant was given credits for participation as partial completion of a course requirement. None of the participants had participated in Experiment 1 or 2.

Apparatus and stimuli

The material was identical to that of Experiment 2.

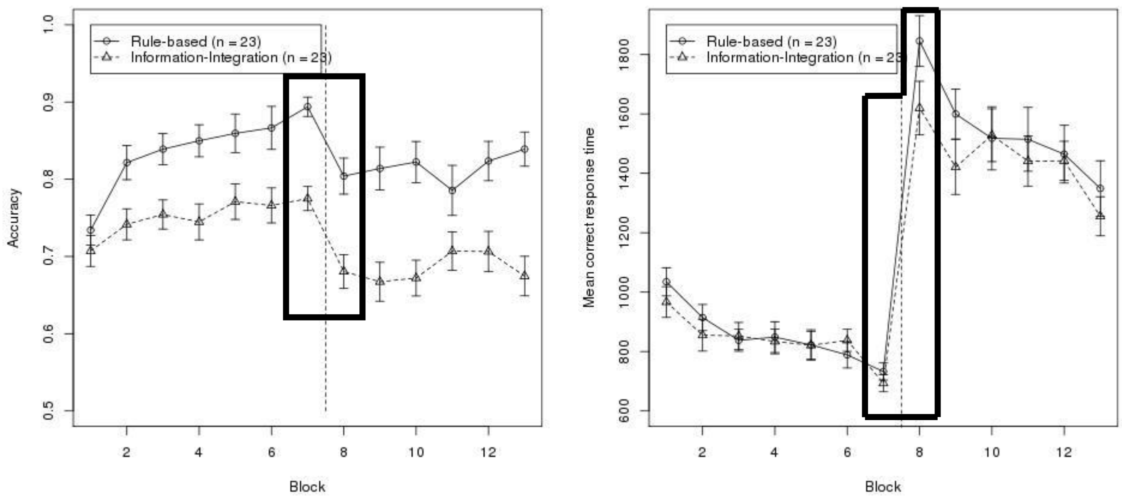
Procedure

The procedure was identical to that of Experiment 2, except for target-present trials in the visual search task. For these trials, each stimulus in the display was equally likely to be a member of category "A" or "B", with the restriction that at least one of the stimuli in the display was from the target category. Participants were informed that more than one target might be present in each display.

Results

Accuracy

The mean accuracy per block for each condition is shown in Figure 6a. As in Experiment 2, Blocks 1 to 7 are from the categorization phase while Blocks 8 to 13 are from the visual search phase. Again, participants in both conditions improved with practice in the categorization phase, but not much in the visual search phase. In both tasks, the RB condition was easier than the II condition. We performed separate ANOVAs for each phase. For the categorization phase, a 2 (RB vs. II) × 7 (Block) ANOVA confirmed the previous observations. The effects of Block ($F(6, 258) = 13.76$, $p < .001$) and Condition ($F(1, 43) = 14.49$, $p < .001$) both reached statistical significance. However, these main effects must be interpreted with care since the interaction also reached statistical significance ($F(6, 258) = 2.30$, $p < .05$). We proceeded to decompose the effect of Block within each level of Condition. Participants improved their performance in the RB task ($F(6, 126) = 10.37$, $p < .05$), with accuracies increasing from 73.4% in Block 1 to 89.4% in Block 7. In the II task, participants also improved their accuracies ($F(6, 132) = 3.62$, $p < .01$), with accuracies increasing from 70.7% in Block 1 to 77.5% in Block 7. The interaction was thus produced by a larger improvement in the RB condition.



(a) (b)

Figure 6. (a) Mean accuracy for each condition in Experiment 3. (b) Mean correct response time in Experiment 3. In both panels, the vertical dashed line separates the categorization phase from the visual search phase. The rectangle highlights the relevant comparison to test for transfer. Error bars represent one standard error of the mean.

In the training phase, we note that participants in the RB condition are more accurate in Experiment 3 (a significant Condition effect) but slower (compare Figure 6b with Figure 4b). Hence, it is possible that the participants in this experiment understood the experimental instructions as implying a stronger emphasis on accuracy relative to

Experiment 2. This strategy shift (slower, more accurate responding) suggests that RB participants may have been more cautious (although the instructions were the same). This may impact the visual search performed at transfer, reducing the likelihood of a premature stop. If this explanation is correct, the miss rate should be reduced in the VSC task, thus restoring an asymmetry between the false alarm rate and the miss rate.

A 2 (RB vs. II) × 6 (Block) ANOVA was computed for the visual search task. Similar to the categorization task, accuracies were higher in the RB condition than in the II condition ($F(1, 44) = 15.99$, $p < .001$). Participants in both conditions did not improve much with practice ($F(5, 220) = 1.21$, *n.s.*), but the interaction between the factors reached statistical significance ($F(5, 220) = 3.23$, $p < .01$). We again proceeded to decompose the effect of Block within each level of Condition. For the RB condition, the effect of Block was statistically significant ($F(5, 110) = 2.90$, $p < .05$), showing that the lowest accuracy (78.6%, Block 11) differed from the highest accuracy (83.9%, Block 13). In contrast, the effect of Block was not statistically significant in the II condition ($F(5, 110) = 0.44$, *n.s.*).

As in Experiment 2, we also tested transfer performance from the last block of categorization to the first block of visual search (the black rectangle in Figure 6a). A 2 (RB vs. II) × 2 (Block 7 vs. Block 8) ANOVA was performed on accuracies. As suggested by the Figure, accuracies decreased in both conditions when transferring from the categorization phase to the visual search phase ($F(1, 44) = 39.52$, $p < .001$). The effect of Condition also reached statistical significance, with accuracies in the RB condition being higher than accuracies in the II condition both before and after transfer ($F(1, 44) = 29.75$, $p < .01$). Finally, the Condition × Block interaction failed to reach statistical significance ($F(1, 44) = 0.03$, *n.s.*).

<u>Response times</u>

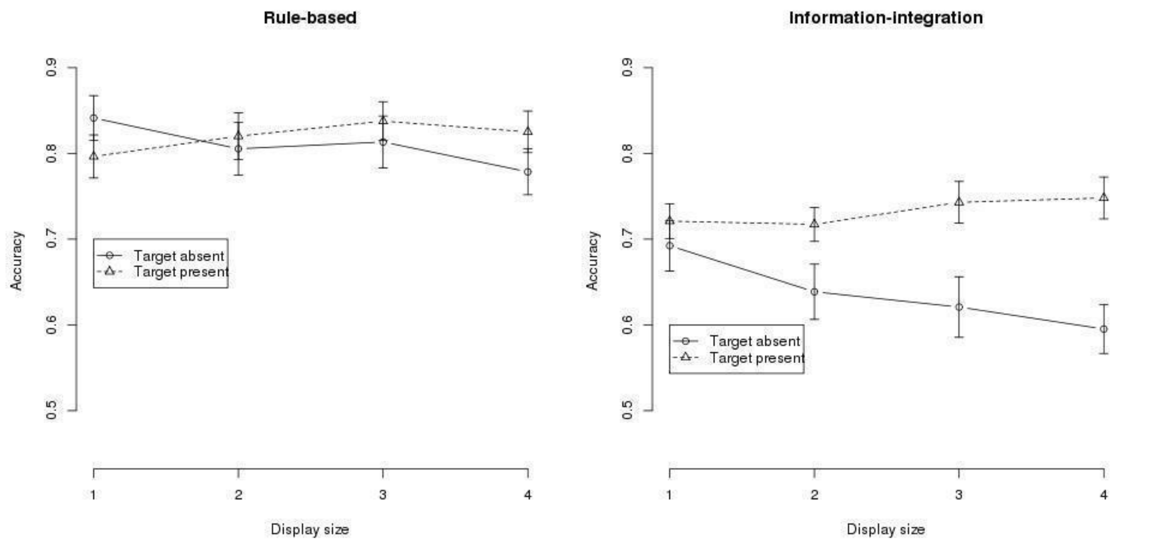
The mean correct RTs are shown in Figure 6b. As can be seen, participants became faster with training in both phases in both conditions. Similar to accuracies, we computed separate ANOVAs for the categorization and visual search phases. For the categorization task, a 2 (RB vs. II) × 7 (Block) ANOVA confirmed the previous observation. The effect of Block was statistically significant ($F(6, 258) = 17.35$, $p < .001$), with mean correct RTs decreasing from 1,000 ms (Block 1) to 713 ms (Block 7). However, unlike accuracies, the effect of Condition ($F(1, 43)$ 0.11, *n.s.*), and its interaction with Block ($F(6, 258) = 0.98$, *n.s.*), both failed to reach statistical significance. This is different from Experiment 2 where the effect of Condition was also significant.

A similar 2 (RB vs. II) × 6 (Block) ANOVA was computed for the visual search phase. The pattern of results was analogous to that of the categorization phase. Mean correct RTs decreased with practice in both conditions ($F(5, 175) = 12.71$, $p < .001$). Mean correct RTs decreased from 1,739 ms (Block 8) to 1,311 ms (Block 13). As in the categorization phase, the effect of Condition ($F(1,35) = 1.55$, *n.s.*) and its interaction with Block ($F(5, 175) = 1.49$, *n.s.*) both failed to reach statistical significance. As was seen in Experiment 2, RTs are more variable in the VSC than in the categorization task (as seen by larger error bars).

Finally, we tested transfer performance from the last block of categorization to the first block of visual search (the black rectangle in Figure 6b). A 2 (RB vs. II) × 2 (Block 7 vs. Block 8) ANOVA was performed on correct RTs. As suggested by the Figure, RTs increased in both condition when transferring from the categorization phase to the visual search phase ($F(1, 35) = 266.77$, $p < .001$). The effect of Condition also reached statistical significance, showing that correct responses in the II condition were faster than those in the RB condition ($F(1,35) = 5.08$, $p < .05$). In contrast, the interaction between the factors failed to reach statistical significance ($F(1, 35) = 3.38$, $p = .07$).

Display size in the VSC task

Figure 7 shows visual-search accuracies across all blocks for target-present and target-absent trials as a function of display size for each category structure. As in Experiment 2, accuracies for target-present trials seemed to be relatively stable across display size for both category structures. Regarding RB, there is little difference between the target-absent and target-present accuracies. Accuracies are quite high, but with similar and near-zero slopes (the slope is +0.7% per additional items for target-present ± 0.6%, and -1.9% per additional items for target-absent, ± 0.6%). The fact that the miss rate is smaller than the false alarm rate is more in line with the literature on visual search task. This is in agreement with the possible results from the categorization phase suggesting that participants might be more exhaustive in the present experiment relative to Experiment 2. This is not however a focus of this article, just a small strategic change, as seen by the limited influence on accuracy.



(a) (b)

Figure 7. The effect of display size on (a) rule-based and (b) information-integration accuracies in Experiment 3. Error bars represent one standard error of the mean.

Regarding II, we see a marked distinction between target-present and target-absent trials, as was the case for Experiment 2. For target-present trials, accuracy improved by about +0.9% ± 0.8% for every additional item whereas for target-absent trials, accuracy decreases by about 3.1% for every additional item, ± 0.7%. These slopes are qualitatively similar to those of Experiment 2 in the same conditions (-0.1% and -5.0% per item respectively for target-present and target-absent trials). There is an offset of 1.5% towards more accurate responses in the slopes on average (relative to Experiment 2) but this offset is present in both target-present and target-absent trials.

To verify these observations, we performed a separate 2 (target-present vs. target-absent) × 4 (Display Size) ANOVA for each category structure. For the RB category structures, the effects of Display Size ($F(3, 66) = 2.19$, $p < .10$) and target presence/absence ($F(1, 22) = 0.83$, *n.s.*) both failed to reached statistical significance. However, the interaction between the factors reached statistical significance ($F(3, 66) = 5.98$, $p < .01$). We decomposed the interaction by looking at the effect of Display Size separately for target-present and target-absent trials. For target-present trials, the effect of Display Size was statistically significant ($F(3, 66) = 2.75$, $p < .05$), with higher levels of complexity leading to higher accuracy (increasing from 79.6% to 82.5%; as noted above, a slope of +0.7% per item, ± 0.6%). In contrast, the effect of Display size was also statistically significant for target-absent trials ($F(3, 66) = 5.99$, $p < .01$), but this time higher levels of complexity were associated with lower accuracy (decreasing from 84.1% to 77.9%); a slope of -1.9% per item, ± 0.6%).

For II categories, there was an effect of Display Size ($F(3, 66) = 2.46$, $p < .05$), target presence/absence ($F(1, 22) = 16.07$, $p < .001$), and a significant interaction ($F(3, 66) = 7.25$, $p < .001$). We again decomposed the interaction by looking at the effect of Display Size separately for target-present and target-absent trials. Similar to Experiment 2, the effect of Display Size did not reach statistical significance for target-present trials ($F(3, 66) = 1.38$, *n.s.* with a slope of 0.9% per item, ±0.6%), but there was an effect of Display Size for target-absent trials ($F(3, 66) = 6.00$, $p < .01$ with a slope of -3.1% per item, ±0.7%).

Discussion

The results in Experiment 3 are similar to those obtained in Experiment 2. Specifically, overall accuracy decreases as display size increases, especially for target-absent trials with II stimuli. Target redundancy did not help much with II categorization. This is perhaps not surprising given that the collapse with II categorization was caused by target-absent trials in Experiment 2, and target-absent trials lack redundant targets (by definition). However, the collapse could have been countered by increased accuracies with larger display size in target-present trials. This is possible because if the display size is 4 and all 4 stimuli are target, the probability of 4 misses is smaller than the probability of 1 miss when only one target is present. However, this result was not observed in Experiment 3, suggesting that the reduced probability of misses in a redundant environment was absent or insufficient in these more complex environments. This result challenges the assumption that each categorization decision is strictly independent. A partially parallel search (such as Guided Search; Wolfe, 1994) or a parallel search with severely limited capacity (Cousineau & Shiffrin, 2004) could be applicable to the present results. An "odd-man out" strategy does not capture the results. Whatever the details of the models, none of them explain the collapse in accuracies observed exclusively in II conditions during a VSC task.

General Discussion

This article presents the results of three experiments exploring the effect of display size on perceptual categorization as a function of category structures. In Experiment 1, we showed that participants can transfer from an A/B training categorization paradigm to a YES/NO categorization paradigm regardless of category structures. Maddox and his colleagues (2004) and Ell et al. (2017) have shown that II category structures are difficult to learn using a YES/NO categorization paradigm [but see Hélie et al. (2017) for unimpaired YES/NO learning with II categories]. Hence, if participants are treating the visual search phase of the VSC task as a sequence of YES/NO categorization judgments, then the category structures could be responsible for the collapse of category knowledge in complex displays observed by Smith et al. (2005). However, the results from Experiment 1 suggest that category structures are not as important in performing the YES/NO task if participants are already familiar with the categories (which was the case in Smith et al., 2005). Hence, Experiment 1 ruled out this possibility, and also suggests that the incompatibility of the YES/NO task with II category structures observed in some earlier studies only applied to initial category learning, not categorization performance.

Experiment 2 further explored the effect of category structures on display size by using the VSC task with two different category structures, namely an RB and an II condition. The results show that display size affects performance in the VSC task mostly with II stimuli, suggesting that the complex polygon stimuli used by Smith et al. (2005) may have required the integration of several stimulus dimensions at a pre-decisional level. The results further show that the collapse in categorization accuracy is driven mostly by false alarms in target-absent II trials. While Smith et al. did not provide separate analyses for target-present and target-absent trials, their results may also be driven by false alarms. Another possibility is that the stimuli overlapped in some of Smith et al.'s experiment, which was not the case in this article. Experiment 3 reproduced Experiment 2 but allowed for target redundancy. The results were similar to those obtained in Experiment 2, suggesting that target redundancy did not play a major role in the VSC task. Experiment 3 also suggests that while each stimulus needs to be individually categorized in order to perform the VSC task, the individual categorization decisions may not be strictly independent.

Why is the visual-search categorization task so difficult?

One important property of the visual search task performed by the participants was that the searched-for categories alternated randomly across trials. On a given trial, a participant may search for a member of an A category and on another trial, that same participant may search for a member of a B category. Alternation of the target categories randomly across trials is called a “Categorical Varied Mapping” condition (CVM; Shiffrin & Schneider, 1977; Schneider & Schiffrin, 1977; Cousineau & Larochelle, 2004). This training schedule is known to be very difficult, and in fact, if the two alternating categories are not pre-existing in the participant, levels of performance are similar to those obtained in a Varied Mapping condition: the search process is said to be controlled, slow and strongly affected by display size. In light of this distinction, it is not surprising that Maddox et al. (2004) and Ell et al. (2017) found limited category learning for the II stimuli. In the present Experiments 2 and 3, the participants were given 7 blocks of training before being transferred to a CVM visual search task. Is this period long enough to have strong enough category representations? Additional experiments in which a single category is the target category for a given participant might be needed.

Another important characteristic of the VSC task is that participants are not shown the target at the beginning of each trial: they were only shown a category label. While Yang and Zelinsky (2009) have shown that search can be driven by a categorically defined target, Vickery, King, and Jiang (2005) have shown that providing a target cue that does not exactly match the target is detrimental. Among all the conditions tested in Vickery et al., providing a verbal label was the worst type of target cue. Given the additional difficulty from the (1) limited training with the categories, (2) CVM training schedule, and (3) verbal target cues, it might not be all that surprising that the VSC task is difficult for participants. We tried to make the task easier by reducing both the overlap between the stimuli as well as the display size, but these differences were not sufficient to make the VSC task accessible with II stimuli. Given these constraints, participants may be more successful if an example category member is shown as a target cue at the beginning of each trial.

Theoretical implication

The results in Experiments 2 and 3 suggest that while II categorization accuracy collapses with display size, the collapse is mostly seen by an increased proportion of false alarms in larger display sizes. This means that the complex polygon stimuli were likely treated as II stimuli. This conclusion is further supported by the fact that abstract polygons have been used in past visual search experiment as example stimuli that "could not be verbally labeled" (Vickery et al., 2005, p. 82). The possibility (or impossibility) of verbally describing a rule to separate categories has often been used as a heuristic to distinguish RB (verbal) from II (non-verbal) categories. However, more research is needed to confirm that complex polygons are treated as II categories.

In Smith et al. (2005), the only factor that reduced the rate of collapse was increasing the within-category similarity, so this implies that reducing the variance of the II categories may reduce the false alarm rate in complex displays. The results in the VSC task also further support the hypothesis that RB and II learning produce different types of category representations that can be used and generalized differently. Similar to Hélie and Ashby (2012), it seems that the representation built from RB stimuli is more easily amenable to transfer in a new task context. One possible explanation is that in the case of RB stimuli, repetitive rule application may produce rule priming, resulting in better than expected result. Another possibility is that visual search performance relies on the feasibility of distinguishing between targets and distractors (Yang & Zelinsky, 2009), and Hélie et al. (2017) showed that RB categorization yields mental representations that contain between-category information (i.e., distinguishing features between the categories) whereas II categorization yields mental representations that contain within-category information (i.e., what is common among category members). Hence, the mental representation learned with RB categorization may be more appropriate for the goal of distinguishing targets (one category) from distractors (the other category). Finally, it is also possible that, because the RB representation may be more digital (Hélie & Cousineau, 2015), participants may be applying rules separately on each stimulus dimension (Ashby, Alfonso-Reese, Turken, & Waldron, 1998) and focus the search on stimuli that have at least one dimensional value within the range of the target category (Treisman & Gelade, 1980, Treisman & Sato, 1990, Wolfe, 1994, Wolfe & Gancarz, 1997). This strategy, which would facilitate and speed-up the search, would not be easily available when searching for II stimuli because II category representations may be more analog and not decompose the stimuli into separate dimensions, which would force the participant to search the whole display (Lefebvre, Cousineau, & Larochelle, 2008, Cousineau & Shiffrin, 2003).

Limitations and future work

Real-world categorization is often conducted in complex scenes and identifying factors that can help category learning in complex displays is critically important. In this article, we showed that category structure is an important factor to consider when categorizing stimuli from complex displays. However, one important limitation of the present experiment is that the RB categories were easier to learn than the II categories. One consequence of this is that the verbal label designating the target may have produced a noisier search template in the II condition. This may have biased participants to be more likely to accept distractors as potential targets (and produce false alarms). A replication with RB and II categories that are matched for difficulty would eliminate this possibility. Another limitation is that Smith et al. (2005) allowed for overlapping stimuli in the display.

While adding overlap is unlikely to make the task easier, it may affect target-present and target-absent trials differently. Hence, future work should focus on adding perceptual noise and occlusion to explore how perceptual ambiguity interacts with display size and category structure.

Acknowledgements

This work was supported, in part, by the National Science Foundation under Grant #1349677-BCS to SH. All data from all experiments can be obtained by contacting the first author at shelie@purdue.edu.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.

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.

Barlow, H. B. (1997). The knowledge used in vision and where it comes from. *Philosophical Transactions of the Royal Society B*, *352*, 1141-1147.

Brainard, D. H. (1997). Psychophysics software for use with MATLAB. *Spatial Vision, 10* , 443-446.

Cousineau, D., & Larochelle, S. (2004) Visual-Memory search: An integrative perspective. *Psychological Research*, *69*, 77-105.

Cousineau, D., & Shiffrin, R. M. (2004) Termination of a visual search with large display size effects. *Spatial Vision*, *17*, 327-352.

Ell, S. W., Smith, D. B., Peralta, G., & Hélie, S. (2017). The impact of category structure and training methodology on learning and generalizing within–category representations. *Attention, Perception, & Psychophysics*, *79*, 1777-1794.

Faul, F., Erdfelder, E., Lang, A. –G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Hélie, S. (2017). The effect of integration masking on visual processing in perceptual categorization. *Brain and Cognition*, *116*, 63-70.

Hélie, S., & Ashby, F. G. (2012). Learning and transfer of category knowledge in an indirect categorization task. *Psychological Research, 76*, 292-303.

Hélie, S., & Cousineau, D. (2015). Differential effect of visual masking in perceptual categorization. *Journal of Experimental Psychology: Human Perception and Performance, 41*, 816-825.

Hélie, S., Shamloo, F., & Ell, S. W. (2017). The effect of training methodology on knowledge representation in categorization. *PLOS ONE, 12,* e0183904.

Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics, 72*, 1013-1031.

Lefebvre, C., Cousineau, D., & Larochelle, S. (2008). Determinants of automaticity in visual-memory search tasks: Pitting target-distractor similarity against stimulus-response mapping. *Attention, Perception & Psychophysics*, *70*, 1401-1415.

Maddox, W.T., Bohil, C.J., & Ing, A.D. (2004). Evidence for a procedural learning-based system in category learning. *Psychonomic Bulletin & Review, 11*, 945-952.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing. I. Detection, search, and attention. *Psychological Review*, *84,* 1–66.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, *84*, 127–190.

Smith, J. D., Redford, J. S., Gent, L. C., & Washburn, D. A. (2005). Visual search and the collapse of categorization. *Journal of Experimental Psychology: General*, *134*, 443–460.

Treisman, A. & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459-478.

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

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*, 8.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202-238.

Wolfe J. M. & Gancarz G. (1997) Guided search 3.0. In: V. Lakshminarayanan (Ed.). *Basic and Clinical Applications of Vision Science* (pp. 189-192). Dordrecht, Netherlands: Springer.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*, 2095–2103.