2020

The effect of pre-exposure on overall similarity categorization

Milton, F

http://hdl.handle.net/10026.1/14753

10.1037/xan0000226
Journal of Experimental Psychology: Animal Learning and Cognition
American Psychological Association

The effect of pre-exposure on overall similarity categorization.

Fraser Milton[1], I.P.L. McLaren[1], Edward Copestake[1], David Satherley[1] &

Andy J. Wills[2]

[1] Discipline of Psychology, University of Exeter, Exeter, U.K.
[2] School of Psychology, University of Plymouth, Plymouth, U.K.

Address for correspondence:
Fraser Milton
Washington Singer Laboratories,
Perry Road,
Exeter
U.K.
EX4 4QG

Tel: +44 1392 725884
E-mail: f.n.milton@exeter.ac.uk

Abstract

This paper examines the effect that prior exposure to perceptual stimuli has on the prevalence of overall similarity (family resemblance) categorization. Experiment 1 demonstrated that participants who had previously encountered stimuli produced more overall similarity sorting when asked to free classify them than participants who were pre-exposed to different stimuli to those they later classified. Experiments 2a and 2b showed that this effect is modulated by the perceptual difficulty of the stimuli - pre-exposure statistically increased overall similarity sorting for perceptually easy stimuli but not for perceptually difficult stimuli. Overall similarity sorting was also significantly higher for perceptually easy stimuli than for perceptually difficult stimuli. Experiment 2b additionally showed that pre-exposure increased the discriminability of the perceptually easy stimuli but this effect was not statistically detectable for perceptually difficult stimuli. Experiment 3 established that the pre-exposure effect is also influenced by the spatial separateness of the stimulus dimensions - pre-exposure significantly elevated overall similarity sorting when the dimensions were integrated into a coherent object but not when they were spatially separated. Similarly, there was a statistically significant increase in the perceptual discriminability of the spatially integrated stimuli after pre-exposure but not for the spatially separate stimuli. Taken together, these results demonstrate that pre-exposure can elevate overall similarity sorting and provide insight into the conditions under which the effect will occur.

Key words: overall similarity; free classification; perceptual learning; pre-exposure; categorization.

The ability to group items into meaningful categories is a fundamental cognitive process that helps us make sense of the world we live in. For example, it allows us to make inferences about objects that we have never seen before and to treat different objects in the same way, greatly simplifying the complex environment that we live in. However, due to the vast number of different items we encounter outside the lab, this process must inevitably be highly constrained. A key question, then, is how do we acquire the categories that we have?

Categorization has traditionally been examined using supervised learning procedures where participants are required to learn experimenter-defined categories and are provided with trial-by-trial feedback on their responses (e.g., Medin & Schaffer, 1978; Shepard, Hovland, & Jenkins, 1961). While this approach has undoubtedly provided great insight, there is a growing acknowledgement that it is important also to examine other conditions in which we acquire categories (e.g., Love, 2002). In unsupervised categorization – also known as free classification (e.g., Imai & Garner, 1965), free sorting (e.g., Bersted, Brown, & Evans, 1969), category construction (e.g., Medin, Wattenmaker, & Hampson, 1987), and spontaneous categorization (e.g., Pothos & Close, 2008) - participants are given a set of stimuli and asked to sort them in the way that seems most sensible and natural to them with no feedback provided on their responses. This approach is ideal for providing insight into the way people naturally choose to create categories.

One reasonable assumption is that people would have a preference to create categories that are consistent with the way items are organized outside the laboratory. The “classical” view proposes that categories are organized around necessary and jointly sufficient defining features (e.g., Bruner, Goodnow, & Austin, 1956) – as long as an item possesses the particular defining feature (or features) it is a member of that category regardless of the properties of the remaining features. However, this theory has become less influential due to the difficulty of finding defining features for many natural categories. Instead, many natural categories appear to possess a family resemblance or overall similarity structure (e.g., Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) where categories are organized around a number of characteristic but not defining features. If an item has enough features characteristic of a category, it can be considered a member of that category even if it does not have a particular feature. One advantage of overall similarity categories is that they are typically believed to be a more information-rich structure than unidimensional categories and have been considered especially useful for identification, inference, problem-solving, and other cognitive tasks (Murphy, 2002). For example, Hoffman and Murphy (2006) note that classifying an object as a robin allows one to infer perhaps as many as 100 properties that are characteristic of the category “robin”.

Surprisingly, however, initial work indicated that when participants are asked to group items without any feedback they have a strong tendency to create categories based on a single dimension and rarely sort by overall similarity (e.g., Ahn & Medin, 1992; Ashby, Queller, & Berretty, 1999; Imai & Garner, 1965; Medin et al., 1987). This approach appears more consistent with the classical view than with a family resemblance theory. More recent work has revealed a more nuanced picture with manipulations of stimulus presentation method (Regehr & Brooks, 1995), the spatial integration of stimulus dimensions (Milton & Wills, 2004; Milton & Wills, 2009), the category structure (Pothos & Close, 2008), instructions (Wills, Milton, Longmore, Hester, & Robinson, 2013), and background knowledge (Spalding & Murphy, 1996) all influencing the prevalence of overall similarity categorization. Nevertheless, even in these studies overall similarity sorting is typically far from ubiquitous. An important question, therefore, is to understand why the categories we prefer to create do not reflect the commonly assumed underlying structure of natural world categories.

One notable aspect of many of the studies cited above is that participants had little or no exposure to the stimuli prior to classifying a very limited number of items (e.g., Ahn & Medin, 1992; Medin et al., 1987; Milton & Wills, 2004). This appears atypical of categorization outside the lab where we usually have had a great deal of exposure to the objects we categorize. One possibility, therefore, is that this lack of familiarity with the stimuli is contributing to the dearth of overall similarity categorization in these studies. Instead, limited experience with the stimuli may pre-dispose participants to fall back on a simplistic, unidimensional, strategy as they have not had sufficient experience with the stimuli to identify the overall similarity structure that organizes them. One prediction, then, is that if participants receive substantial pre-exposure to the stimuli prior to classification this may increase the probability that they will group them by overall similarity (see Milton & Wills, 2004, for an earlier discussion of this prediction).

While there is an extensive body of work looking at the influence of pre-exposure on the ability to differentiate perceptual stimuli, the vast majority of this work has used response accuracy to measure its effect (e.g., McLaren, 1997). This work typically demonstrates that prior experience has a beneficial effect on accuracy (e.g., McLaren, Leevers, & Mackintosh, 1994). A related, but to date surprisingly neglected, question is the extent to which pre-exposure can actually change the *nature* of the categories that we create. One of the few published studies to look at this was by Wills and McLaren (1998) who used a free classification procedure to show that pre-exposure can influence the number of categories people use. While intriguing, that work, which used complex checkerboard stimuli, provides no insight into the current question of whether pre-exposure can modulate the level of overall similarity categorization.

More directly related to this question is Experiment 3 of Spalding and Murphy's (1996) paper that examined the influence of background knowledge on overall similarity sorting. In this particular experiment, participants were provided with verbal labels giving details about the features of an instance relating to domains such as vehicles (e.g., Made in Africa/Made in Norway; Lightly insulated/Heavily insulated; Green/White; Drives in jungles/Drives on glaciers) which were either organized so that there was a meaningful theme (i.e., vehicles suitable for driving in Africa or Scandinavia) or were arranged so that there was no coherent theme. Spalding and Murphy found that when there was a coherent integrated theme connecting the dimensions, participants who previewed the stimuli prior to categorization produced a greater level of overall similarity categorization than those who had no preview. This difference was not present, though, when there was no coherent theme connecting the dimensions – no overall similarity sorting was observed in either condition. Spalding and Murphy argued that the preview provided participants with a greater opportunity to detect the coherent theme and the inter-correlation of features which would then facilitate overall similarity categorization.

On the other hand, Milton and Wills (2004), using stimuli with perceptual dimensions rather than verbal labels, compared the prevalence of overall similarity categorization for participants who were pre-exposed to the stimuli via a matching-pairs task and those who were not. In the matching-pairs task, participants were provided with two copies of each of the ten stimuli in the set and were asked to put them into identical pairs. This ensured that they could identify all of the stimulus dimensions that varied and gave them some initial exposure to the stimuli prior to categorization. However, this manipulation had no effect on the level of overall similarity sorting. This may reflect that the limited amount of time that participants were exposed to the perceptual stimuli (around 2-3 minutes on average) was not sufficient to change the nature of their classifications.

By what mechanism might pre-exposure be expected to elevate overall similarity categorization for a set of perceptual stimuli? One of the most extensively documented effects of pre-exposure is that it leads to perceptual learning which can be defined as the enhanced ability to discriminate between stimuli as a consequence of experience with them or related stimuli (e.g., McLaren, Graham, & Wills, 2010). There are numerous theories of perceptual learning (e.g., Goldstone, 1998; Hall, 1991; Lavis & Mitchell, 2006; McLaren, Kaye, & Mackintosh, 1989; Mundy, Honey, & Dwyer, 2007; Seitz & Watanabe, 2005). For example, the MKM model (McLaren, Kaye, & Mackintosh, 1989; see also McLaren & Mackintosh, 2002; Livesey, & McLaren, 2011) assumes that stimuli are represented by a number of elements. Items that share many common elements will be more difficult to discriminate than items that share few elements. One of the key assumptions of this model is that when elements co-occur, there will be a reduction in the salience of these elements (often referred to as latent inhibition). Consequently, one of the principal effects of pre-exposure is that elements which frequently co-occur reduce in salience more quickly than elements that rarely co-occur. This means that the unique elements that discriminate one stimulus from another will tend to be higher in salience than the common elements that both stimuli share (because the common elements will have been presented more often and because they are good predictors of one another). This preferential processing of the unique elements, which discriminate between items, compared to the common elements, which do not, is what, according to the MKM model, leads to the increased differentiation of stimuli after pre-exposure.

An increased ability to differentiate stimuli might be expected to encourage overall similarity sorting as it would enable people to not only facilitate the processing of the different feature-values of those dimensions (within-dimension differentiation) but also to detect more easily the dimensions of variation between the stimuli (between-dimension differentiation). It could also make the inter-correlation of features and the overall similarity structure more salient in a manner analogous to that which Spalding and Murphy (1996; see also Lassaline & Murphy, 1996) proposed occurs when there is a meaningful theme underlying the category structure. The four experiments presented in this paper provide the first detailed investigation of this hypothesis. If our assumption is correct then it would provide a valuable illustration of the way that pre-exposure can actually change the nature of the decisions made.

Experiment 1

Method

Participants and Apparatus. Fifty students from the University of Exeter were recruited to take part in this experiment either for course credits or £5. For this and subsequent experiments, all participants were aged 18-35 and there was a strong female bias. Participants were tested inside individual cubicles within a multi-testing lab (with up to eight participants being run simultaneously) and the experiment was run using E-prime on a Dell PC with a 17-in. monitor and a standard computer keyboard. No participant took part in more than one experiment in this paper. Ethical approval for this and all subsequent studies was obtained from the University of Exeter School of Psychology Ethics Committee.

Stimuli. The two stimulus sets employed had the same abstract stimulus structure to that used by Medin et al. (1987). This structure is shown in Table 1. Stimuli possessed four binary-valued dimensions (D1-D4) and the stimuli were organized around two prototypes, each representative of one of the categories. These prototypes were constructed by taking all the positive values on the dimensions for one of the stimuli (1, 1, 1, 1) and all of the zero values on the dimensions (0, 0, 0, 0) for the other category. The rest of the stimuli were mild distortions of the two prototypes in that they had three features characteristic of their category and one atypical feature more characteristic of the other category. In total, there were 10 stimuli in each set. Sorting the stimuli by overall similarity, as shown in Table 1, maximizes within-group similarities and minimizes between-group similarities.

The category prototypes for the two stimulus sets are shown in Figure 1. One stimulus set were artificial lamps first used by Milton & Wills (2004; see also Milton, Wills, & Hodgson, 2009). The four dimensions were the number of dots in the lampshade (few/many), the width of the stem (thin/thick), the color of the top part of the base (light blue/dark blue) and the length of the bottom part of the base (narrow/wide). The boat stimuli were first used by Milton, Longmore, and Wills, (2008) and were inspired by Lamberts (1998). The dimensions were the shape of the flag (square/triangle), the size of the sail (small/large), the shape of the porthole (circle/diamond) and the length of the hull (wide/narrow).

Design. Participants were randomly allocated to one of the two between-subject conditions. In the *same-stimuli* exposure condition, participants were pre-exposed to the same stimuli that they later classified. For example, they were pre-exposed to and then classified the boat stimuli. In the *unrelated-stimuli* exposure condition participants were pre-exposed to different stimuli to those they later classified (e.g., they were pre-exposed to the lamp stimuli and subsequently classified the boat stimuli). The stimulus set that participants classified was randomized. In total, there were 24 participants in the same-stimuli condition (12 classified the boat stimuli, 12 classified the lamp stimuli) and 26 participants in the unrelated-stimuli condition (15 classified boats, 11 classified lamps).

Procedure.

The running-recognition phase. In both the same-stimuli and unrelated-stimuli conditions, participants were pre-exposed to the appropriate set of stimuli via a running-recognition task (e.g., Wills & McLaren, 1998). We chose a running-recognition task as it should encourage participants to actively process the stimuli and attending to all the stimulus dimensions would be needed to perform optimally on the task (we consider the impact that different methods of pre-exposure may have had on our results in the General Discussion). The instructions for this task are displayed in the Appendix. Each of the ten stimuli in the set was presented twice in each block in a random order. Each trial began with a black fixation cross presented in the middle of the screen lasting 500ms. This was immediately followed by one of the stimuli from the set appearing in the middle of the screen for 3000ms. Participants could not respond during this time. After this, the stimulus then immediately disappeared and participants were asked to press "x" if they had seen that stimulus before in that particular block and "m" if they had not. This response was self-paced. Following this, the next trial immediately began. At the end of each block, participants were informed of their accuracy in that block. In total, participants in both conditions completed sixteen blocks of 20 trials.

Categorization phase. The categorization procedure was identical for the same-stimuli and unrelated-stimuli conditions and immediately followed the running-recognition pre-exposure phase. We used a computer-based variation of Regehr and Brooks's (1995) match-to-standards procedure that was the same as that adopted in Milton et al. (2009). Participants were asked to classify a set of stimuli into two categories (participants were not told prior to this point that they would be completing a categorization task). They were informed that there were many ways in which the stimuli could be split and that there was no correct answer. Participants were asked to classify the stimuli in the manner they thought most appropriate. The full instructions are provided in the Appendix.

At the beginning of each trial a black fixation cross was presented for 500ms in the center of the screen. The two category prototypes were then presented at the top of the screen and below these prototypes in the center of the screen one of the ten stimuli in the set (E1-E10 in Table 1) was displayed. Participants categorized this stimulus either into category A by pressing “x” on the keyboard or into category B by pressing “m”. This decision was self-paced and the stimulus immediately disappeared when a response had been made. No feedback was provided on the responses and instead a blank screen was then presented for 1000ms before the next trial began. Each of the stimuli in the set appeared once in each block in a random order. At the end of each block, participants were asked to write down as precisely as possible in a booklet provided how they categorized the stimuli in the previous block. Participants then began the next block when they were ready (in other words, there was no extra pre-exposure of the stimuli in between each of the categorization test blocks). In total, there were six blocks. The inclusion of multiple blocks provided the opportunity to build up a reliable index of an individual's sorting behavior rather than relying on a limited number of responses from a single block. Previous work indicates a close correspondence between multiple block procedures and single block procedures (Milton et al., 2008).

Analysis of results. In all experiments in this paper, each participant was classified as having produced one of the sort types described below. These sort types are similar to those employed by Regehr and Brooks (1995) and are identical to those used by Milton and Wills (2004) and subsequent studies from our lab (e.g., Milton et al., 2008; Wills et al., 2013). To be classified as sorting by either overall similarity or unidimensionally, both the participant's description of their strategy and their behavioral response were required to be consistent with each other. As in previous work (e.g., Wills et al., 2013), each block was categorized independently.

An *overall similarity* sort, also commonly known as a "*family resemblance*" sort, has the same structure as shown in Table 1. In this type of strategy, the participant has to place each of the prototypes, along with their derived one-aways, into separate categories without error. Additionally, they have to describe their strategy as being based either on general similarity or by placing each item into the category with which it had more features in common. Participants were classified as producing a *one-away overall similarity* sort if they grouped items in the same way as for an overall similarity sort but there was a single error in their classification.

A *unidimensional* sort is based on a single dimension of the stimulus. It does not matter which of the dimensions is used as the basis of sorting, so long as all of the positive values for the chosen dimension are placed in one category and all of the zero values for that dimension are in the other category. Additionally, to be classified as a unidimensional sort, the participant has to describe their sort as being based on a single dimension. In a *one-away unidimensional* sort, participants described their decision as being driven by a single dimension but there was a solitary error in their classification.

Any classifications other than those described above were classified as *other* sorts, even if the description given by the participant fitted one of the sorts described above.

Results and Discussion

The raw data for this experiment are publicly available at: https://osf.io/qxtw9/

Running-recognition phase. Recognition accuracy was measured using d' with 0 indicating chance performance. The mean accuracy across blocks is shown in Figure 2a (note that the same-stimuli and unrelated-stimuli conditions are identical in the running-recognition phase so we collapsed across this factor here and for similar analyses in subsequent experiments). Recognition accuracy significantly improved across training, $F(15,735) = 2.04$, $p = .011$, $\eta^2_p = .04$.

Categorization phase. For every block, each participant's sorting strategy was classified according to the sort types described above. One-away unidimensional and one-away overall similarity sorts were classified as unidimensional and overall similarity sorts respectively (cf., Milton & Wills, 2004; Milton et al., 2008).

The mean proportion of overall similarity, unidimensional and other categorizations produced in the same-stimuli and unrelated-stimuli conditions are shown in Figure 2b. There was no statistical difference in categorization behavior between the lamp and boat stimuli for any of the analyses so we therefore collapsed across these two stimulus sets in the subsequent analyses (this is also the case in all subsequent experiments). The proportion of overall similarity categorization was .39 greater in the same-stimuli condition than the unrelated-stimuli condition, $t(48) = 4.15$, $p<.001$, $d = 1.16$.

Unidimensional categorization was .15 less in the same-stimuli condition than the unrelated-stimuli condition but this difference was not statistically significant, $t(48) = 1.31$, $p = .197$, $d = 0.37$. The proportion of unidimensional sorts in which participants used their most commonly selected dimension was .82 (same-stimuli = .84; unrelated-stimuli = .80, $t(28) = .42$, $p = .68$)[1].

Other sorts were .24 greater in the unrelated-stimuli condition than the same-stimuli condition, $t(48) = 2.35$, $p = .023$, $d = 0.67$. Using the self-reports, Other sorts were further classified to examine what strategies participants were attempting. These were divided into three categories: failed overall similarity (.36), failed unidimensional sorts (.22) and other idiosyncratic strategies (.42). Using this information, together with the sorts that had previously been classified as either overall similarity or unidimensional, we calculated the mean proportion of self-reports that were consistent with each participants' most commonly reported strategy. Overall, participants self-reported using the same strategy on .83 of sorts.[2]

[1] For subsequent experiments we report only the mean proportion collapsed across all conditions, but further information at the condition level is provided in Section A of the Supplemental Materials.

[2] Please see Section B of the Supplemental Materials for descriptive statistics displaying the breakdown across conditions for these measures for this and subsequent experiments.

The results of Experiment 1 provide the first demonstration that prior exposure to a set of perceptual stimuli leads to a greater level of overall similarity sorting than exposure to a different set of stimuli to those subsequently categorized. Participants who had not been pre-exposed to the stimuli produced a greater level of Other sorts suggesting that they struggled to apply a coherent strategy consistently without having any previous experience with them. This, of course, is a single demonstration of a novel finding and in Experiment 2 we attempt to generalize the effect to different stimulus sets containing a greater number of dimensions and to explore in more depth the conditions under which pre-exposure influences the nature of the categories we create.

Experiment 2a

Our explanation for the results of Experiment 1 is that same-stimuli pre-exposure increased the discriminability of the stimuli and that this made an overall similarity response easier to perform than in the unrelated-stimuli condition. Previous work has shown that perceptual learning is most pronounced when the stimuli are perceptually difficult to discriminate (e.g., Oswalt, 1972) and that pre-exposure can even slow learning if the stimuli are sufficiently different (e.g., Chamizo & Mackintosh, 1989).

The MKM model (e.g., McLaren et al., 1989) proposes that stimuli are made up of a combination of common elements and unique elements which differentiate the stimuli. In the case, for example, of the ladybird prototype stimuli in Figure 3 (top left), common elements would include the dots that both stimuli share. However, the ladybird stimulus on the left has more dots than the one on the right while the one on the right has a larger red colored surface. These differences would constitute unique elements. According to MKM pre-exposure should draw attention to the aspects of the stimuli that differ. MKM posits that, for perceptually similar items, there are many common elements which will lead to a greater benefit accruing from pre-exposure compared to perceptually different items where there are few common elements. In Experiment 1, the stimuli in each category are relatively similar to each other (e.g., both categories are boats which share a very similar configuration) which is likely to have encouraged perceptual learning to some extent. Nevertheless, it is clearly possible to further increase the similarity between the stimulus sets – for example, the lamp stimuli used in Experiment 1 were the easier to perceptually discriminate of two stimulus sets, sharing the same basic dimensions, employed in Experiment 5 of Milton and Wills (2004). We therefore adopted a similar approach to Milton and Wills by creating a set of perceptually easy stimuli, that were comparable in perceptual difficulty to the stimuli used in Experiment 1 (where there were, for example in the ladybird stimuli, relatively few shared dots) and a perceptually difficult set of stimuli where the differences between the feature values were harder to discriminate because there were more common elements (e.g., there were more shared dots).

We had two main aims with this experiment. First, we wished to generalize the effect we obtained in Experiment 1 with perceptually easy stimuli to new stimulus sets which possessed a different number of dimensions (five rather than four). This change should increase the within-category similarities and increase the inter-correlation of features which could potentially increase the size of the effect. Second, we wished to examine whether we could also obtain the pre-exposure effect with perceptually difficult stimuli and whether it may even be enhanced (on the basis that perceptual learning has been argued to be greater for perceptually difficult discriminations, Oswalt, 1972).

Method

Participants and Apparatus. Students from the University of Exeter participated either for course credits or for a payment of £5. There were sixty-four participants (16 in each of four between-subject conditions) who were tested individually in a quiet testing cubicle. We tested participants using E-prime, on a Dell PC with a 17-in. monitor and a standard computer keyboard.

Stimuli. The four stimulus sets in this experiment had the same basic structure to that used in Experiment 1 with the exception that there were now five, rather than four, binary-valued dimensions (see Table 2). Similar to in Experiment 1, each category was organized around a prototype which possessed all five characteristic features of that category. The remaining stimuli were mild distortions of the two prototypes in that they had four features characteristic of their category and one atypical feature more characteristic of the other category. In total, there were 12 stimuli in each set.

Two of the stimulus sets were based on ladybirds and the other two stimulus sets were based on houses (see Figure 3). The two pairs of stimulus sets were identical except that for one of the sets the binary values for each dimension were relatively easy to distinguish (e.g., for the ladybird stimuli the difference in the leg size was relatively large) and for the other set the differences were relatively difficult to distinguish (e.g., the difference in the leg size was relatively small). We term these sets the perceptually easy and the perceptually difficult stimuli respectively. The five dimensions for the ladybird stimuli were: antennae (length and width between them), the size of the head, the number of dots on the body, the length of the green ovals on the body, and the size of the legs. The five dimensions for the house stimuli were the height of the aerial, the length of the chimney, the number of lines on the roof, the size of the windows, and the size of the door.

Design. The experiment had a 2 x 2 between-subjects factorial design. The first factor was the perceptual difficulty of the stimuli (two levels: *perceptually easy*/*perceptually difficult*). The second factor was the type of pre-exposure, which also had two levels (*same-stimuli*/*unrelated-stimuli*). This led to four conditions: *perceptually easy/same-stimuli* exposure, *perceptually difficult/same-stimuli* exposure, *perceptually easy/unrelated-stimuli* exposure and *perceptually difficult/unrelated-stimuli* exposure. In all conditions, the stimulus set (either ladybirds or houses) that participants classified was counterbalanced.

Procedure. The basic procedure for both the pre-exposure and categorization phases was identical to in Experiment 1. However, because there were now 12 stimuli in the sets rather than 10, there were 16 blocks of 24 stimuli in the running-recognition task, and 6 blocks of 12 stimuli in the categorization task.

Results

The raw data for this experiment are publicly available at: https://osf.io/rgpx6

Running-recognition phase. Mean d' accuracy across conditions is shown in Supplemental Figure 1. Similar to in Experiment 1, accuracy in the running-recognition task improved significantly across blocks, $F(15,930) = 3.27$, $p < .001$, $\eta^2_p = .05$. Mean d' was 0.31 greater in the perceptually easy condition than the perceptually difficult condition, $F(1, 62) = 7.68$, $p = .007$, $\eta^2_p = .11$, but there was no statistically significant interaction between perceptual difficulty and block, F (15, 930) = 0.52, p = .933 η^2_p = .01.

Categorization phase. The mean proportion of overall similarity, unidimensional, and other categorizations produced in the four conditions are shown in Figure 4. The proportion of overall similarity sorting was .39 higher in the perceptually easy condition than the perceptually difficult condition, F(1,60) = 36.07, p <.001, η^2_p =.38. There was also .13 greater overall similarity sorting in the same-stimuli condition than the unrelated-stimuli condition, F(1,60) = 4.22, p = .044, η^2_p = .07, and a significant interaction between perceptual difficulty and pre-exposure type F (1, 60) = 6.40, p = .014, η^2_p = .10. Pairwise comparisons, though, revealed that the direction of the interaction was the opposite to that predicted – for the perceptually easy stimuli, same-stimuli exposure resulted in .29 higher overall similarity categorization than unrelated-stimuli exposure, t(30) = 2.33, p = .027, d = 0.82, but there was no statistical difference between exposure type for the perceptually difficult stimuli, t(30) = -1.38, p = .178, d = 0.49 (mean proportion difference = .03).

The mean proportion of unidimensional sorts was .26 greater for the perceptually difficult stimuli than the perceptually easy stimuli, F(1,60) = 8.58, p = .005, η^2_p = .13. The .11 difference between the exposure conditions was not statistically significant, F(1,60) =1.58, p = .214, η^2_p= .03, but there was a significant interaction between level of perceptual difficulty and exposure type, F(1,60) = 4.38, p = .041, η^2_p = .07. Pairwise comparisons revealed that for the perceptually easy stimuli, unidimensional sorting was .29 lower in the same-stimuli exposure condition than in the unrelated-stimuli exposure condition, t(30) = 2.22, p = .034, d = 0.78. In contrast, the .07 greater unidimensional sorting in the same-stimuli condition compared to the unrelated-stimuli condition was not statistically significant, t(30) = -0.644, p = .528, d

= 0.22. As in Experiment 1, participants who produced unidimensional sorts tended to stick with the same dimension (mean proportion = .86).

The .14 greater Other categorizations in the perceptually difficult condition compared to the perceptually easy condition was not statistically significant, $F(1,60) = 2.87$, $p = .096$, $\eta^2_p = .05$, and neither was the .02 difference between the pre-exposure conditions, $F(1,60) = 0.14$, $p = .71$, $\eta^2_p < .01$. There was also no statistically significant interaction between exposure type and perceptual difficulty, $F(1,60) = 0.51$, $p = .822$, $\eta^2_p = .001$. Other sorts were a mixture of failed overall similarity (.17), failed unidimensional (.59), and other idiosyncratic (.24) strategies. Overall, participants self-reported using the same strategy on .90 of sorts.

Discussion

The results for the perceptually easy stimuli replicate and extend the results of Experiment 1 to different sets of stimuli with a larger number of dimensions - overall similarity categorization was significantly enhanced when participants were pre-exposed to the stimuli prior to sorting them compared to when they were exposed to a different set of stimuli. In contrast, we did not find any statistically significant effect of pre-exposure for the perceptually difficult stimuli indicating that the influence of pre-exposure does not occur for all types of stimuli.

One explanation for our failure to detect an effect of pre-exposure for the perceptually difficult stimuli lies in the finding that there was a significant main effect of perceptual difficulty, with the perceptually easy stimuli evoking more overall similarity sorting than the perceptually difficult stimuli (where overall similarity sorting was close to floor). This result is noteworthy in its own right as a first demonstration of a further factor that modulates the prevalence of overall similarity sorting. It is also in line with our general tenet that overall similarity sorting is more prevalent when the stimuli are easier to differentiate (which can be achieved either via pre-exposure or due to the perceptual characteristics of the stimuli themselves) as this makes such a strategy easier to perform. For the perceptually easy stimuli, pre-exposure further increases the already noticeable differences between the stimuli leading to an additional elevation of overall similarity sorting. For the perceptually difficult stimuli, however, the stimuli are so similar to each other that discriminating the differences across multiple dimensions is an extremely effortful and time consuming process even after pre-exposure which leads participants to categorize based on a subset of the information (e.g., the dimension whose feature values they find easiest to discriminate). This consequently leads to a paucity of overall similarity sorting and a high level of unidimensional classifications.

Experiment 2b

While our explanation for the pattern of findings observed in Experiment 2a appears to have some plausibility, our results were nevertheless different to what we predicted. In Experiment 2b, we therefore aimed to replicate the results of Experiment 2a using different stimulus sets. In addition, we made the assumption both in Experiment 1 and for the perceptually easy stimuli in Experiment 2a that the increase in overall similarity sorting was driven by an elevation in the perceptual discriminability of the stimuli as a result of the relevant pre-exposure. While this supposition appears reasonable given the extensive documentation of perceptual learning effects after pre-exposure (cf., Goldstone, 1998; McLaren & Mackintosh, 2000; Suret & McLaren, 2003), Experiments 1 and 2a provide no direct evidence for this effect. We therefore included a perceptual discrimination task after the classification phase in Experiment 2b to directly test this assumption and, in particular, to compare the relative impact of pre-exposure on differentiating the perceptually easy and the perceptually difficult stimuli.

Method

Participants, Apparatus, and Design. Eighty-three participants took part for either course credits or £5. Three further participants were excluded for failing to complete the experiment. Participants were tested inside individual cubicles within a multi-testing lab (with up to seven participants run simultaneously). As before, the experiment was run using Eprime on a Dell PC with a 17-inch monitor and a standard computer keyboard. Participants were randomly allocated to one of the four between-subjects conditions: *perceptually easy/same-stimuli* exposure (21 participants), *perceptually difficult /same-stimuli* exposure (21 participants), *perceptually easy/unrelated-stimuli* exposure (21 participants) and *perceptually difficult/unrelated-stimuli* exposure (20 participants). In all conditions, the stimulus set (butterflies or lamps) that participants classified was approximately counterbalanced.

Stimuli. The four stimulus sets had the same abstract structure as in Experiment 2a. Two of the stimulus sets were modifications of the lamp stimuli used in Experiment 1 and the other two stimulus sets were modifications of the butterfly stimuli previously employed in Experiment 4 of Milton and Wills (2004). Four of the dimensions of the lamp stimuli were the same as the stimuli used in Experiment 1; the fifth dimension was the size of the triangle on the top of the lampshade. For the butterfly stimuli, the dimensions were: the size of the antennae (long/short), the size of the head (big/small), the number of lines in the top set of wings (many/few), the color of the bottom set of wings (light grey/dark grey) and the length of the tail (long/short). The category prototypes are displayed in Figure 5.

Procedure. The basic procedure of the running-recognition and free classification phases was identical to in Experiment 2a. However, following the categorization phase, participants completed a perceptual discrimination test. Each trial began with a blank screen for 250ms before a black fixation cross appeared in the middle of the screen for 250ms. Immediately after this, two stimuli were presented in the center of the screen, with one being directly above the other. Participants were required to say whether the stimuli were identical (by pressing "x") or different (by pressing "m"). The task was self-paced and the stimuli remained on the screen until participants made their response. Feedback (“Correct” in blue and “Incorrect” in red) was then provided for 750ms. The next trial then immediately began. The stimulus pairs were presented in a random order in a single block of forty-eight trials.

Participants viewed stimuli of the same level of perceptual difficulty as they had encountered previously in the experiment. Twenty-four of these trials displayed a pair of the lamp stimuli and the remaining twenty-four trials presented a pair of the butterfly stimuli. Participants had been pre-exposed to one of these sets during the running-recognition exposure phase (*pre-exposed*) while they had not seen the other stimulus set during the running-recognition task (*not pre-exposed*). Within each of these sets twelve of the stimulus pairs were from one of the categories and the other twelve pairs contained stimuli from the other category. For each category, six of these pairs comprised identical stimuli (e.g., both stimuli were E1; see Table 2) and for the other six pairs the stimuli were different (e.g., one stimulus was E1 and the other was a different stimulus, such as E6, from the same overall similarity category). Pairs that were different varied on one or at most two dimensions. In total, for 24 of the trials the correct answer was "identical" and for the other 24 trials the correct answer was "different". Instructions for this phase are shown in the Appendix.

Results

The raw data for this experiment are publicly available at: https://osf.io/t79sr/

Running-recognition phase. Mean accuracy across conditions is displayed in Supplemental Figure 2. Accuracy again significantly improved across blocks, F (15, 1215) = 7.89, p < .001, η^2_p = .09, and mean d' accuracy was 0.30 better in the perceptually easy condition than the perceptually difficult condition, F(1,81) = 14.45, p< .001, η^2_p = .15. There was no significant interaction between block and perceptual difficulty, F (15,1215) = 0.78, p = .698, η^2_p = .01.

Categorization phase. The mean proportion of overall similarity, unidimensional, and other sorts for each condition are displayed in Figure 6. The proportion of overall similarity sorting was .14 higher in the same-stimuli condition than the unrelated-stimuli condition, F(1,79) = 4.86, p = .030, η^2_p = .06, and .51 greater for the perceptually easy stimuli than the perceptually difficult stimuli, F(1,79) = 61.15, p <.001, η^2_p = .44, with overall similarity sorting again close to floor for perceptually difficult stimuli. There was also a significant interaction between exposure type and perceptual difficulty, F(1,79) = 4.83, p = .031, η^2_p= .06. For the perceptually easy stimuli, the proportion of overall similarity sorts was .29 higher in the same-stimuli condition than the unrelated stimuli condition, t(40) = 2.24, p = .031, d = 0.69. For the perceptually difficult stimuli the proportion of overall similarity sorts was almost identical (a .001 difference), t(39) = 0.04, p = .973, d = 0.01.

The proportion of unidimensional sorting was .55 higher in the perceptually difficult condition than the perceptually easy condition, F(1,79) = 63.03, p <.001, η^2_p= .44. There was no statistical effect of pre-exposure, F(1,79) = 1.16, p = .286, η^2_p = .01, with a mean difference of .07 between the same-stimuli and unrelated-stimuli conditions. However, there was a significant interaction between exposure and perceptual difficulty, $F(1,79) = 4.18$, $p = .044$, $\eta^2_p = .05$. Investigating this interaction further, unidimensional sorting was .22 lower for the perceptually easy stimuli in the same-stimuli condition than the unrelated-stimuli condition although this effect did not reach statistical significance, $t(40) = 1.89$, $p = .067$, $d = 0.58$. In contrast, for the perceptually difficult stimuli unidimensional sorting was .07 higher in the same-stimuli condition than the unrelated-stimuli condition, although this result was again not statistically significant, $t(39) = -0.88$, $p = .385$, $d = 0.28$. The proportion of unidimensional sorts where participants used their most commonly selected dimension was .80.

For Other sorting, there was no statistical effect of perceptual difficulty, $F(1,79) = 0.46$, $p = .498$, $\eta^2_p = .01$ (difference = .03), or exposure type, $F(1,79) = 1.58$, $p = .212$, $\eta^2_p = .02$ (difference = .07), and no statistically significant interaction between exposure type and perceptual difficulty, $F(1,79) < 0.01$, $p = .968$, $\eta^2_p < .001$. According to the self-reports, other sorts were a mixture of failed overall similarity (.33), failed unidimensional (.30), and other idiosyncratic (.36) strategies. Overall, the self-reports indicated that participants attempted the same strategy on .90 of sorts.

Perceptual discrimination test. As for the running-recognition task, we used d’ as our measure of accuracy for the perceptual discrimination task. The mean accuracy for the same-stimuli and unrelated-stimuli conditions is shown in Figure 7a. We conducted a mixed-design three-way ANOVA with the between-subjects factors being level of perceptual difficulty (perceptually easy/perceptually difficult), categorization set (same-stimuli/unrelated-stimuli) and the within-subjects factor being whether the stimulus set had been pre-exposed via running-recognition or if it had not[3]. Mean d' was 0.95 greater in the perceptually easy condition than the perceptually difficult condition, $F(1,79) = 27.00$, $p <.001$, $\eta^2_p = .26$, while accuracy was also 0.39 higher in the same-stimuli condition than the unrelated-stimuli condition, $F (1,79) = 4.53$, $p = .036$, $\eta^2_p = .05$. The 0.07 greater accuracy for the stimuli pre-exposed in running-recognition compared to those not pre-exposed was not statistically significant, $F(1,79) =0.56$, $p = .457$, $\eta^2_p < .01$. However, the interaction between the same-stimuli/unrelated-stimuli factor and the pre-exposure factor was significant, $F(1,79) = 5.79$, $p = .018$, $\eta^2_p = .07$. The remaining interactions were not statistically significant.

We then analyzed the data for the same-stimuli and unrelated-stimuli conditions separately to better characterize the nature of the significant interaction we observed. For the same-stimuli condition, accuracy was 0.29 greater when stimuli had been pre-exposed during the running-recognition task than when they had not been, and this effect was marginally significant, $F(1,40) = 3.94$, $p = .054$, $\eta^2_p = .09$. Mean d' was 1.04 greater in the perceptually easy condition than the perceptually difficult condition, $F(1,40) = 15.34$, $p <.001$, $\eta^2_p = .28$.There was no significant interaction between running-recognition exposure and perceptual difficulty, $F(1,40) = 2.96$, $p = .093$, $\eta^2_p = .07$. We then ran a priori follow-up comparisons for the perceptually easy and perceptually difficult conditions separately as a key reason for including the perceptual discrimination task was to test whether pre-exposure had a differential effect on these groups. For the perceptually easy stimuli, the beneficial effect of pre-exposure was 0.53, $t(20) = 3.09$, $p = .006$, $d = 0.68$, while for the perceptually difficult

[3] We also divided the task into two halves (24 trials in each half) and ran a four-way ANOVA including this additional factor to assess whether there was an impact of learning on the task. This indicated that d' was .24 higher in the second half of the task than the first half, $F (1, 79) = 9.33$, $p = .003$, $\eta^2_p = .11$. There was no statistically significant interaction between session half and any of the other factors, with the exception of a four-way interaction, $F(1,79) = 7.01$, $p = .01$, $\eta^2_p = .08$, between session half, pre-exposure in running recognition, categorization set (same-stimuli/unrelated-stimuli), and perceptual difficulty,

stimuli the .03 advantage of relevant pre-exposure was not statistically significant, t (20) = 0.17, p = .871, d = 0.03.

For the unrelated-stimuli condition, d' accuracy for the perceptually easy stimuli was 0.84 greater than for the perceptually difficult stimuli, F(1,39) = 11.74, p < .001, η^2_p = .23. The stimuli not exposed during running-recognition had 0.15 higher accuracy than the stimuli pre-exposed during running-recognition but this effect was not statistically significant, F(1,39) = 1.90, p = .176, η^2_p = .05. There was also no statistically significant interaction between pre-exposure in running-recognition and perceptual difficulty, F(1,39) = 0.15, p = .698, η^2_p < .01 - there was no statistical difference between the stimuli pre-exposed during running-recognition and stimuli not pre-exposed for either the perceptually easy stimuli, t(20) = -1.27, p = .218, d = 0.28 (d' difference = 0.19), or the perceptually difficult stimuli, t(19) = -0.69, p = .501, d = 0.13 (d' difference = 0.11). This difference in the pattern of results between the same-stimuli and unrelated-stimuli conditions may reflect the fact that in the unrelated-stimuli conditions, participants had previously seen both sets of stimuli (one during running-recognition and the other via categorization) prior to the discrimination test. Contrastingly, in the same-stimuli condition, participants had only viewed one set of stimuli, but in this case over both the running-recognition and categorization phases. As a consequence, the difference in exposure was maximal for the same-stimuli condition, but minimal for the unrelated-stimuli condition. This variation in pre-exposure is one plausible explanation for the significant interaction between categorization set (same-stimuli/unrelated-stimuli) and running-recognition.

As secondary analyses, we examined reaction time (RT) in the same manner as we did for accuracy (see Figure 7b for the descriptive data). In an initial 3-way

ANOVA[4], the perceptually easy stimuli were responded to 2158ms quicker than the perceptually difficult stimuli, $F(1,79) = 24.77$, $p < .001$, $\eta^2_p = .24$, while RT was 214ms quicker for stimuli that had been pre-exposed during running-recognition than stimuli that had not been and this effect was marginally significant, $F(1,79) = 3.72$, $p = .057$, $\eta^2_p = .05$. Participants in the same-stimuli condition responded 458ms slower than participants in the unrelated-stimuli condition but this difference was not statistically significant, $F(1,79) = 1.11$, $p = .295$, $\eta^2_p = .01$. As for the accuracy analysis, there was a significant interaction between running-recognition exposure type and categorization set (same-stimuli/unrelated-stimuli), $F(1,79) = 6.66$, $p = .012$, $\eta^2_p = .08$, although none of the other interactions were statistically significant.

For the same-stimuli condition, perceptually easy stimuli were responded to 2424ms faster than perceptually difficult stimuli, $F(1,40) = 12.75$, $p <.001$, $\eta^2_p = .24$. RT was also 498ms faster for stimuli that had been pre-exposed compared to non-pre-exposed stimuli, $F(1,40) = 7.06$, $p = .011$, $\eta^2_p = .15$. There was no statistically significant interaction between pre-exposure type and perceptual difficulty, $F(1,40) = 0.31$, $p = .578$, $\eta^2_p < .01$.

For the unrelated-stimuli condition, RT was 1893ms quicker for the perceptually easy stimuli than the perceptually difficult stimuli, $F(1,39) = 12.48$, $p = .001$, $\eta^2_p = .24$. However, there was no statistical difference between exposure conditions, $F(1,39) = 0.40$, $p = .533$, $\eta^2_p = .01$ (difference = 72ms) and no statistically significant interaction, $F(1,39) = 2.60$, $p = .115$, $\eta^2_p = .06$.

Discussion

The classification results of Experiment 2b replicated the main findings of Experiment 2a. Specifically, there was a significant interaction, with same-stimuli

[4] As for the accuracy analyses, we also considered the effect of session half together with the other factors in a four-way ANOVA. Participants were 964ms slower in the first half of the task than the second half, $F(1,79) = 33.47$, $p < .001$, $\eta^2_p = .30$. There were no statistically significant interactions between session half and any of the other factors.

exposure significantly increasing the prevalence of overall similarity sorting relative to unrelated pre-exposure for the perceptually easy stimuli but there was no such statistical effect for the perceptually difficult stimuli. We also again observed greater overall similarity sorting for the perceptually easy stimuli than the perceptually difficult stimuli. Furthermore, in the same-stimuli (but not the unrelated-stimuli) conditions we observed a broadly similar pattern in the perceptual discrimination task. Participants who viewed the perceptually easy stimuli responded more quickly and accurately on the stimuli they had viewed during the running-recognition task than the stimuli they had not previously seen but this effect did not emerge for the perceptually difficult stimuli although this interaction did not reach statistical significance.

In summary, Experiment 2b again showed that there was a significant effect of pre-exposure for perceptually easy stimuli but no significant effect for perceptually difficult stimuli. In addition, it suggests that the increased ability to differentiate the perceptually easy stimuli following pre-exposure may be driving the corresponding elevation in overall similarity sorting for these stimuli. For the perceptually difficult stimuli, there was no detectable perceptual learning and also no elevation in overall similarity sorting. This is perhaps because, even after pre-exposure, the differences between the dimensions were sufficiently small to prohibit easy multi-dimensional processing of the stimuli and instead encouraged participants to selectively attend to the dimension they found easiest to differentiate.

Experiment 3

Experiment 3 takes a different angle from Experiments 2a and 2b by examining how the pre-exposure effect is affected by manipulating the level of spatial integration of the stimulus dimensions. Previous work has indicated that stimuli with spatially separate dimensions evoke a greater proportion of overall similarity sorting than stimuli where the dimensions are spatially integrated into a coherent object (e.g., Milton & Wills, 2004, 2009). Milton and Wills (2004) argued that spatially separating out the dimensions made it easier to extract the relevant dimensions and to differentiate the stimuli which would, consequently, make it easier for participants to apply a multidimensional, overall similarity rule. We propose that pre-exposure could work in a similar way to this by making it easier for participants to extract the different stimulus dimensions. This assumption is supported by Goldstone and Steyvers (2001) who created stimuli with arbitrary dimensions by blending photographs of faces in different proportions. Over time, participants learned to identify the relevant dimensions and use this information to increase their categorization accuracy.

If one of the things pre-exposure does is to help identify and process the individual dimensions (all of which in the current experiments are equally relevant), then one would predict that it would have more of an impact for stimuli whose dimensions are spatially integrated where the dimensions are difficult to extract than for spatially separate stimuli where this should be relatively straightforward to do even without pre-exposure. Accordingly, we predicted an interaction between pre-exposure and the level of spatial integration of the stimuli with the effect of pre-exposure on overall similarity sorting being greater for spatially integrated stimuli than for spatially separate stimuli. As in Experiment 2b we included a perceptual discrimination task after the categorization phase to examine whether relevant pre-exposure improved the differentiation of the stimuli.

Method

Participants and Apparatus. Eighty-seven students from the University of Exeter took part in the experiment either for course credits or for £5. Four additional participants were excluded for having incomplete data. Participants were tested in individual cubicles within a multi-testing lab (with up to eight participants tested simultaneously) using E-prime, on a Dell PC with a 17-in. monitor and a standard computer keyboard.

Stimuli. The four stimulus sets were closely based on the perceptually easy stimuli employed in Experiment 2b. For both the lamp and butterfly stimuli one of the sets had the dimensions integrated into a coherent object (spatially integrated), while the other set had the dimensions separated out (spatially separate). The category prototypes for the four sets of stimuli are shown in Figure 8.

Design. The experiment had a 2 x 2 between-subjects factorial design. The first factor was the spatial separateness of the stimulus dimensions (two levels: *spatially separate/spatially integrated*). The second factor was the type of pre-exposure, which also had two levels (*same-stimuli*/*unrelated-stimuli*). This led to four conditions: *spatially integrated/same-stimuli* exposure (24 participants), *spatially separate/same-stimuli* exposure (20 participants), *spatially integrated/unrelated-stimuli* exposure (23 participants) and *spatially separate/unrelated-stimuli* exposure (20 participants). In all conditions, the stimulus set (either lamps or butterflies) that participants classified was randomized.

Procedure. The procedures for the running-recognition and categorization phases were identical to those in Experiments 2a and 2b. The perceptual discrimination task also had the same structure as in Experiment 2b, although here participants were only tested on stimuli of the same level of spatial integration that they had encountered earlier in the experiment.

Results

The raw data for this experiment are publicly available at: https://osf.io/5cd34/

Running-recognition phase. Mean accuracy across conditions is displayed in Supplemental Figure 3. As in previous experiments, accuracy improved across blocks, $F (15, 1260) = 3.32$, $p < .001$, $\eta^2_p = .04$. Mean d' was also 0.20 greater overall in the spatially integrated condition than the spatially separate condition, $F (1, 84) = 7.95$, $p = .006$, $\eta^2_p = .09$. There was a significant interaction between spatial integration and block, $F (15, 1260) = 1.82$, $p = .028$, $\eta^2_p = .02$, with performance rising more sharply across blocks in the spatially integrated than the spatially separate condition.

Categorization. The mean proportion of overall similarity, unidimensional and other categorizations produced in the four conditions are shown in Figure 9. There was a marginally significant effect of the level of spatial integration, $F(1,83) = 3.92$, $p = .051$, $\eta^2_p = .05$, with the proportion of overall similarity sorting .17 higher in the spatially separate condition than the spatially integrated condition. There was no statistical difference between the pre-exposure conditions, $F(1,83) = 1.64$, $p = .203$, $\eta^2_p = .02$ (mean difference = .11), but there was a significant interaction between spatial integration and pre-exposure type, $F(1,83) = 4.33$, $p = .041$, $\eta^2_p = .05$. This interaction reflected that overall similarity sorting was .28 greater in the same-stimuli condition than the unrelated-stimuli condition for the spatially integrated condition, $t(45) = 2.61$, $p = .012$, $d = 0.76$, but there was no statistical difference for the spatially separate stimuli, $t(38) = -0.51$, $p = .610$, $d = 0.16$ (difference = .07). In addition, for the marginally significant main effect of spatial integration we looked at the same-stimuli and unrelated-stimuli groups separately. Overall similarity sorting was .34

higher in the spatially separate condition than the spatially integrated condition under unrelated-stimuli pre-exposure, $t(41) = 2.86$, $p < .001$, $d = 0.87$, but there was no statistical difference between the spatial integration conditions for the same-stimuli exposure condition, $t(42) = 0.07$, $p = .944$, $d = 0.02$ (mean difference = .01).

For unidimensional sorting, the .11 difference between the spatially integrated and spatially separate conditions was not statistically significant, $F(1,83) = 1.76$, $p = .188$, $\eta^2_p = .02$, nor was the .05 difference between the pre-exposure conditions, $F(1,83) = 0.25$, $p = .616$, $\eta^2_p < .001$.There was also no statistically significant interaction between spatial integration and pre-exposure type, $F(1,83) = 2.29$, $p=.134$, $\eta^2_p = .03$. For those participants who had some unidimensional sorts there was again a tendency to consistently use a single dimension (mean proportion = .75).

For Other responding, there was no statistically significant effect of spatial integration, $F(1,83) = 1.16$, $p = .285$, $\eta^2_p = .01$ (difference = .06), or pre-exposure, $F(1,83) = 1.59$, $p = .210$, $\eta^2_p = .02$ (difference = .06), and no statistically significant interaction between spatial integration and pre-exposure type, $F(1,83) = 0.89$, $p = .49$, $\eta^2_p = .01$. Other sorts, according to classification of the self-reports, were again a mixture of failed overall similarity (.34), failed unidimensional (.26), and other idiosyncratic (.41) strategies. According to the self-reports, participants attempted the same strategy on .86 of classifications.

Perceptual discrimination test. The mean accuracy (d') in the perceptual discrimination test for the same-stimuli and unrelated-stimuli conditions is shown in Figure 10a. Analyses were run in the same way as in Experiment 2b.[5] The mean d' of stimuli that had been exposed during the running-recognition phase was 0.30 higher than for stimuli that had not been pre-exposed during running-recognition, F(1,83) =

[5] As before, we also ran a four-way ANOVA, splitting the task into two halves as an additional factor to ascertain whether there were any learning effects. Mean d' was 0.03 higher in the first half than the second half but this was not statistically significant, $F(1, 83) = .16$, $p = .694$. There were also no statistically significant interactions between session half and any of the other three factors.

11.70, $p < .001$, $\eta^2_p = .12$. There was no statistical difference between the same-stimuli and unrelated-stimuli conditions, $F(1,83) = 0.10$, $p = .757$, $\eta^2_p = .001$ (mean difference = 0.05). The 0.24 greater accuracy for the spatially separate stimuli than the spatially integrated stimuli was also not statistically significant, $F(1,83) = 1.81$, $p = .182$, $\eta^2_p = .02$. There was likewise no statistically significant interaction between the running-recognition factor and level of spatial integration, $F(1,83) = 1.51$, $p = .222$, $\eta^2_p = .02$, but there was a significant interaction between the running-recognition factor and categorization set (same-stimuli/unrelated-stimuli), $F(1,83) = 13.24$, $p < .001$, $\eta^2_p = .14$, and a three-way interaction between running-recognition exposure, categorization set and spatial integration, $F(1,83) = 6.07$, $p = .016$, $\eta^2_p = .07$.

As before, we then analyzed the results for the same-stimuli and unrelated-stimuli conditions separately to characterize the nature of this factor's interaction with pre-exposure. For the same-stimuli conditions, mean d' was 0.62 higher for stimuli that had been pre-exposed during running-recognition than stimuli which had not been pre-exposed, $F(1,42) = 21.78$, $p < .001$, $\eta^2_p = .34$. While accuracy was 0.31 higher for the spatially separate stimuli than the spatially integrated stimuli, this effect was not statistically significant, $F(1,42) = 0.96$. $p = .332$, $\eta^2_p = .02$. However, there was a significant interaction between level of spatial integration and running-recognition pre-exposure, $F(1,42) = 5.964$, $p = .019$, $\eta^2_p = .12$. Investigating this interaction further, for the spatially integrated stimuli, accuracy was 0.94 higher for the stimuli pre-exposed during running recognition than the stimuli that were not pre-exposed, $t(23) = 5.39$, $p <.001$, $d = 1.10$, while the 0.29 difference between pre-exposure conditions for the spatially separate stimuli was not statistically significant, $t(19) = 1.47$, $p = .158$, $d = 0.32$.

For the unrelated-stimuli condition, the d' difference of 0.02 between the stimuli pre-exposed during running-recognition and the stimuli that were not pre-exposed was not statistically significant, $F(1,41) = 0.03$, $p = .867$, $\eta^2_p = .001$. There was also no statistical difference between the spatially integrated and spatially separate stimuli, $F(1,41) = 0.87$, $p = .357$, $\eta^2_p = .02$ (difference = 0.21), and no statistically significant interaction between spatial integration and running-recognition exposure, $F(1,41) = 0.90$, $p = .349$, $\eta^2_p < .001$.

The mean RT across conditions are shown in Figure 10b. The 78ms difference between the spatially integrated and spatially separate stimuli was not statistically significant, $F(1,83) = 0.16$, $p = .687$, $\eta^2_p = .002$. There was also no statistical difference between the pre-exposure conditions, $F(1,83) = 2.32$, $p = .131$, $\eta^2_p = .03$ (difference = 97ms) or between the same-stimuli and unrelated-stimuli categorization conditions, $F(1, 83) = 0.49$, $p = .490$, $\eta^2_p = .01$ (difference = 134ms). There was, however, a significant interaction between type of pre-exposure and the same/unrelated stimuli factor, $F(1,83) = 13.95$, $p < .001$, $\eta^2_p = .14$, although none of the other interactions were statistically significant.[6]

For the same-stimuli condition, RT was 333ms quicker for stimuli that had been pre-exposed during running-recognition compared to stimuli that had not been, $F(1,42) = 24.76$, $p < .001$, $\eta^2_p = .37$, but there was no statistical difference between the spatially integrated and spatially separate stimuli, $F(1,42) = 0.85$, $p = .362$, $\eta^2_p = .02$ (difference = 298ms). There was also no statistically significant interaction between pre-exposure and spatial integration, $F(1,42) = 0.75$, $p = .391$, $\eta^2_p = .02$.

For the unrelated-stimuli condition, the difference of 139ms between the stimuli that had been pre-exposed and the stimuli that had not been was not

[6] We again ran an additional ANOVA also including session half as an extra factor. Participants were 516ms slower in the first half of the task than the second half, $F(1, 83) = 63.97$, $p < .001$, $\eta^2_p = .44$. There were no statistically significant interactions between session half and any of the other factors.

statistically significant, $F(1, 41) = 1.67$, $p = .203$, $\eta^2_p = .04$. There was also no significant effect of spatial integration, $F(1, 41) = 0.48$, $p = .49$, $\eta^2_p = .01$ (mean difference = 143ms), and no statistically significant interaction between pre-exposure type and spatial integration, $F(1,41) = 0.20$, $p = .656$, $\eta^2_p = .01$.

Taken together, as in Experiment 2b, we found a notable difference in the pattern of results for the same-stimuli and unrelated-stimuli conditions, with the effect of pre-exposure much more marked for the same-stimuli condition. This plausibly reflects the fact that participants in the unrelated-stimuli condition had, by necessity, pre-exposure to both sets of stimuli, one in the running-recognition pre-exposure phase and the other set in the categorization phase. This gave participants the opportunity to learn about the dimensions of both stimulus sets prior to the perceptual discrimination task. Given this, one might expect that in the unrelated-stimuli condition, participants who produced a higher level of overall similarity sorting (and consequently demonstrating awareness of multiple dimensions) would perform better in the perceptual discrimination task which requires use of all the dimensions for optimal accuracy than participants who sorted by a single dimension.

To investigate this possibility, we first collapsed across the perceptual discrimination tasks in the unrelated-stimuli conditions of Experiments 2b and 3 to increase power and then correlated the proportion of overall similarity sorts that participants produced against their d’ accuracy for the stimuli they had not viewed during running-recognition. This revealed a significant positive correlation, $r(85) = 0.35$, $p = .001$ (looking at the experiments individually the results were: Experiment 2b, $r(41) = 0.25$, $p = .115$; Experiment 3, $r(43) = 0.29$, $p = .064$).

On the other hand, a similar pattern emerged for the stimuli that the unrelated-stimuli condition participants had been pre-exposed to during running-recognition, r (84) = .30, p = .006 (Experiment 2b, r (41) = .17, p = .285; Experiment 3, r (43) = .22, p = .161). The same pattern emerged as well as in the same-stimuli condition for the items that had been pre-exposed in running recognition, r(86) = .52, p < .001 (Experiment 2b, r (42) = .56, p < .001; Experiment 3, r (44) = .38, p = .01) and, perhaps most intriguingly, in the same-stimuli condition for stimuli which had not been pre-exposed during running-recognition, r (86) = .43, p < .001 (Experiment 2b, r (42) = .45, p = .002; Experiment 3, r (44) = .37, p = .013). This pattern of results is of interest in itself as it suggests that participants who utilize more of the dimensions during categorization are likely to perform better in a subsequent perceptual discrimination task that requires use of all the dimensions for optimal accuracy. Furthermore, the results of the same-stimuli/ not-pre-exposed condition suggest that participants who categorized by overall similarity in the categorization phase can transfer a multidimensional approach to stimuli in the perceptual discrimination task that they have not encountered before (for a somewhat analogous result, see Milton & Wills, 2009). However, this pattern does not provide much insight into why there is a difference in perceptual discrimination accuracy between the same-stimuli and unrelated-stimuli conditions. As such, the precise reason for the difference must remain a matter for speculation, but on the basis of the two experiments we ran the result appears robust.

Discussion

Consistent with our predictions, we found a significant interaction between pre-exposure and the level of spatial integration of the stimulus dimensions. As in Experiments 1 and 2, when the stimuli were spatially integrated and the within-dimension differences relatively easy to distinguish we observed a significant elevation of overall similarity sorting in participants who had been pre-exposed to the stimuli compared to those who had been pre-exposed to a different set of stimuli. In contrast, when the dimensions were spatially separate we did not find a significant effect of pre-exposure. This pattern of findings is consistent with the idea that pre-exposure enables participants to identify and more easily process the relevant dimensions which is necessary for participants to sort by overall similarity. When the dimensions are spatially separate, they are relatively easy to differentiate even without prior experience and the benefit of pre-exposure is consequently attenuated.

This explanation is supported by the results of the perceptual discrimination test which was conducted after the categorization phase to directly assess whether perceptual learning had occurred. The results, for the same-stimuli but, as in Experiment 2b, not the unrelated-stimuli condition, mirrored the pattern observed in the classification task. For the spatially separate stimuli, there was no difference in discrimination accuracy between stimuli which had been pre-exposed via running-recognition and stimuli which had not been pre-exposed. In contrast, for the spatially integrated stimuli, perceptual discrimination was better and RT was quicker for stimuli which had been viewed during pre-exposure than stimuli which had not been. This pattern of results, therefore, is consistent with the idea that the increased ability to differentiate the spatially integrated stimuli after pre-exposure is driving the corresponding elevation in overall similarity sorting and this leads to an elimination of the spatial integration effect described by Milton and Wills (2004) where spatially separate stimuli evoke more overall similarity responding than spatially integrated stimuli.

General Discussion

This paper presents a series of experiments which provide the first evidence that prior exposure to perceptual stimuli can increase the prevalence of overall similarity categorization. This effect was found for perceptually easy stimuli but not for perceptually difficult stimuli (Experiments 2a and 2b) and was mirrored in Experiment 2b by the findings of a perceptual discrimination task which revealed that pre-exposure enhanced the ability of participants to differentiate between the perceptually easy stimuli but not the perceptually difficult stimuli (although this only emerged for the same-stimuli condition but not the unrelated-stimuli condition). Finally, in Experiment 3 we found that pre-exposure significantly elevated overall similarity categorization for spatially integrated stimuli but not for spatially separate stimuli with pre-exposure similarly improving differentiation for the spatially integrated stimuli but not for the spatially separate stimuli (again, though, this effect only emerged for the same-stimuli condition and not the unrelated-stimuli condition).

An obvious first thing to consider is the process by which pre-exposure might lead to an increased likelihood of overall similarity categorization. According to Combination theory (Wills et al., 2015; see also Milton & Wills, 2004), overall similarity sorting is the result of an effortful combination of information from the various stimulus dimensions. Participants process all the stimulus dimensions individually and use a rule to place the item into the category with which it has the most features in common. One necessary pre-requisite for overall similarity sorting of this kind is the ability to identify the individual dimensions so that the information from them can then be combined. Furthermore, the quicker that participants can process the individual dimensions and differentiate the feature values, the easier such a strategy would be to perform, which should lead to an elevation of overall similarity sorting. A direct prediction of Combination theory, therefore, is that an increased ability to differentiate stimuli as a result of pre-exposure should make an overall similarity strategy easier and, consequently, a more commonly applied strategy.

An alternative process by which overall similarity sorting has been thought to occur is what we have previously termed Differentiation theory (Wills et al., 2015; see also J.D. Smith & Kemler Nelson, 1984; Ward, 1983). This account posits that stimuli are first processed as an undifferentiated whole and only later and with effort can they be broken down into their constituent parts. This account therefore assumes that overall similarity sorting should be a quick, more effortless, process than unidimensional sorting. According to this account, pre-exposure should make it easier for participants to process dimensions at an individual level which should consequently make unidimensional sorting easier. The increased overall similarity sorting and reduced unidimensional sorting that we observed as a result of pre-exposure appears therefore to be more consistent with Combination theory than Differentiation theory.

One unexpected finding was that pre-exposure increased overall similarity sorting for the perceptually easy stimuli but not for the perceptually difficult stimuli, which was the reverse pattern to that we predicted a priori. Nevertheless, it is consistent with the key tenet of Combination theory that a greater ability to differentiate the stimulus dimensions will lead to increased overall similarity categorization. The perceptually easy stimuli are relatively straightforward to distinguish which makes overall similarity sorting a relatively easy and quick strategy to perform. Pre-exposure further increases the differences between the stimuli which leads to an additional elevation of overall similarity sorting. With the perceptually difficult stimuli, it is a difficult and time-consuming process to differentiate the feature-values for all the dimensions which makes an overall similarity strategy very effortful and time consuming to conduct. Instead, it is easier and quicker for participants to base their categorizations on a single dimension. We suspect, in hindsight, that the reason we failed to detect any effect for the perceptually difficult stimuli was because they were so similar. This meant that the level of exposure provided was insufficient for perceptual learning to be detectable because even after pre-exposure the stimuli are likely to still look very similar with the differences being difficult and time-consuming to detect. Instead, it seems plausible during the categorization phase that participants in the perceptually difficult condition participants selectively attended to the dimension that they found most salient and paid less attention to the rest of the stimulus where variations were less easy to detect and this would also be the case in the perceptual discrimination task. In this regard, it appears that a necessary condition for pre-exposure to influence category behavior is for the stimuli to be conducive to overall similarity sorting – the extremely low levels of overall similarity sorting for the perceptually difficult stimuli in both Experiments 2a and 2b suggests that this was not the case for these stimulus sets.

One possibility is that with more extended pre-exposure the perceptually difficult stimuli would become much easier to differentiate, which might then potentially facilitate overall similarity sorting. On the other hand, it is possible that with this additional pre-exposure participants will simply find it easier to differentiate the subset of the dimensions they initially focused on, and that this would encourage them to persist with only using these dimensions, rather than learning more about the other dimensions. Future work distinguishing between these two possibilities would be of value.

One further important question is why our perceptual difficulty result appears inconsistent with previous work that indicates that the effect of pre-exposure is most pronounced for perceptually difficult stimuli (e.g., Oswalt, 1972). Indeed, the rationale for making the perceptually difficult stimuli so similar was based on past work and theories of perceptual learning. One explanation is that the multi-dimensional stimuli and the small differences in the feature-values that we used in this experiment are unlike the types of stimuli that have previously been used to investigate the relationship between pre-exposure and perceptual difficulty. For example, in the classic study by Oswalt (1972), rats were trained on either easily discriminable stimuli (horizontal and vertical striations) or difficult to distinguish stimuli (circles and triangles). Similarly, Chamizo and Mackintosh (1989; see also Trobalon et al., 1991) trained rats in intra-maze and extra-maze discriminations and found that pre-exposure aided learning when the cues shared many common features but when the differences between the cues were increased (e.g., by painting the walls of the arms in the maze black or white) pre-exposure retarded learning. We suspect that the effect of perceptual difficulty on pre-exposure may be more complex than is typically recognized and perhaps follows a non-linear function. The present work indicates that it is difficult to observe perceptual learning for very hard discriminations (at least with the amount of pre-exposure we provided) but equally it appears likely that at the other extreme where the feature-values are extremely different, perceptual learning would also be negligible as the discriminations would be easy even without pre-exposure. It would, therefore, be useful in future work to more systematically characterize the relationship between perceptual difficulty and pre-exposure than has previously been done.

The finding that pre-exposure increased overall similarity categorization for spatially integrated stimuli but not for the spatially separate stimuli was consistent with our initial hypothesis. Spatially separating the dimensions has previously been found to increase overall similarity sorting and, consistent with Combination theory, it has been argued that this is because it makes it easier to identify and process the dimensions of variation (Milton & Wills, 2004). We predicted that pre-exposure may act in a similar way by increasing between-dimension discriminability. According to this account, pre-exposure has a greater impact for spatially integrated stimuli than for spatially separate stimuli where the dimensions should be relatively easy to identify even without pre-exposure (in a similar manner to what we propose would happen with extremely easy perceptual discriminations). Consistent with this idea, we found greater overall similarity sorting for the spatially separate stimuli than the spatially integrated stimuli when they had not been pre-exposed but this effect was not present when the stimuli had been pre-exposed. This explanation also receives direct support from the results of the perceptual discrimination test where pre-exposure significantly increased the differentiation of the spatially integrated stimuli but not the spatially separate stimuli.

Is it possible that the only thing that pre-exposure does is enable participants to identify better the dimensions which are varying? For example, if participants have only identified a single dimension of variation within the stimuli then it would be impossible to categorize them by overall similarity. While we believe that pre-exposure does facilitate this, and the results of Experiment 3 in particular are supportive of this, we think that it is unlikely that this is the only thing that is occurring as a result of pre-exposure. One reason for this is that previous work has shown that being able to identify all the dimensions (via a matching-pairs task described in the introduction) did not, on its own, yield an increase in overall similarity categorization (Milton & Wills, 2004). Instead, we suspect that improving the identification of dimensions is just one of potentially several processes resulting from pre-exposure that is elevating overall similarity categorization.

For example, it has been extensively documented that pre-exposure leads to an enhanced ability to make within-dimension discriminations (e.g., Gibson & Walk, 1956; for reviews see Goldstone, 1998; McLaren & Mackintosh, 2000; see our summary of the MKM model in the introduction as one theory that explains this result). This would be consistent with the perceptual difficulty effect we identified in Experiments 2a and 2b where stimuli with greater differences in the feature-values had increased levels of overall similarity sorting compared to stimuli with little difference in the feature values. Pre-exposure may act in a similar way to the perceptual difficulty manipulation we applied in Experiment 2 – both are likely to enhance within-dimension discriminations, which should as one consequence aid between-dimension discriminations, leading to a facilitation of overall similarity sorting according to Combination theory.

Additionally, pre-exposure may aid participants in identifying the inter-correlation of features that is present in the category structures used in the present experiments (see Lassaline & Murphy, 1996 for a related illustration of this). For example, for the structure employed in Experiments 2 and 3 (shown in Table 2), knowing the value on Dimension 1, allows one to predict the value on any other dimension with 67% accuracy. In this regard, pre-exposure may serve a similar purpose to relevant background knowledge which has been taken to encourage overall similarity sorting because it increases the salience of inter-dimension relationships that are otherwise difficult to discover (e.g., Lassaline & Murphy, 1994; Spalding & Murphy, 1996). An alternative, but perhaps complementary, effect is that the inter-correlation of features could itself enhance perceptual learning. For example, one assumption of the MKM model (McLaren et al, 1989) is that the effects of pre-exposure will be greater when features reliably co-occur than when they do not. This is because a greater level of feature inter-correlation would enhance the contribution of salience modulation that is assumed to underlie perceptual learning.

A related possibility is that pre-exposure may lead to unitization, a process whereby individual dimensions or units can be bound into a single perceptual configuration (e.g., Goldstone, 2000; Schyns & Rodet, 1998; Welham & Wills, 2011). If items are perceived holistically then this would likely encourage overall similarity categorization as this should be less effortful than breaking down the holistic object into its constituent parts as is likely needed for unidimensional categorization. It seems unlikely that unitization is driving the current pattern of results as informal inspection of the reaction times (RT) in all experiments (the sample size was too small in some cells to effectively run formal analyses) indicated that in both the same-stimuli and unrelated-stimuli conditions overall similarity responding took longer than unidimensional sorting (see Section C of the Supplemental Materials for the descriptive data). If unitization had occurred, one might expect that the RT would be quicker in overall similarity sorting than unidimensional sorting under same-stimuli conditions. One important caveat is that the stimuli we have used here are quite different from those that have been used in previous demonstrations of unitization (e.g., the blob stimuli of Schyns & Rodet,

1998) which may make them less conducive to such a process. Nevertheless, the precise conditions under which unitization occurs are still not well understood so it remains plausible that it could emerge in this context with more extended pre-exposure and/or with certain types of stimuli.

As noted earlier, there has been debate about whether overall similarity sorting is best characterized as the result of a time-consuming, effortful, deliberative, process (Combination theory e.g., Milton et al., 2008; Wills et al., 2015) or as the result of a quick, holistic, non-deliberative process (Differentiation theory e.g., Smith & Kemler Nelson, 1984; Ward, 1983; Ward, Foley, & Cole, 1986). One intriguing possibility is that there may be a transition from a deliberative to a non-deliberative approach with increasing exposure to stimuli via a process such as unitization. While speculative, this would be an interesting question for future research (see Milton et al., 2008 and Wills et al., 2015 for examples of how this could be done).

A notable aspect of our experiments is that they all used a running-recognition task to pre-expose the stimuli. One reason for using an active pre-exposure task like this was to encourage participants to process the stimuli which arguably may be more likely than if participants had just been asked to passively view the stimuli. Nevertheless, the running-recognition task should not be considered a canonical or neutral method for pre-exposing stimuli and there are many other ways in which it could have been done effectively. For one thing, the current task loads more heavily on declarative memory processes than other tasks one might use. (such as a pleasantness rating, counting the number of dimensions present, passive viewing). It also seems likely that different pre-exposure tasks will encourage attention to all of the stimulus dimensions to a greater or lesser extent.

An important implication of this is that some pre-exposure tasks may potentially elevate overall similarity categorization more than others. One study that illustrates this was conducted by Lassaline and Murphy (1996) who found that a pre-exposure task where participants had to count the number of dimensions led to less overall similarity sorting than a task where participants had to make inductive inferences during the pre-exposure phase. Clearly, then, an important goal for future research would be to understand better the impact that different types of pre-exposure can have on overall similarity sorting. This is of interest in itself but is likely to also provide insight into the mechanisms by which the pre-exposure effect we have observed here operates.

In summary, one of the most notable findings in unsupervised categorization research is that people have a strong tendency to form single-dimension categories and will rarely spontaneously group items according to overall similarity (e.g., Ahn & Medin, 1992; Medin et al., 1987). In recent years, there has been a growing appreciation that certain manipulations such as stimulus presentation method (Regehr & Brooks, 1995), the spatial separateness of the stimulus dimensions (e.g., Milton & Wills, 2004), instructional manipulations (Wills et al., 2013), and background knowledge (e.g., Spalding & Murphy, 1996) can increase overall similarity sorting. The present study provides evidence for a further two manipulations that can be added to this list. The first of these is that stimuli whose dimensions are easy to discriminate lead to more overall similarity sorting than stimuli where the differences are difficult to discriminate. While this in many ways seems intuitive, it does underscore the fact that in many related studies the categories used are arguably more similar to each other than is typically the case outside the lab (unless one is dealing with subordinate categories which non-experts can often have difficulty acquiring). Second, while we have identified boundary conditions for the effect of pre-exposure (i.e., it does not appear to be present for spatially separate stimuli and for stimuli where the stimuli are extremely similar to each other), we have provided clear evidence that relevant pre-exposure can significantly elevate overall similarity sorting in perceptual stimuli. Again, this effect, while previously not documented in the literature, appears relatively intuitive. Outside the lab we typically have a great deal of exposure to the stimuli we are required to categorize. Given this, it is surprising that nearly all extant studies have provided participants with little or no pre-exposure before asking them to classify the stimuli in a meaningful way. Taken together, our results provide new insight into the relationship between the characteristics of the stimuli and perceptual learning and identify two new factors that facilitate overall similarity sorting.

References

Ahn, W.-K., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16*, 81-121.

Ashby, F.G., Queller, S., & Berretty, P.M. (1999) On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178–1199.

Bersted, C. T., Brown, B. R., & Evans, S. H. (1969). Free sorting with stimuli in a multidimensional attribute space. *Perception & Psychophysics, 6B,* 409–413.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking.* New York: Wiley.

Chamizo, V. D., & Mackintosh, N. J. (1989). Latent learning and latent inhibition in maze discriminations. *Quarterly Journal of Experimental Psychology B, 41***,** 21–31.

Gibson, E. J., & Walk, R. D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative & Physiological Psychology*, 49, 239-242.

Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology*, *49*, 585-612.

Goldstone, R.L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 86-112.

Goldstone, R.L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*, 116-139.

Hall, G. (1991). *Perceptual and associative learning.* Oxford, England: Clarendon Press/Oxford University Press.

Hoffman, A.B & Murphy, G.L. (2006). Category dimensionality and feature knowledge: When more features are learned as easily as fewer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 301-315.

Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology, 69*, 596–608.

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

Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic Bulletin & Review, 3*, 95–99.

Lavis Y., Mitchell C. (2006). Effects of pre-exposure on stimulus discrimination: An investigation of the mechanisms responsible for human perceptual learning. *Quarterly Journal of Experimental Psychology, 59*, 2083-2101.

Livesey, E.J., & McLaren, I.P.L. (2011). An elemental model of associative learning and memory. In E.M. Pothos and A.J. Wills (Eds.). Formal approaches in categorization. Cambridge University Press.

Love, B.C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review, 9*, 829-835.

McLaren, I.P.L. (1997). Categorization and perceptual learning: an analogue of the face inversion effect. *Quarterly Journal of Experimental Psychology B, 50,* 257-273.

McLaren, I.P.L., Wills, A.J., & Graham, S. (2010). Attention and Perceptual Learning. In C. Mitchell, & M. Le Pelley (Eds.). *Attention and associative learning: From brain to behaviour*. Oxford: Oxford University Press.

McLaren, I.P.L., Kaye, H., & Mackintosh, N.J. (1989). An associative theory of the representation of stimuli: applications to perceptual learning and latent inhibition. In Morris R.G.M. (ed) *Parallel Distributed Processing - Implications for Psychology and Neurobiology*, Oxford: Oxford University Press.

McLaren, I., Leevers, H., & Mackintosh, N. (1994). Recognition, categorization, and perceptual-learning (or, how learning to classify things together helps one to tell them apart). *Attention and Performance XV, 15*, 889-909.

McLaren, I.P.L., & Mackintosh, N.J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning and Behavior, 28,* 211-246.

McLaren, I.P.L., & Mackintosh, N.J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior, 30*, 177-200.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238.

Medin, D.L., Wattenmaker, W.D. & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.

Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 676 – 692.

Milton, F., & Wills, A.J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *30*, 407-415.

Milton, F., & Wills, A. J. (2009). Long-term persistence of sort strategy in free classification. *Acta Psychologia, 130*, 161-167.

Milton, F., Wills, A. J., & Hodgson, T. L. (2009). The neural basis of overall similarity and single-dimension sorting. *Neuroimage*, *46*, 319-26.

Mundy, M. E., Honey, R.C., & Dwyer, D.M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of Experimental Psychology: Animal Behavior Processes*, *33*, 124-138.

Murphy, G.L. (2002). *The big book of concepts.* Cambridge: Massachusetts.

Oswalt, R. M. (1972). Relationship between level of visual pattern difficulty during rearing and subsequent discrimination in rats. *Journal of Comparative & Physiological Psychology, 81***,** 122–125.

Pothos, E.M., & Close, J., (2008). One or two dimensions in spontaneous classification: a simplicity approach. *Cognition* 107, 581–602.

Regehr, G. & Brooks, L.R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 347-363.

Rosch, E. & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

Schyns, P.G., Goldstone, R.L., & Thibaut, J-P. (1998). The development of features in object concepts. *Behavior and Brain Sciences, 21,* 1-54.

Seitz, A. & Watanabe, T. (2005). A unified model for perceptual learning. *Trends in Cognitive Sciences, 9,* 329-334.

Shepard, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. *Psychological Monographs, 75,* Whole No. 517.

Smith, J. D., & Kemler Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General, 113,* 137 – 159.

Spalding, T. L., & Murphy, G. L. (1996). Effect of background knowledge on category construction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 525–538.

Suret, M.B., & McLaren, I.P.L. (2003). Representation and discrimination on an artificial dimension. *Quarterly Journal of Experimental Psychology, 56B*, 30-42.

Trobalon, J.B., Sansa, J., Chamizo, V.D., & Mackintosh, N.J. (1991). Perceptual learning in maze discriminations. *Quarterly Journal of Experimental Psychology, 43*, 389-402.

Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 103 – 112.

Ward, T. B., Foley, C. M., & Cole, J. (1986). Classifying multidimensional stimuli: Stimulus, task, and observer factors. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 211 – 225.

Welham, A.K., & Wills, A.J. (2011). Unitization, similarity, and overt attention in categorization and exposure. Memory & Cognition, 39, 1518-1533.

Wills, A.J., & McLaren, I.P.L. (1998). Perceptual learning and free classification. *Quarterly Journal of Experimental Psychology B, 51*, 235-270.

Wills, A.J., Milton, F., & Inkster, A. (2015) Combination or differentiation? Two theories of processing order in categorization. *Cognitive Psychology, 80*, 1-33.

Wills, A.J., Milton, F., Longmore, C.A., Hester, S., & Robinson, J. (2013). Is overall similarity classification less effortful than single-dimension classification? *Quarterly Journal of Experimental Psychology, 66*, 299-318.

Appendix

Instructions for the running-recognition phase

Thank you for agreeing to take part in this study. You will see a number of pictures. Your task is to decide whether you have seen an identical version of that picture previously. If you have previously seen the picture within the current block, press X. If you have not seen the picture previously within the current block, press M. There will be 16 blocks in total, with 24 trials per block. You must treat each block independently from the others. In other words only say that you have seen that picture if you have seen it previously in that block. Please ask the experimenter if you have any questions. Press the spacebar to continue.

Instructions for the categorization phase

Please read the following instructions carefully. Two pictures will be displayed at the top of the screen. One of these (on the left) will be characteristic of category A and the other (on the right) will be characteristic of category B. These two characteristic pictures will be present throughout the experiment. Directly under these two characteristic pictures, another will be presented. Your task is to put this lower picture into either Category A (by pressing X) or into category B (by pressing M). There are many ways in which these pictures can be split and there is no correct answer. We are just interested in what you think is the most appropriate way to sort these pictures. There is no time limit and you are encouraged to take as much time as you need to complete the task. In total, there will be 12 pictures to categorize in a 'block' and there will be 6 blocks in total. There will be an opportunity to rest at the end of each block, if you so wish. If you have any questions, please ask the experimenter before you start the task. Please press the spacebar to continue.

Instructions for the categorization response booklet

For block (x), please note down how you categorized the stimuli in as much detail as possible.

Instructions for the perceptual discrimination task (Experiments 2b and 3 only)

In the final part if this study, there will be 48 trials. On each trial you will see two pictures. Your task is to say whether the pictures are identical or whether they differ in some way. If you think they are identical please press x. If you think that they differ please press m. You will be provided feedback about whether you are correct at the end of each trial. Please ask the experimenter if you have any questions and then press the space bar when you are ready to continue.

Acknowledgements

Correspondence should be addressed to Fraser Milton, Discipline of Psychology, University of Exeter, Exeter, EX4 4QG. E-mail: f.n.milton@ex.ac.uk. Experiment 2a was presented in the Proceedings of the 36th Annual Conference of the Cognitive Science Society, and was a development of an earlier study presented at the 30th Annual Conference. We think the Experimental Psychology Society for their support in this research. We also thank Paula McDonald and Tobias Stevens for their help in collecting the data and Ciro Civile for his help in analyzing the data.

Table 1

The Abstract Stimulus Set Used in Experiment 1.

Category A					Category B				
	D1	D2	D3	D4		D1	D2	D3	D4
E1	1	1	1	1	E6	0	0	0	0
E2	1	1	1	0	E7	0	0	0	1
E3	1	1	0	1	E8	0	0	1	0
E4	1	0	1	1	E9	0	1	0	0
E5	0	1	1	1	E10	1	0	0	0

Note. Each row (within each category) describes a different stimulus. D = dimension: 1 and 0 represent the values of each dimension.

Table 2

The Abstract Stimulus Set Used in Experiments 2 and 3.

Category A						Category B				
D1	D2	D3	D4	D5		D1	D2	D3	D4	D5
1	1	1	1	1	E7	0	0	0	0	0
1	1	1	1	0	E8	0	0	0	0	1
1	1	1	0	1	E9	0	0	0	1	0
1	1	0	1	1	E10	0	0	1	0	0
1	0	1	1	1	E11	0	1	0	0	0
0	1	1	1	1	E12	1	0	0	0	0

Note. Each row (within each category) describes a different stimulus. D = dimension: 1 and 0 represent the values of each dimension.

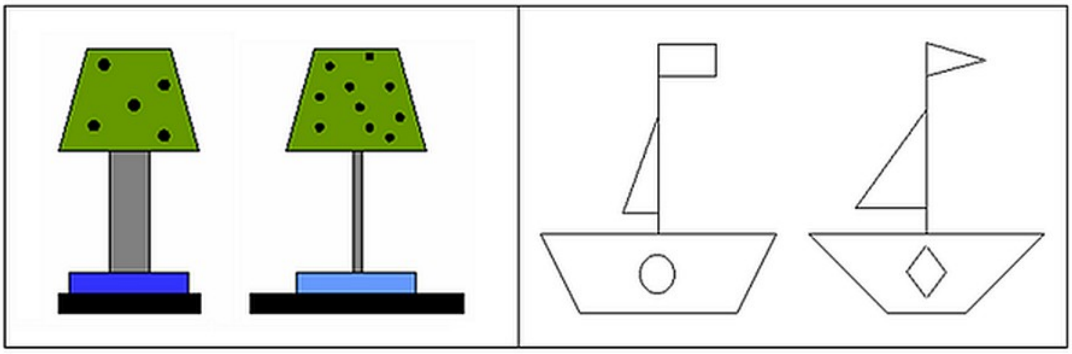

Figure 1. The category prototypes for the two stimulus sets used in Experiment 1.

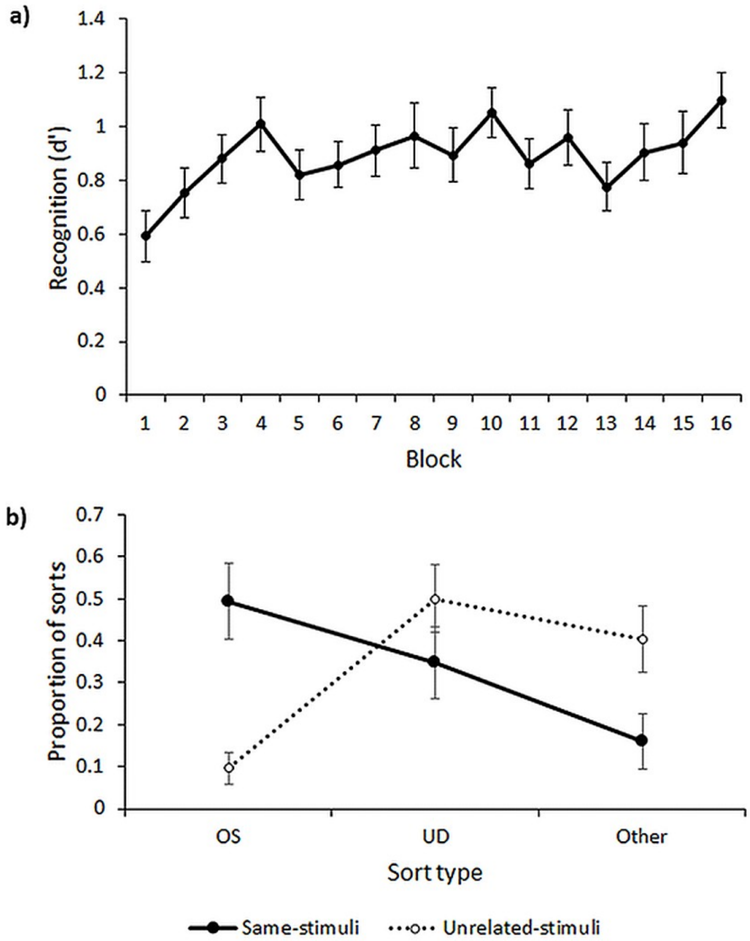

Figure 2. a) Mean accuracy across blocks in the running-recognition phase for both the same-stimuli and unrelated-stimuli conditions; b) The mean proportion of overall similarity (OS), unidimensional (UD) and other sorts for the same stimuli and unrelated-stimuli conditions in Experiment 1. Error bars represent between-subjects variability and were calculated as +/- 1 standard error of the mean.

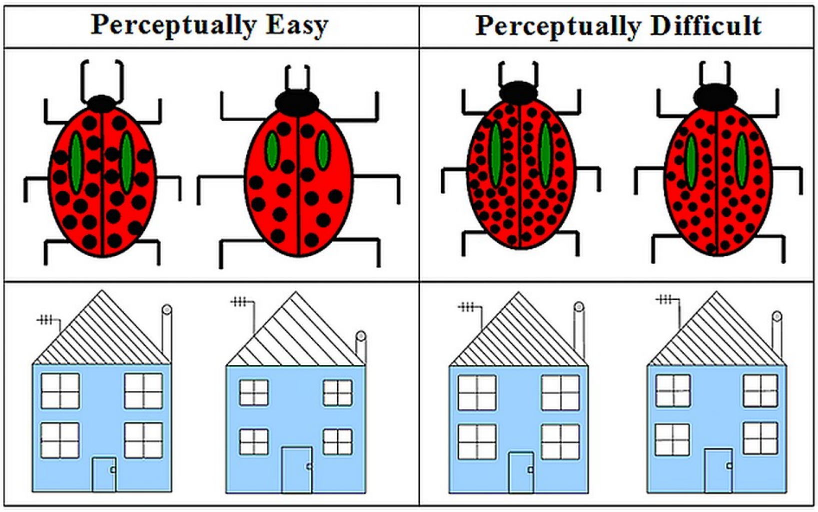

Figure 3. The category prototypes for the stimulus sets used in Experiment 2a.

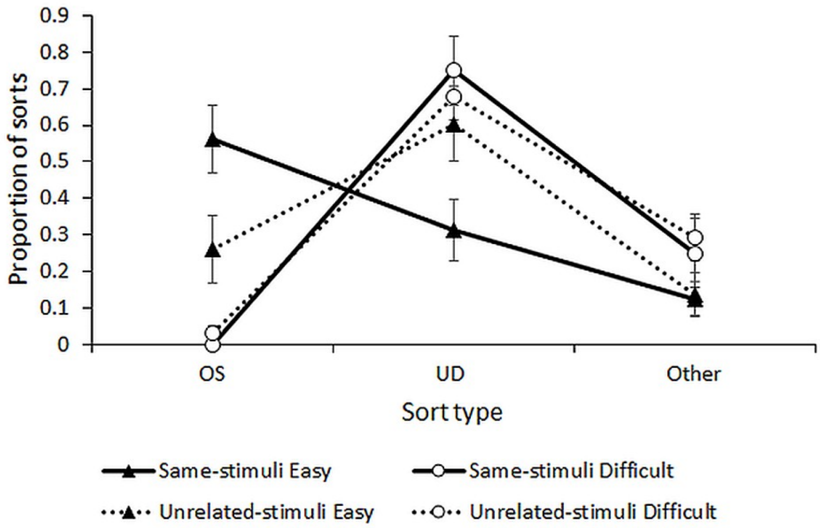

Figure 4. Mean proportion of overall similarity (OS) unidimensional (UD) and Other sorts for all conditions in Experiment 2a. Error bars represent between-subjects variability and were calculated as +/- 1 standard error of the mean.

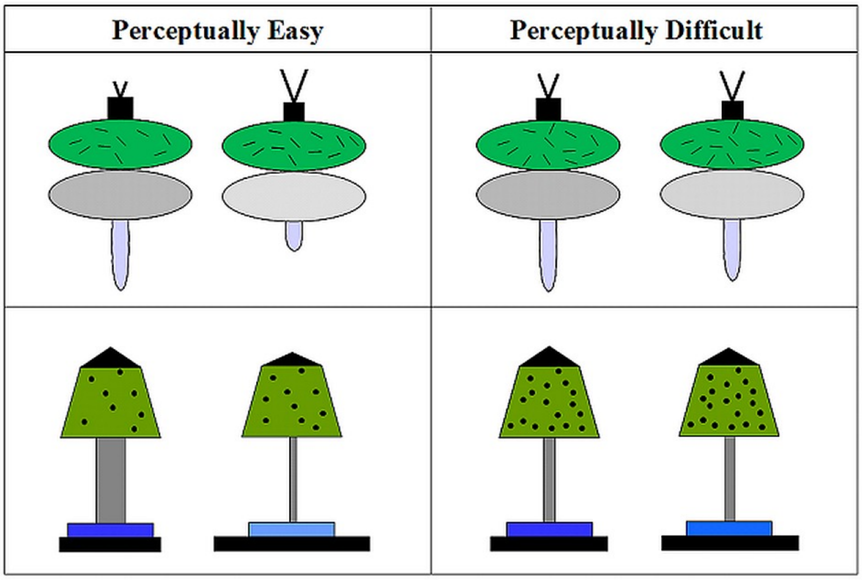

Figure 5. The category prototypes for the stimulus sets used in Experiment 2b.

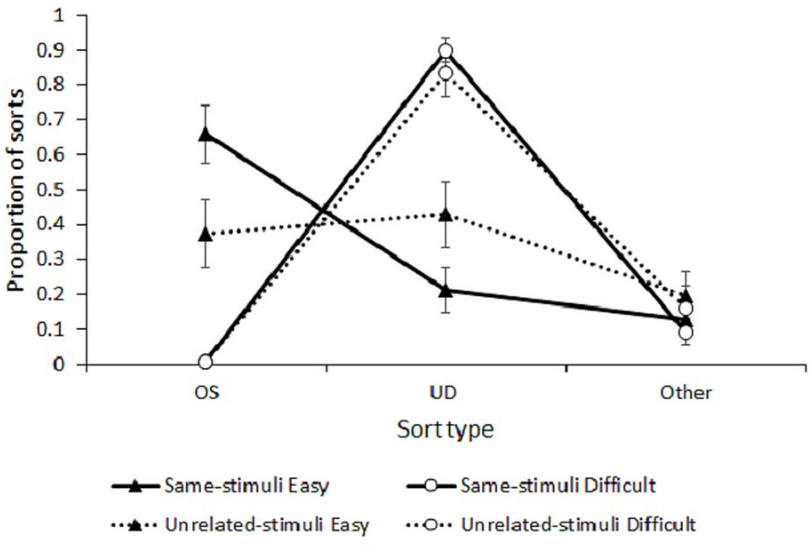

Figure 6. Mean proportion of overall similarity (OS) unidimensional (UD) and Other sorts for all conditions in Experiment 2b. Error bars represent between-subjects variability and were calculated as +/- 1 standard error of the mean. For the perceptually difficult conditions, overall similarity sorts had a low variance so the error bars are difficult to perceive.

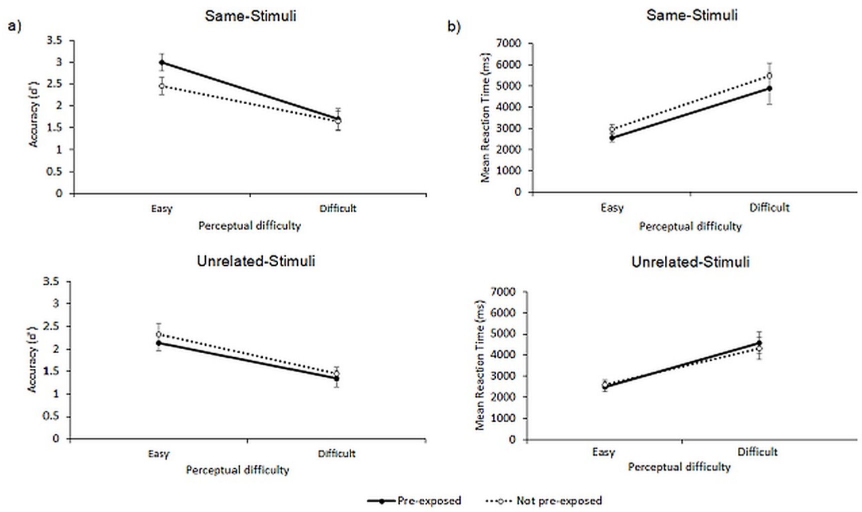

Figure 7. a) The mean proportion accuracy in the perceptual discrimination task in Experiment 2b for participants in the same-stimuli conditions and participants in the unrelated-stimuli conditions; b) Mean reaction time in the perceptual discrimination task in Experiment 2b for participants in the same-stimuli conditions and participants in the unrelated-stimuli conditions. Error bars represent between-subjects variability and were calculated as +/- 1 standard error of the mean.

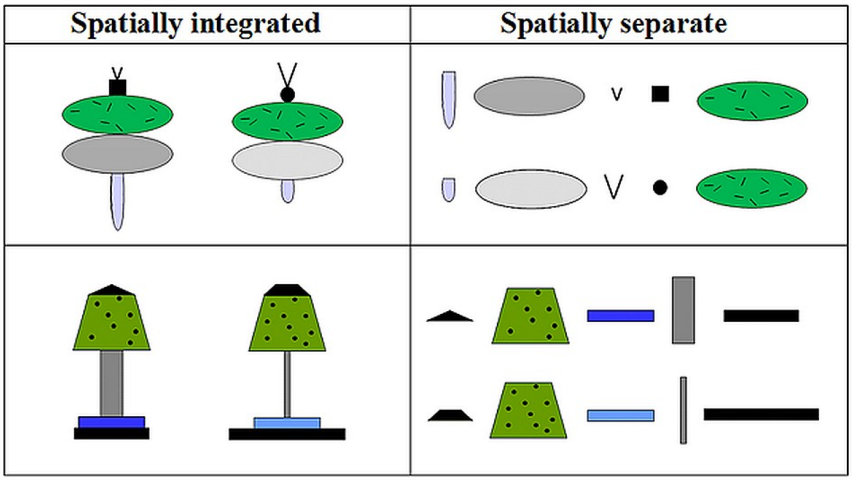

Figure 8. The category prototypes for the stimulus sets used in Experiment 3.

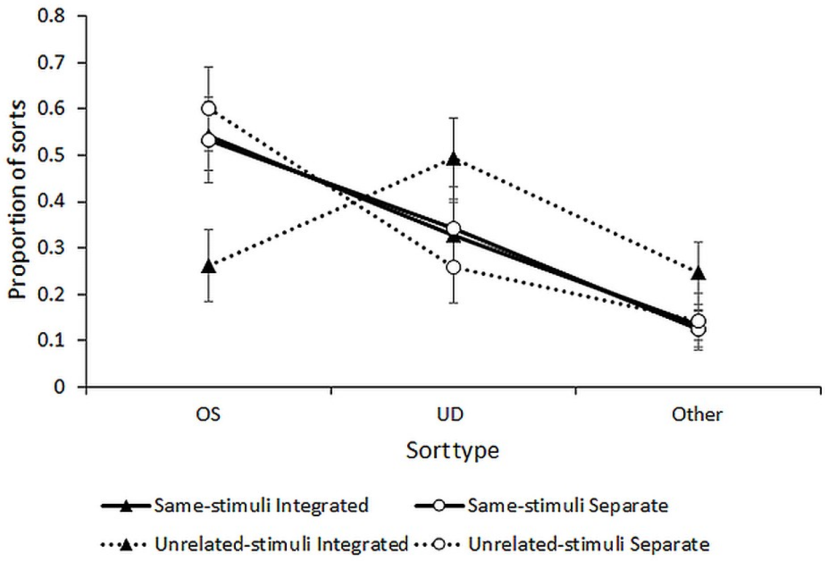

Figure 9. Mean proportion of overall similarity (OS), unidimensional (UD) and Other sorts for all conditions in Experiment 3. Error bars represent between-subjects variability and were calculated as +/- 1 standard error of the mean.

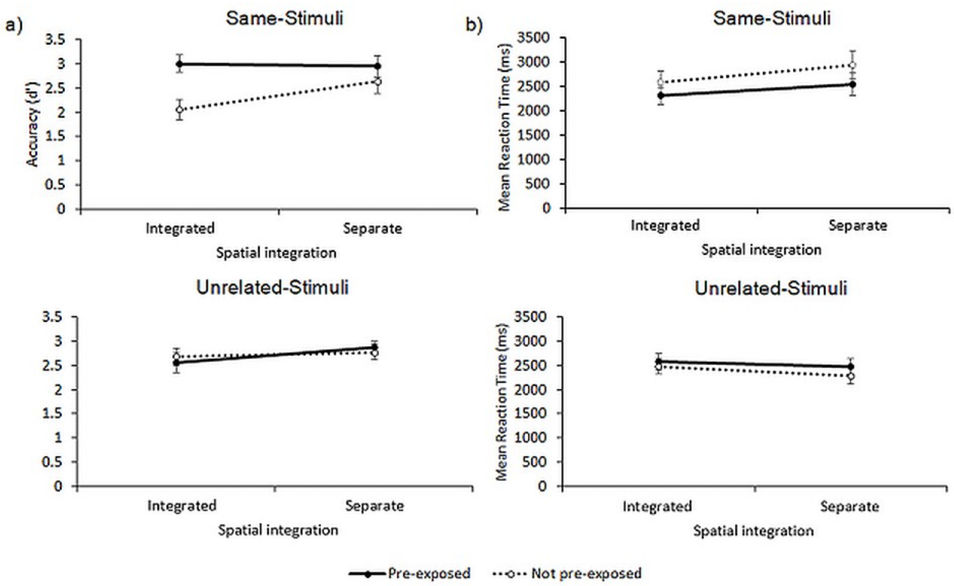

Figure 10. For the perceptual discrimination task in Experiment 3: a) The mean proportion accuracy (d') for participants in the same-stimuli conditions and participants in the unrelated-stimuli conditions. b) The mean reaction time for participants in the same-stimuli conditions and participants in the unrelated-stimuli conditions. Error bars represent between-subjects variability and were calculated as +/- 1 standard error of the mean.