Journal of Experimental Psychology:
Human Perception and Performance
2008, Vol. 34, No. 3, 676–692

0096-1523/08/$12.00 DOI: 10.1037/0096-1523.34.3.676

Processes of Overall Similarity Sorting in Free Classification

Fraser Milton, Christopher A. Longmore, and A. J. Wills
University of Exeter

The processes of overall similarity sorting were investigated in 5 free classification experiments. Experiments 1 and 2 demonstrated that increasing time pressure can reduce the likelihood of overall similarity categorization. Experiment 3 showed that a concurrent load also reduced overall similarity sorting. These findings suggest that overall similarity sorting can be a time-consuming analytic process. Such results appear contrary to the idea that overall similarity is a nonanalytic process (e.g., T. B. Ward, 1983) but are in line with F. N. Milton and A. J. Wills's (2004) dimensional summation hypothesis and with the stochastic sampling assumptions of the extended generalized context model (K. Lamberts, 2000). Experiments 4 and 5 demonstrated that the relationship between stimulus presentation time and overall similarity sorting is nonmonotonic, and the shape of the function is consistent with the idea that the three aforementioned processes operate over different parts of the time course.

Keywords: speeded categorization, free classification, analytic–nonanalytic processing, triad, match-to-standards

How do individuals come to have the categories that they do? Even a relatively small number of items can be categorized in a large number of ways; for example, two category labels can be assigned to 10 items in over 1,000 different ways. What factors determine the mappings that one actually produces?

One answer is that people classify objects by the actions they require. This view of categorization appears to underlie supervised categorization studies, in which participants acquire an experimenter-defined category structure from stimulus-specific feedback. Although such an approach has advanced understanding of categorization, it reveals little about how people construct categories in situations where feedback is rarer, nontrial specific, or absent. This question is important because, outside the laboratory, people rarely receive the level of feedback that occurs in supervised categorization experiments. Free classification (e.g., Imai & Garner, 1965)—also known as free sorting (e.g., Bersted, Brown, & Evans, 1969), category construction (e.g., Medin, Wattenmaker, & Hampson, 1987), and spontaneous categorization (e.g., Pothos & Chater, 2002)—is a methodology for investigating categorization preferences in the absence of feedback. Participants are given a set of stimuli and asked to sort them in the way that seems most sensible and natural to them. No feedback is given.

It might seem reasonable to assume that the categories people prefer to construct in such situations would reflect the underlying structure of the categories they encounter in the real-world environment. Over the years, there has been a shift from the "classical" view that natural categories are made up of necessary and jointly sufficient features (e.g., Bruner, Goodnow, & Austin, 1956; E. E. Smith & Medin, 1981) toward the idea of an overall similarity ("family resemblance") structure (e.g., Ryle, 1951; Wittgenstein, 1958). In an overall similarity structure, the category is 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. This theory has been supported by the finding that many natural categories appear to be organized around an overall similarity structure (e.g., Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

It is therefore perhaps surprising that early free sorting studies found that people rarely sort by overall similarity and instead have a strong tendency to sort on the basis of a single dimension (e.g., Ahn & Medin, 1992; Imai & Garner, 1965; Medin et al., 1987). More recent work, however, has demonstrated that although single-dimension sorting is common in free classification, overall similarity sorting can be encouraged in a number of ways. For example, background knowledge, which allows features to be integrated in a meaningful and coherent way, increases overall similarity sorting (e.g., Ahn, 1990, 1991; Kaplan & Murphy, 1999; Spalding & Murphy, 1996), as does using sequential rather than simultaneous presentation of stimuli (Regehr & Brooks, 1995). Overall similarity sorting can also be encouraged by incidental training conditions (Kemler Nelson, 1984) and by the presence of a concurrent task (J. D. Smith & Kemler Nelson, 1984). However, the main focus of the current article is the effect that processing time has on the prevalence of overall similarity sorting.

Fraser Milton, Christopher A. Longmore, and A. J. Wills, School of Psychology, University of Exeter, Exeter, United Kingdom.

This research was supported by Economic and Social Research Council (ESRC) Grants PTA-030-2003-00287 and PTA-026-27-1256 awarded to Fraser Milton and by Biotechnology and Biological Sciences Research Council Grant 9/S17109, EC Framework 6 project Grant 516542 (NEST), and ESRC Grant RES-000-22-1779 awarded to A. J. Wills. We thank Gareth Croft, Alice Welham, Thomas Beesley, Mark Suret, Koen Lamberts, and Emmanuel Pothos for their advice. We also thank Robert Nosofsky and Lee Brooks for their helpful comments and suggestions on earlier versions of this article. Related research can be found at http://www.willslab.co.uk

Correspondence concerning this article should be addressed to Fraser Milton, School of Psychology, Washington Singer Laboratories, University of Exeter, Exeter EX4 4QG, United Kingdom. E-mail: fraser@willslab.co.uk

In a typical free classification task, participants are under no explicit time pressure and can basically take as long as they wish over each decision. However, outside the laboratory, categorization often is, or has to be, performed quickly. This is illustrated by Thorpe and Imbert's (1989) calculation that classification decisions often take only about 100 ms. Ward's (1983) influential integral-to-separable model predicts that if participants are encouraged to spend less time categorizing, they are more likely to use overall similarity than a single dimension. The rationale behind this prediction is that overall similarity sorting is the product of a holistic, nonanalytic process, whereas single-dimension sorting is the product of a dimensional, analytic process, and that the latter takes more time than does the former. This idea receives converging support from developmental (L. B. Smith, 1979; L. B. Smith & Kemler, 1978) and incidental learning studies (Kemler Nelson, 1984). A general overview of the analytic–nonanalytic distinction can be found in Brooks (1978), and a formalized model that accounts for the overall similarity to dimensional responding shift is described in L. B. Smith (1989).

Ward's integral-to-separable model has been supported by the results of a minimal free classification procedure known as the triad task (e.g., J. D. Smith & Kemler Nelson, 1984; Ward, 1983). In a typical experiment, participants are presented with three stimuli that vary on two dimensions, the formal structure of which is shown in Figure 1. Participants are then asked to pick the two stimuli that go together best. Only three responses are possible. Choosing Stimuli A and B is described as a dimensional response because A and B are identical on one of the dimensions but very different on the other. Typically, on half the triads, (Figure 1a) Dimension 1 possesses two stimuli that share identical values, whereas the other half possess shared values on Dimension 2 (Figure 1b). Choosing Stimuli B and C is considered an overall similarity response as B and C are similar but not identical on both dimensions. Typically, the perceptual differences between B and C are small but relatively easy to discriminate. The third option of choosing A and C is considered a haphazard response as this neither makes use of dimensional identity nor maximizes overall similarity.

Ward (1983) classified his triad-task participants as either fast or slow responders according to whether their median response latency was greater or less than the group median. People who were classified as fast responders made significantly more overall similarity sorts than did those who were classified as slow responders. When a time-pressure manipulation was introduced, participants in the high time-pressure condition produced significantly more overall similarity sorts than did those in the low time-pressure condition. J. D. Smith and Kemler Nelson (1984) and Ward, Foley, and Cole (1986) showed that this effect was robust across a variety of different stimuli.

However, more processing time does not always increase the prevalence of overall similarity sorting. For example, J. D. Smith and Shapiro (1989) manipulated the amount of time participants were exposed to stimuli before free sorting them (although the free sort itself was self-paced). This manipulation did not affect the prevalence of overall similarity sorting. Somewhat related to this, in at least one free sorting experiment, participants have maintained a preference for dimensional responding in the face of substantial time pressure (J. D. Smith and Kemler Nelson, 1984, Experiment 4).

There are even reasons to suppose that restricting processing time might decrease overall similarity sorting under some conditions. Ward and Scott (1987) demonstrated that participants in a self-paced categorization task who used a single-dimension-plus-exception strategy sorted more rapidly than did those participants using an overall similarity strategy. This suggests that single-dimension sorting may be less time consuming than overall similarity sorting and, by extension, that increasing time pressure on participants might, under certain conditions, decrease the prevalence of overall similarity sorting. The idea that increasing time pressure might reduce overall similarity sorting under certain conditions is also supported by stochastic sampling models of categorization (e.g., Cohen & Nosofsky, 2003; Lamberts, 2000) and by the dimensional summation hypothesis (Milton & Wills, 2004). We discuss each in turn below.

Stochastic sampling models assume that the formation of an object's representation is a time-consuming process that involves the gradual accrual of perceptual information (e.g., Lamberts, 2002). Specifically, such models assume that the probability of inclusion of a stimulus dimension into the decision stage of the categorization process is an increasing function of (a) time from stimulus onset and (b) dimensional salience. Stochastic sampling models are able to successfully account for much previous speeded categorization data (cf. Lamberts, 2000), and although such models were developed to explain supervised categorization, their

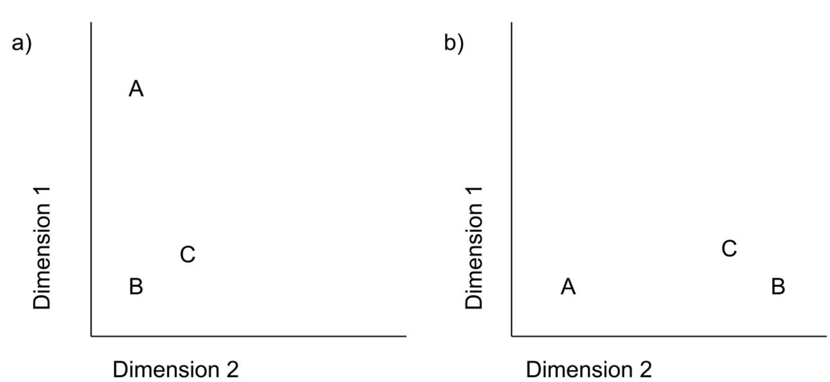

Figure 1. The abstract structure of a classification triad. In Panel a, Stimuli A and B share a value on Dimension 1; in Panel b, Stimuli A and B are identical on Dimension 2.

prediction for free classification nonetheless seems clear. Under sufficiently high time pressure, only the most salient stimulus dimension will be available to the decision process, and hence, single-dimension classification will dominate (assuming discriminability between dimension mismatches is sufficiently high). In the absence of time pressure, all dimensions enter the decision process, and overall similarity sorting is the likely result (assuming that the attention weights are approximately evenly distributed across the four dimensions).

Turning to the dimensional summation hypothesis (Milton & Wills, 2004), this theory proposes (in agreement with Ward & Scott, 1987) that both single dimension and overall similarity sorting can sometimes result from verbal, analytic processes. This theory suggests that overall similarity sorting is simply the result of a more complex and time-consuming analytic strategy than single dimension sorting. Specifically, this hypothesis predicts that overall similarity sorting can result from a process of dimensional summation where participants explicitly consider each stimulus dimension in turn and place each stimulus into the category for which it has more characteristic features. The dimensional summation hypothesis was proposed to explain situations where, contrary to previous research (e.g., Garner, 1974; Handel & Imai, 1972; Lockhead, 1972), increasing the separability of stimulus dimensions increases the prevalence of overall similarity sorting (Milton & Wills, 2004). It is plausible that spatially separating the stimulus dimensions makes them easier to differentiate, which could make a dimensional summation strategy easier to apply. In contrast, there is no obvious explanation for the observed integration effect under a nonanalytic process.

In summary, although it is well documented that increasing time pressure increases the prevalence of overall similarity sorting in free classification, there are empirical and theoretical reasons to suppose that the opposite result can also be found. In our first experiment, we attempted to find conditions under which this would be observed.

Experiment 1

The idea that increased time pressure can elevate overall similarity sorting is well supported by experiments using the triad task methodology (e.g., Ward, 1983) described in the introduction. In contrast, our contention that there may be conditions under which time pressure can reduce overall similarity sorting comes from the results of, and theories pertaining to, a number of somewhat different procedures (e.g., sequential free classification, supervised categorization). It therefore seemed logical to start our investigation by using one of these alternative procedures. In order to maximize the chances of observing an effect, it was important to use a classification procedure known to be amenable to both overall similarity and single-dimension sorting. The methodology we selected was the match-to-standards procedure previously used by Regehr and Brooks (1995) and known to produce, dependent on the testing conditions, both overall similarity and single-dimension sorting (Milton & Wills, 2004).

Method

Participants. Participants were students from the University of Exeter (Exeter, United Kingdom) who took part either for £2 or for course credit. There were 24 participants (12 in each condition) who were tested individually in a quiet cubicle.

Stimuli. The stimuli had the same structure as that used by Medin et al. (1987). The abstract stimulus structure can be seen in Table 1. The stimulus set consisted of four binary-valued dimensions (D1–D4), and the stimuli were organized around two prototypes, each representative of the two categories. We constructed these prototypes by taking all the positive values on the dimensions for one of the stimuli (1, 1, 1, 1) and all the zero values on the dimensions (0, 0, 0, 0) for the other category. The rest of the stimuli (one-aways) 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 the set. Sorting the stimuli by overall similarity, as shown in Table 1, maximizes within-group similarities and minimizes between-group similarities. The stimuli (shown in Figure 2) were the perceptually simple, spatially integrated stimuli (originally based on a schematic butterfly) used in Experiment 4 of Milton and Wills (2004). The dimensions varied in the length of the "antennae," the quantity and size of the dots in the top oval (controlled so that each type has an equal amount of black ink), the hue of the lower oval, and the length of the "body." Both categories possessed characteristic features: One category was characterized by a long body, an oval of a darker blue, a smaller V, and fewer dots, and the other category was characterized by a smaller body, an oval of a lighter blue, a larger V, and more dots.

Procedure. At the beginning of the experiment, participants were introduced to the stimuli with a pre-sort matching-pairs procedure used previously by Milton and Wills (2004). In this procedure, 20 cards were randomly spread out in an array in front of the participant. These cards consisted of the 10 stimuli in the set and an identical copy of each of them. The participant then had to match these stimuli into identical pairs without feedback. When the participant felt that the pairs had been matched correctly, the matches were then examined by the experimenter; if there were any mistakes, participants were asked to match the pairs again. The purpose of this task was to ensure that the participant could fully distinguish the four feature pairs, because if the participant had not identified all of the features, an overall similarity sort might be difficult to carry out.

After this pre-sort procedure, the method of stimulus presentation for the categorization task was similar to that used by Regehr and Brooks (1995) and identical to that used by Milton and Wills (2004). The two prototypes were placed side by side on a table throughout the experiment, and participants were informed that

Table 1
Abstract Stimulus Set Used in Experiments 1–4

Category A				Category B			
D1	D2	D3	D4	D1	D2	D3	D4
1	1	1	1	0	0	0	0
1	1	1	0	0	0	0	1
1	1	0	1	0	0	1	0
1	0	1	1	0	1	0	0
0	1	1	1	1	0	0	0

Note. D = dimension.

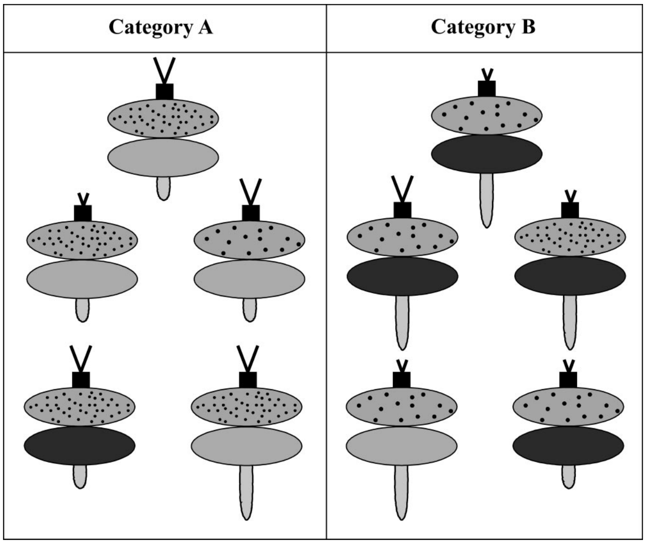

Figure 2. The stimuli used in Experiment 1, organized into their overall similarity groups. Each category consists of one prototype and four items that have three features characteristic of their category and one atypical feature.

these were characteristic of Category A and Category B (which category each prototype represented was randomized). Participants were then given all 10 stimuli in the set (i.e., the 2 category prototypes and the 8 one-away stimuli), which had been shuffled into a random order, and were asked to look at each stimulus individually and to put it into the group for which they felt it was most representative. Once they had made their decisions, participants placed each card face down directly below the category they felt it most resembled. Participants were told that there were many ways in which the stimuli could be split and that there was no one correct answer. Participants were also told that the two groups did not have to be of equal sizes and that the only constraints were that each stimulus had to be placed into one of the two groups and that they were not allowed to look through the cards that remained to be sorted or to change previous responses.[1]

At the beginning of the task, participants in the high time-pressure condition were informed that the sort had to be completed within 30 s; those participants in the low time-pressure condition were told that the task must be completed within 15 min. The time taken was measured with a manually operated stopwatch. Apart from this time-pressure manipulation, there were no differences between conditions. Once they had finished sorting the stimulus set, participants were asked to explain as precisely as possible the way in which they had classified the 10 items.

Dependent measure. Each participant was classified as having produced one of the sort types described below. These sort types are very similar to those used by Regehr and Brooks (1995) and identical to those used by Milton and Wills (2004).

A *single-dimension sort* is based on a single dimension of the stimulus. It did not matter which dimension was used as the basis of sorting, so long as all of the positive values for the chosen dimension were in one category and all of the zero values for that dimension were in the other category. Additionally, in order to receive this classification, the participant had to describe his or her sort as being based on a single dimension.

Participants were considered to have produced a *one-away single-dimension sort* if they described their sorting as being driven by a single dimension but there was a solitary error in their classification. This means that nine of the items were classified on the basis of a single dimension but that the other item was placed into the wrong category.

An *overall similarity sort*, also commonly known as a “family resemblance” sort (e.g., Medin et al., 1987), has a structure identical to that shown in Table 1. In order to receive this classification, participants had to place each of the prototypes, along with their derived one-aways, into separate categories without error. Additionally, they had to describe their strategy as being based either on general similarity or on the premise that they placed each item into the category with which it had more features in common.

A *one-away overall similarity sort* is similar to the one-away single-dimension sort with the exception that the error occurred in a sort that was otherwise arranged by overall similarity.

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

Results and Discussion

The time-pressure manipulation was effective, with participants in the high time-pressure condition ($M = 26.86$ s, $SD = 3.97$ s) taking significantly less time to complete the task than participants in the low time-pressure condition ($M = 73.64$ s, $SD = 34.07$ s), $t(22) = 4.72$, $p < .0001$ (the Welch–Satterthwaite correction for unequal variances was applied). The effects of time pressure on the prevalence of overall similarity and single-dimension sorting are shown in Table 2. As in Milton and Wills (2004), the overall similarity and single-dimension categories were combined with their respective one-aways to produce sum single-dimension and sum overall similarity categories. A chi-square test (using the sum single-dimensional and the sum overall similarity categories) demonstrated that there was a significant effect of time pressure on sort type, $\chi^2(1, N = 12) = 6.75$, $p < .01$,[2] with participants in the low time-pressure condition sorting by overall similarity to a significantly greater extent than those in the high time-pressure condition. In the low time-pressure condition, a Mann–Whitney test revealed that overall similarity sorters took significantly longer to

[1] Note that the current procedure is not free sorting in its strictest sense (e.g., Bersted et al., 1969)—participants were given the category prototypes and told how many categories should be used (i.e., two). Nevertheless, participants were provided with no feedback on their responses and were encouraged to sort the stimuli in the way that seemed most natural. In line with Regehr and Brooks (1995) and with our own previous usage (Milton & Wills, 2004), we describe the current procedure as an example of *free classification*.

[2] Yates’s correction for 2×2 chi-square tables has not been applied. Research has shown that the conventional chi-square for 2×2 designs, without correction for continuity, is sufficient to prevent Type I errors (Overall, 1980). Also, the assumption of fixed marginals made by Yates’s correction was not applicable for this data set. Additionally, no corrections have been applied for the low expected frequencies of some of the cells. It has been found that small expected frequencies do not increase the chance of Type I errors (Bradley, Bradley, McGrath, & Cutcomb, 1979). A general discussion of these issues can be found in Howell (2002, pp. 151–152).

Table 2
Sort Frequencies for Experiment 1

Condition	Sort strategy: Single-dimensional sort: 1-a 1-D	1-D	Sum 1-D	Overall similarity sort: 1-a OS	OS	Sum OS	Misc.: others
High time pressure	2	9	11	1	0	1	0
Low time pressure	0	5	5	0	7	7	0

Note. 1-a = one-away; 1-D = single-dimensional; OS = overall similarity; Misc. = miscellaneous.

complete the task than did single-dimension sorters, U (5, 7) = 2, $p < .02$ (an equivalent test is not possible in the high time-pressure condition as only 1 participant produced an overall similarity sort).

The current experiment provides the first demonstration that increases in time pressure can significantly reduce the prevalence of overall similarity sorting in a free classification procedure. Although the demonstration is novel, one could reasonably have predicted the existence of such an effect from the related work of Ward and Scott (1987), from stochastic sampling models (Cohen & Nosofsky, 2003; Lamberts, 2000), and from the dimensional summation hypothesis of overall similarity sorting (Milton & Wills, 2004). At the same time, one might have predicted the opposite result on the basis of Ward's (1983) integral-to-separable model or on the basis of the results of the triad classification tasks from which the model was formulated (J. D. Smith & Kemler Nelson, 1984; Ward, 1983). As a first step, we therefore attempted to replicate this somewhat surprising finding. To this end, in Experiment 2 we used a different stimulus set and a somewhat different procedure in an attempt to increase the generality of the findings of Experiment 1.

Experiment 2

Experiment 1 provides the first demonstration that time pressure can reduce the level of overall similarity sorting under a free classification procedure. As such, it provides support for the theory of Milton and Wills (2004) that overall similarity sorting can be due to an effortful, analytic process. This appeared a potentially significant finding, and it therefore seemed important to establish its generality. To do this, in Experiment 2 we introduced a variation on the match-to-standards procedure used in Experiment 1 that controlled for a number of factors that one might argue should ideally be more tightly constrained (e.g., per-stimulus presentation time, type of motoric response emitted, and participant–experimenter interactions). In addition, Experiment 2 required a larger number of classification responses from each participant, which should further increase confidence in the generality of the findings. The stimulus set used in Experiment 2 was also different to that used in Experiment 1. Our hypothesis was that Experiment 2 would show the same pattern of results as in Experiment 1. In other words, an increase in time pressure would result in a reduction in the prevalence of overall similarity sorting. Such a result would demonstrate that the findings of Experiment 1 were replicable and were robust against certain procedural and stimulus variations.

One of the more substantial differences between Experiments 1 and 2 was the method by which time pressure was applied. In Experiment 1, we applied time pressure by reducing the total amount of time available. In Experiment 2, we applied time pressure individually to each trial, and we achieved this by reducing the amount of time each stimulus was presented. After the stimulus disappeared, there was no limit on the amount of time allowed to make the classification decision. This procedure has been used successfully in past research to study time constraints in supervised categorization (e.g., Lamberts & Freeman, 1999).

We selected a presentation time manipulation over a response signal ("hurry up!") procedure (e.g., Lamberts, 1998) for two reasons. First, a presentation time manipulation arguably allows a better characterization of the form of time pressure being applied. A response signal procedure potentially affects both stimulus-dependent processes (e.g., perceptually based processing) and stimulus-independent processes (e.g., guessing), whereas the time pressure under a presentation time procedure is likely to predominantly affect stimulus-dependent properties. Second, the presentation time procedure is arguably less similar to the Experiment 1 methodology than a response-signal procedure, and hence, if the effect persists, the case for its generality would increase.

Method

Participants and apparatus. Participants were students from the University of Exeter who took part either for course credit or for £3. There were 28 participants (14 in each condition) who were tested individually in a quiet testing cubicle. We tested participants with E-prime (Psychological Software Tools, 2002), which was run on a Pentium III PC with a 17-in. monitor and a standard computer keyboard. Participants sat approximately 0.5 m away from the screen.

Stimuli. The stimuli (shown in Figure 3) were line drawings of boats closely modeled on those used by Lamberts (1998). The stimuli took the same abstract structure as that used in Experiment 1 (see Table 1). As before, there were 10 stimuli in the set. Each stimulus was 10 cm high and 8.5 cm wide. The four stimulus dimensions were the flag (rectangular vs. triangular), the size of the sail (small vs. large), the porthole (circular vs. diamond shaped), and the hull (long base vs. short base).

Procedure. Participants were randomly allocated to one of the two between-subject conditions. The two conditions were identical except for the stimulus presentation times, which were 1,024 ms in the high time-pressure condition and 4,096 ms in the low time-

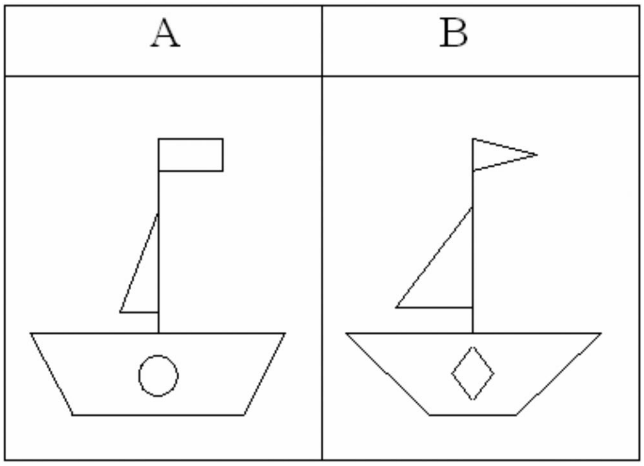

Figure 3. The prototypes of the stimulus set used in Experiments 2 and 3.

pressure condition. Participants were informed at the beginning of the experiment that they were to take part in a categorization task. They were told that there were many ways in which the stimuli could be split and that there was no one correct answer. They were also told that the groups did not have to be of equal sizes and that they should classify each stimulus in the category of which they felt it was most representative.

At the beginning of each trial, the two prototypes (shown in Figure 3) were presented on the screen. The prototype of Category A was presented in the middle left portion of the screen, and the prototype of Category B was presented in the middle right portion of the screen. Directly above each prototype was a label that indicated which category (i.e., A or B) that prototype represented. The category that each prototype represented was fixed across participants. These prototypes remained on the screen until the participant pressed the space bar to continue. The screen then went blank for 500 ms, and this was followed by a central fixation cross for 500 ms, before the target stimulus, which was the same size as the prototypes, was presented in the center of the screen for the appropriate duration (i.e., 1,024 or 4,096 ms). The target stimulus was immediately followed by a medium grayscale mask (11.5 cm high and 9.5 cm wide), which stayed on the screen until participants chose a response (pressing either *C* or *M* on the keyboard for Categories A and B, respectively). The screen then went blank for 1,000 ms before the next trial began.

Participants were presented sequentially with a total of 120 stimuli, in 12 blocks of 10 trials. In each block, each stimulus in the set (the abstract stimulus structure of which is shown in Table 1) was presented once in a random order. Participants were given the chance to pause at the end of each block. During this time, they were also asked to write down, as precisely as possible, the way in which they had sorted the stimuli in the previous block.

Results and Discussion

For each participant, the sort type for each of the 12 blocks was analyzed separately in the same manner as in Experiment 1. These sorts were then placed into one of three categories: overall similarity, single dimension, and other (as in Experiment 1, the overall similarity and single-dimension categories were combined with their respective one-aways).

The mean proportion of overall similarity sorts was 0.18 for participants in the 1,024-ms condition and 0.58 for participants in the 4,096-ms condition. An independent-samples *t* test confirmed that this difference was significant, $t(26) = 2.96$, $p < .01$. The mean proportion of single-dimension sorts for each condition showed a corresponding effect (0.61 for the 1,024-ms condition versus 0.29 for the 4,096-ms condition), $t(26) = 2.32$, $p < .05$. The effect of stimulus presentation time on the frequency of other sorts was not significant, $t(26) = 1.05$, $p > .3$ (0.22 for the 1,024-ms condition and 0.13 for the 4,096-ms condition).

In summary, Experiment 2 replicated the effect shown in Experiment 1 with a different method of applying time pressure, a larger quantity of trials, and a different stimulus set. Experiment 2 therefore supports and extends the generality of the findings from Experiment 1 that time pressure can lead to a reduction of overall similarity sorting and an increase in single-dimension sorting.

Although the main focus of this article is the effect time pressure has on sorting behavior, our hypothesis that overall similarity sorting can be the result of an effortful, analytic process is relatively novel. It therefore appears important to establish a converging line of evidence to support our analytic account of the overall similarity sorting we observed before attempting to elucidate the reasons for the discrepancy between Experiments 1 and 2 and previous work (e.g., J. D Smith & Kemler Nelson, 1984; Ward, 1983). This seems particularly important when one considers that support for a nonanalytic system of overall similarity sorting comes from a number of converging lines of evidence, such as developmental (e.g., L. B. Smith & Kemler, 1978) and incidental learning studies (e.g., Kemler Nelson, 1984). To this end, in Experiment 3 we introduced a concurrent load task to further test the hypothesis that overall similarity sorting can be the result of an effortful, analytic process.

Experiment 3

In a classic experiment, J. D. Smith and Kemler Nelson (1984) demonstrated that the prevalence of overall similarity sorting in a triad classification task is increased by the introduction of a concurrent cognitive load. This result is consistent with the idea that the overall similarity sorting they observed was the result of a relatively automatic nonanalytic process. If, as we hypothesize, the overall similarity sorting observed in the current research is the result of an effortful, analytic strategy, a reversal of J. D. Smith and Kemler Nelson's finding might be predicted. In other words, one might anticipate that the introduction of a concurrent load would reduce the level of overall similarity sorting compared with a no-load condition.

Method

Participants, apparatus, and stimuli. Forty-two students from the University of Exeter took part either for course credit or for £5 payment. Twenty-one participants were allocated to each of the two between-subject conditions (load and no load). The apparatus and stimuli were identical to those used in Experiment 2.

Procedure. The procedure was a modification of that used in Experiment 2. As before, participants had to classify 12 blocks of

10 stimuli; however, the stimulus presentation time was 1,500 ms for all participants, and concurrent load was manipulated via a digit-probe task. In the load condition, the two prototypes (shown in Figure 3) were presented on the screen at the beginning of each trial as in Experiment 2. These prototypes remained on the screen until the participant pressed the space bar. The screen then went blank for 500 ms, and participants were presented with six random numbers over headphones. These numbers ranged between 1 and 9, and one number was presented every 330 ms. Hence, the screen remained blank for a total of 2,480 ms (6 numbers × 330 ms, plus the initial blank screen of 500 ms). A fixation cross then appeared on the screen for 500 ms followed by a blank screen for 500 ms. The target stimulus was then presented for 1,500 ms and was immediately followed by a medium grayscale mask that remained on the screen until participants made their classification response. A digit previously presented in the trial (although never the last one) was then presented on the screen. Participants were then required to press the number on the keyboard that came after the number presented on the screen in the auditory sequence. For example, if the auditory sequence was 3 . . . 5 . . . 2 . . . 8 . . . 6 . . . 7 and the number on the screen was 5, the correct response was to press *2* on the keyboard. The next trial then began immediately. Participants were not provided with feedback on their performance on the task.

The no-load condition differed from the load condition in two ways: First, no digits were presented over headphones in the no-load condition (the duration between the offset of the prototypes and the onset of the target stimuli was still 2,480 ms, however). Second, after the classification response, participants in the no-load condition were simply asked to press the space bar to continue (rather than press a number key).

In both conditions, participants were given the chance to pause at the end of each block. During this period, they were asked to write down, as precisely as possible, the way they sorted the stimuli in the previous block.

Results and Discussion

Performance on the digit probe task was good, with a mean accuracy of 76.38% ($SD = 17.17$). As in Experiment 2, the sort types for each participant were analyzed for each block and placed into one of three categories: overall similarity, single dimension, and other.

The mean proportion of overall similarity sorts produced was 0.17 for participants in the load condition and 0.46 for participants in the no-load condition. An independent-samples *t* test confirmed that this difference was significant, $t(40) = 2.44$, $p < .05$. The mean proportion of single-dimension sorts for each condition showed a corresponding pattern, with participants in the load condition producing significantly more single-dimensional sorts ($M = 0.72$) than participants in the no-load condition ($M = 0.45$), $t(40) = 2.15$, $p < .05$. There was no significant difference in the level of other sorts between the load ($M = 0.11$) and the no-load ($M = 0.09$) conditions, $t(40) = 0.47$, $p > .5$.

In summary, Experiment 3 shows that the introduction of a concurrent cognitive load can significantly reduce overall similarity sorting and increase single-dimensional responding (for a related demonstration in supervised category learning, see Zeithamova & Maddox, 2006). Our result is consistent with the idea that the introduction of a concurrent cognitive load inhibits the application of a dimensional-summation analytic strategy (which would produce overall similarity sorting, but is hypothesized to require significant working memory capacity) while still leaving sufficient resources to perform a simple single-dimensional analytic strategy. As such, the results of Experiment 3 appear to provide converging evidence for the case made in Experiments 1 and 2—in other words, that overall similarity sorting can be the result of an effortful, analytic strategy.

Although Experiment 3 provides converging support for Experiments 1 and 2, it leaves a crucial question unanswered: Why are the results of the time pressure manipulations in Experiments 1 and 2 essentially opposite to a number of previously reported findings (e.g., J. D. Smith & Kemler Nelson, 1984; Ward, 1983)? In Experiments 4 and 5, we pursued this line of inquiry by attempting to elucidate the important differences between the methodologies under which these two contrasting sets of findings occur.

Experiments 4 and 5: Time Pressure, Total Time, and Task Demands

Previous experiments by, for example, Ward (1983) and J. D. Smith and Kemler Nelson (1984), clearly indicate that increasing time pressure elevates the prevalence of overall similarity sorting, implicating a quick, nonanalytic processing strategy. Experiments 1 and 2 of the current article show the opposite result—namely, that increasing time pressure decreases the prevalence of overall similarity sorting, implicating a time-consuming, analytic system of overall similarity sorting. We believe that the difference between these two sets of studies is more likely to lie in differences in methodology between the two procedures than in the unreliability of one of them. The purpose of Experiments 4 and 5 was to begin to uncover the important differences and, by so doing, increase understanding of the processes of free classification.

There seem to be two main candidate explanations for the difference between the former studies and the current ones. The first explanation is based on differences in task demands. More specifically, the complexity of the behavior that is required to emit a dimensional response differs between the two procedures. In the match-to-standards procedure, a simple single-dimension rule is easily applied and can be maintained consistently throughout the sort. This means that a single-dimension strategy is likely to require less time than does an overall similarity strategy carried out via dimensional summation. In contrast, for the triad task, as J. D. Smith and Shapiro (1989) have previously highlighted, an identical-attribute dimensional response involves repeated switching between dimensions, as the dimension on which two of the stimuli are identical continually changes. Switching dimensions has been shown to require time (e.g., Proctor & Fisicaro, 1977), so one possibility is that in a triad task, the identical-attribute dimensional response is more time consuming than an overall similarity response. If this hypothesis is correct, then both sets of experiments suggest that when time pressure is limited, people opt for the less time-consuming strategy (single-dimension for the match-to-standards procedure and overall similarity for the triad task), and this explains the apparently opposing results found. This explanation will be examined in Experiment 5.

The second explanation is based on differences in level of time pressure—more specifically, that although the current studies and those of Ward (1983) and J. D. Smith and Kemler Nelson (1984) both manipulate time pressure, the level of time pressure is arguably higher in the previous studies than in ours. For example, J. D. Smith and Kemler Nelson (1984) allowed an average of about 1 s per response in their high time-pressure condition. Although this is a comparable amount of time to the high time-pressure condition of our Experiment 2, in a triad task three stimuli have to be examined before a response can be made, whereas in our procedure only one stimulus was presented on each trial. In the case of our Experiment 1, three stimuli were presented on each trial; however, participants were allowed an average of 3 s per response in the high time-pressure condition. Hence, it is possible that moderate time pressure (as we hypothesize is present in our studies) leads to a reduction in overall similarity sorting whereas more severe time pressure (as we hypothesize was present in previous studies) leads to an increase in overall similarity sorting.

The nonmonotonic relationship between time pressure and overall similarity sorting proposed by this time pressure explanation is far from intuitive, but it is testable and is not a prediction of the task demands explanation (which assumes that, within a given procedure, the relationship between time pressure and overall similarity sorting is monotonic). In contrast, the proposed nonmonotonic relationship emerges as a consequence of assuming that Ward's (1983) integral-to-separable model and Milton and Wills's (2004) dimensional summation hypothesis are both correct but operate over different parts of the time course. These two theories can be integrated in the following way: Early in processing, stimuli are treated nonanalytically as integral "blobs," and this leads to overall similarity sorting (Ward's "integral" phase). Somewhat later on, analytic processing begins, the stimuli are separated into their constituent dimensions, and this permits dimensional responding (Ward's "separable" phase). Milton and Wills's (2004) dimensional summation hypothesis concurs that, at this stage, sorting will be single dimensional because time pressure is still relatively high, and so, although there is sufficient time to apply a dimensional rule, there is insufficient time to apply the more complex and time-consuming dimensional-summation rule. However, as time available for processing continues to increase, it becomes possible to apply the dimensional summation rule, and hence, the prevalence of overall similarity sorting begins to increase again.

What such an explanation so far neglects are the predictions of stochastic sampling models of categorization (Cohen & Nosofsky, 2003; Lamberts, 2000). Some of the most convincing evidence for these theories comes from studies of supervised categorization where stimulus presentation time is extremely short. For example, Experiment 2 of Lamberts and Freeman (1999) manipulated stimulus presentation time in the region of 33 ms to 200 ms. This level of time pressure is higher than in either the triad or the match-to-standard tasks discussed so far and raises the possibility that stochastic sampling models are also correct but operate at the earliest parts of the time course. If this possibility is correct, then at very short presentation intervals, reducing presentation time should reduce the prevalence of overall similarity sorting in free classification. This is because at very short intervals, only the most salient dimension will have been sampled, and so systematic responding can be based only on a single dimension.

In summary, we are suggesting that stochastic sampling models, the integral-to-separable model, and the dimensional-summation hypothesis are all correct but that they operate on different parts of the categorization time course. If this suggestion is correct, the resultant process might be characterized as follows: The process starts by the perceptual sampling of stimulus dimensions, with the information so acquired being used by a fast (and presumably nonanalytic) system to build up a holistic stimulus representation. At a later stage, another (presumably analytic) system breaks this holistic representation down into its components. This later system is highly flexible and can apply many different classification rules, but complex rules such as dimensional summation take longer to apply than do such simple rules as single-dimension responding.

This combined theory is complex, but it also makes a clear and complex prediction about the relationship between time pressure and the prevalence of overall similarity sorting. At very short stimulus-presentation times, single-dimension responding will dominate (because only one dimension has been sampled). As presentation time is gradually increased, the prevalence of overall similarity sorting will rise (as the holistic representation is formed), fall (as analysis begins), and rise again (as time begins to permit dimensional summation). A complementary pattern should be seen in the prevalence of single-dimension sorting. This complex a priori prediction was tested in Experiment 4.

Experiment 4

In Experiment 4, as in Experiment 2, we used a match-to-standards procedure with time pressure manipulated by varying stimulus presentation time. Detection of the pattern described above requires a minimum of four time pressure conditions. Selection of the presentation times to be used is clearly critical, but in the absence of directly comparable previous studies, it must be based on a number of extrapolations, generalizations, and assumptions. Bearing these caveats in mind, our rationale for the four presentation times selected (64 ms, 256 ms, 384 ms, and 640 ms) is described below.

The shortest presentation time (64 ms) was based on Lamberts and Freeman (1999), where presentation times in the order of 60 ms were the smallest that produced above-chance responding for all stimuli. Although their procedures are somewhat different than ours, the stimuli being presented are of approximately the same complexity.

The next presentation time (256 ms) was also based on Lamberts and Freeman (1999) and, to a lesser extent, on J. D. Smith and Kemler Nelson (1984). In Lamberts and Freeman, presentation times in the order of 200–300 ms produced responding that seems to be explicable only if all four stimulus dimensions influence responding. In J. D. Smith and Kemler Nelson, the high time-pressure condition involved classifying about one triad per second. In that time, three stimuli have to be inspected and a response emitted. If one assumes that each of the three stimuli is inspected for approximately the same amount of time and that emitting a response takes an additional 100–400 ms, a presentation time for a single stimulus in the order of 200–300 ms is again suggested. Because 256 ms is close to the center of this range, we hypothesized that holistic responding would be observed at this time point.

Whereas 256 ms is a presentation time that we estimated would produce holistic responding, 4,096 ms is an interval where we propose dimensional summation sorting occurred (see Experiment 2). A reduction in overall similarity sorting with reduced time

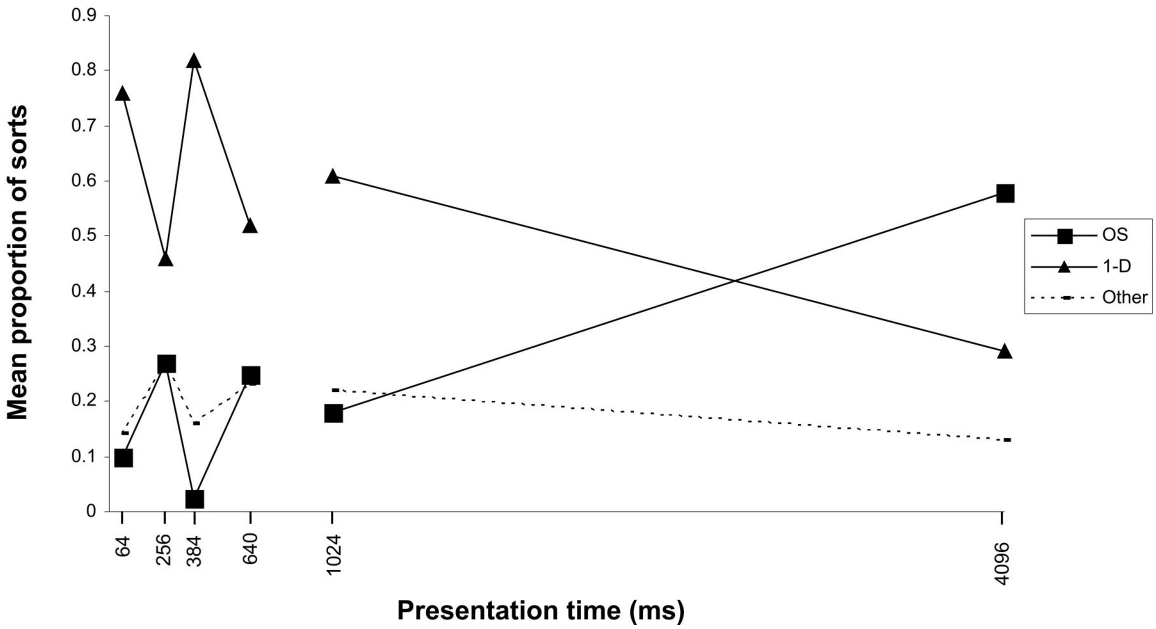

Figure 4. The frequency of overall similarity (OS), single-dimensional (1-D), and other sorts for each condition in Experiments 2 and 4.

pressure has never been observed in a match-to-standards procedure. Demonstrating a drop in overall similarity sorting subsequent to 256 ms therefore seemed critical to our hypothesis but also easy to miss as some overall similarity sorting is already detectable at 1,024 ms (Experiment 2). According to our account, single-dimension analytic responding should be maximized somewhere between 256 ms and 1,024 ms. We decided to investigate presentation times of 384 ms and 640 ms in this interval in our attempt to observe a drop in overall similarity sorting.

To summarize, our a priori prediction was that the prevalence of overall similarity sorting in a match-to-standards procedure would rise, fall, and rise again as presentation time increased. Prevalence of single-dimension sorting should show a complementary pattern. Demonstration of the predicted pattern of results would be consistent with the hypothesis that stochastic sampling models (e.g., Lamberts, 2000), Ward's integral-to-separable model, and the dimensional summation hypothesis (Milton & Wills, 2004) are all correct but operate over different parts of the categorization time course.

Method

Participants, apparatus, and stimuli. Fifty-six undergraduates from the University of Exeter took part either for course credit or for £3 payment. There were 14 participants allocated to each of the four between-subject conditions. The apparatus and stimuli were identical to those used in Experiments 2 and 3.

Procedure. The procedure was basically identical to that used in Experiment 2. As before, participants had to classify 12 blocks of 10 stimuli in one of a number of presentation time conditions. In the current experiment, there were four presentation time conditions: 64, 256, 384, and 640 ms.

Results and Discussion

For each participant, as in Experiments 2 and 3, the sort types were analyzed for each block separately and placed into one of the three categories: overall similarity, single dimension, and other.

The mean proportion of overall similarity sorts produced for the four conditions are shown in Figure 4, with the results for Experiment 2 (1,024 ms and 4,096 ms) also included for comparison. A between-subjects analysis of variance (ANOVA) of the conditions of Experiment 4 revealed that there was a significant main effect of condition on overall similarity sorting, $F(3, 52) = 3.53, p < .05, \eta_p^2 = .17$. Trend analyses showed that there was a significant cubic (polynomial order 3) trend, $F(1, 52) = 9.84, p < .005$, but no significant linear or quadratic trends (both $ps > .4$).[3] Post hoc Fisher's least-significant difference (LSD) analyses revealed that the difference between the 64-ms and the 256-ms conditions approached significance ($p = .07$), with participants in the 256-ms condition producing a higher proportion of overall similarity sorts than did participants in the 64-ms condition. Participants in the 256-ms condition produced a significantly higher proportion of overall similarity sorts than did those in the 384-ms condition ($p < .01$), and participants in the 640-ms condition produced a significantly higher proportion of overall similarity sorts than did participants in the 384-ms condition ($p < .02$).

The mean proportion of single-dimension sorts showed a corresponding pattern (see Figure 4). A between-subjects ANOVA revealed that there was a significant main effect of condition on the

[3] Trend analyses were carried out with millisecond intervals as well as with equal intervals. The findings for both the overall similarity trend analyses and the single-dimensional trend analyses (presented subsequently in the main text) were the same for both types of analyses. Only the fixed interval results have been reported.

proportion of single-dimensional sorting, $F(3, 52) = 4.49$, $p < .01$, $\eta_p^2 = .21$, with trend analyses showing a significant cubic trend, $F(1, 52) = 12.53$, $p < .005$, but no significant linear or quadratic trends (both ps $> .3$). Post hoc Fisher's LSD tests revealed that participants in the 64-ms condition produced a significantly higher proportion of single-dimensional sorts than did those in the 256-ms condition ($p < .02$). Participants in the 384-ms condition sorted single dimensionally to a significantly greater extent than did those in the 256-ms condition ($p < .005$), and participants in the 384-ms condition produced a significantly higher proportion of single-dimensional sorts than did those in the 640-ms condition ($p < .02$). There was no significant effect of time pressure on the quantity of other sorts, $F(3, 52) = 1.30$, $p > .2$.

Independent-samples *t* tests comparing sorting behavior in the 640-ms condition with the 1,024-ms and 4,096-ms conditions in Experiment 2 were also conducted. Although cross-experiment comparisons must be treated with caution, in this case the comparisons seem justified on the basis that the experiments were run with the same stimuli and procedure and because participants were from the same population. There were no significant differences between the 640-ms condition and the 1,024-ms condition in the level of overall similarity sorting, $t(26) = 0.62$, $p > .5$, or single-dimensional sorting, $t(26) = 0.67$, $p > .5$. Participants in the 4,096-ms condition sorted by overall similarity to a significantly greater extent than did those in the 640-ms condition, $t(26) = 2.53$, $p < .02$, whereas the difference between the two conditions in terms of single-dimension sorting approached significance, $t(26) = 2.00$, $p = .056$, with participants in the 640-ms condition producing more single-dimensional sorts than did those in the 4,096-ms condition.

In summary, the results of Experiment 4 show that the frequency of overall similarity sorting is initially extremely low; it then rises (although note $p = .07$), is followed by a significant fall, and is then followed by a significant rise. Conversely, the frequency of single-dimension sorting starts high and then decreases significantly, before increasing significantly, before significantly decreasing again. Taken together these results provide strong evidence for the idea that the relationship between stimulus presentation time and overall similarity sorting is nonmonotonic. More specifically, these results suggest that the discrepancy between, on the one hand, the results presented in Experiments 1 and 2 and, on the other hand, certain previous findings (e.g., Ward, 1983) could be due to the differences in the level of time pressure applied. Although these results do not, of course, rule out the possibility that differing task demands between the two sets of studies could also have an influence on sorting, such an account appears unable to explain the nonmonotonic pattern of overall similarity sorting observed in the current experiment. The results of Experiment 4 are in accordance not only with previous findings that time pressure leads to an increase in overall similarity sorting (e.g., J. D. Smith & Kemler Nelson, 1984; Ward, 1983) but also with the results of Experiments 1 and 2, which showed that low time pressure can increase overall similarity sorting, and with the result that under the most extreme time pressure, categorization is based only on the most salient dimension (e.g., Lamberts & Freeman, 1999).

Experiment 5

In Experiment 4, we investigated the hypothesis that the discrepancy between the results of Experiments 1 and 2 and previous findings (e.g., J. D. Smith & Kemler Nelson, 1984; Ward, 1983) was due to differences in the levels of time pressure applied. This explanation proposed that overall similarity sorting can be due to both analytic and nonanalytic processes, depending on the level of time pressure applied. This hypothesis is consistent with the nonmonotonic relationship between time pressure and overall similarity sorting observed in Experiment 4. In contrast, such a result does not intuitively emerge from the idea that the discrepancy is due to variations in task demands between the two procedures. Nevertheless, the more general case that there are important task demand differences between the triad and match-to-standards procedures, and that these differences may lead to different behavior in the two procedures, remains plausible and merits further investigation.

To this end, in Experiment 5 we explored the time course of overall similarity sorting under a triad task procedure. Previous work suggests that in a triad task, the level of overall similarity sorting simply rises with increasing time pressure (e.g., Ward, 1983). Therefore, according to a task-demands hypothesis, the time course of overall similarity sorting under the triad procedure might be expected to be monotonic. In contrast, according to the levels-of-time-pressure account, time pressure should also have a nonmonotonic effect on the prevalence of overall similarity sorting in the triad task. Specifically, the levels-of-time-pressure account predicts that overall similarity sorting will start relatively low. This is based on stochastic sampling models (e.g., Lamberts, 2000), which assume that only one dimension will be sampled under a very high level of time pressure. This makes it difficult to respond consistently by overall similarity. Somewhat later on, both dimensions can now be sampled, and the stimuli are treated nonanalytically as integral blobs, leading to a rise in overall similarity sorting. With decreasing time pressure, the stimuli are analyzed via their constituent dimensions, and this will lead to a reduction in overall similarity sorting.

The levels-of-time-pressure hypothesis and the task-demands hypothesis should not be considered to be mutually exclusive, and it is at the lowest levels of time pressure that we suspect that the greatest effect of differing task demands may be observed. Specifically, we observe that in the match-to-standards procedure both single-dimensional and overall similarity sorting can make use of the presence of identical values on dimensions (e.g., that the shape of the hull is identical to the shape in one of the two prototypes). In contrast, in the triad task, only dimensional sorting can make use of the presence of identical values (e.g., that two of the stimuli, although they differ in brightness, are identical in size). Overall similarity responding in a triad task requires two stimuli to be grouped together that do not match precisely on any dimension. Under nonanalytic sorting, the absence of identical matches is perhaps not important as one might hypothesize that participants are acting holistically and not analyzing the constituent dimensions. However, the absence of identical matches may have an important effect for highly analytic responders (as we assume those at the lowest levels of time pressure are). We suggest that under a triad structure the absence of shared values for overall similarity sorting and, just as important, the presence of shared values for the dimensional rule make an overall similarity strategy a less attractive option than a dimensional response. In a match-to-standards procedure, identical matches can be utilized either to produce a single-dimension classification or to produce an overall similarity classification, and under these conditions, highly analytic responders may find it unsatisfactory to ignore most of the stimulus properties and produce a single-dimension sort.

In summary, we are proposing that both the level-of-time-pressure hypothesis and the task-demands hypothesis are correct. The level-of-time-pressure hypothesis leads us to predict that the relationship between time pressure and overall similarity sorting will be nonmonotonic in a triad task, whereas a form of task-demands hypothesis leads us to suspect that the final increase in overall similarity sorting seen at low time pressures in a match-to-standards task may not be seen in a triad task. Specifically, we hypothesize that overall similarity sorting will start low; it will then rise, before falling again.

The five stimulus presentation time conditions we chose to investigate were 640 ms, 1,024 ms, 2,048 ms, 3,072 ms, and 7,500 ms. These presentation times were longer than those used in Experiment 4 due to the greater number of stimuli presented on each trial in the triad procedure compared with the version of the match-to-standards procedure used in Experiment 4. In Experiment 5 (using the triad procedure), three stimuli were presented on the screen on each trial, whereas only one stimulus appeared per trial in Experiment 4 (which used the match-to-standards procedure). Furthermore, the stimuli in Experiment 5 covered a greater proportion of the screen than in Experiment 4 (due to the increased number of stimuli presented on each trial). This means that more time may be required to fixate all the information in the current experiment. Pilot work using the same stimuli and procedure as in Experiment 5 but with a stimulus presentation time of 288 ms ($N = 22$) revealed sorting behavior that did not differ significantly from chance responding. We therefore suggest that in a triad procedure, the 640-ms presentation time used in Experiment 5 may be close to the shortest time in which the current stimuli can be processed sufficiently for participants to respond in a coherent manner.

In summary, in Experiment 5 we used five time conditions to explore the time course of overall similarity sorting within a triad task procedure. It is perhaps noteworthy that this number of time conditions goes considerably beyond the amount used in previous speeded triad classification studies. The current experiment should therefore not only better characterize the time course of overall similarity sorting under a triad task procedure but also help further elucidate the discrepancy between the results of Experiments 1 and 2 and those of previous studies (e.g., Ward, 1983).

Method

Participants and apparatus. One hundred fifty students from the University of Exeter were recruited to take part either for course credit or for £3. Participants were tested individually in a quiet testing cubicle. There were 30 participants in each of the five between-subject conditions. The experiment was run on a Pentium IV PC with a 17-in. monitor and a standard computer keyboard.

Stimuli. The stimuli were modified versions of the boats used in Experiments 2, 3, and 4 and are illustrated in Figure 5. The porthole and flag dimensions were removed, and four values were created for both the hull and the sail dimensions. For each dimension, there were two extreme values and two neighboring values that were close but not identical to one of the extreme values. In total, eight unique stimuli (all sized 5.4 cm high by 7.7 cm wide) were constructed from these values.

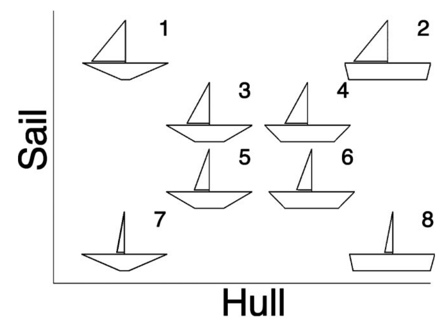

Figure 5. The stimuli used to construct the triads in Experiment 5.

Given the formal structure of the triad task (which is shown in Figure 1), eight stimulus triads can be created from these eight stimuli: 1-3-7, 1-5-7, 2-4-8, 2-6-8, 1-2-3, 1-2-4, 5-7-8, and 6-7-8 (see Figure 5). Each stimulus triad was presented in a triangular formation around the center of the screen; one stimulus to the left, one to the right, and one at the top centered directly between the left and right stimuli. There was 1.1 cm between the bottom of the top boat and the top of the left and right boats and 2.7 cm between the edges of the left and right boats. There are six ways in which the three stimuli can be placed in three spatial locations, and participants saw all six permutations of all eight triad types. In other words, each of the eight triad types was presented six times, making 48 triads in total.

Procedure. Participants were randomly allocated to one of the five between-subject conditions, which were identical except for stimulus presentation time. The presentation times used were 640 ms, 1,024 ms, 2,048 ms, 3,072 ms, and 7,500 ms. The procedure was similar to that used in previous triad classification studies (e.g., Ward, 1983), although it was computer based for more tightly controlled stimulus presentation times. The 48 triads were presented in a random order, and for each triad, participants were required to choose the two stimuli that went together best. At the beginning of each trial, a fixation cross was presented for 500 ms, followed by a blank screen for 500 ms. The triads remained on the screen for the appropriate duration and were immediately followed by a rectangular, medium grayscale mask (sized 20 cm high by a maximum 20 cm wide), with the message "Please respond now" presented underneath the mask. The mask remained on the screen until participants made a response (by pressing either the left, right, or up cursor key). To make their decision, participants pressed the cursor key on the keyboard that corresponded to the stimulus that they were excluding (as in, for example, J. D. Smith & Kemler Nelson, 1984).[4] For instance, if they felt that the left and right stimuli went together best, they pressed the up cursor key, and if they felt that the top and right stimuli went together best, they pressed the left cursor key. The screen then went blank for 1,500 ms before the next trial began.

[4] This is arguably another example of differing task demands between the triad and match-to-standards procedures. In the triad task, participants exclude one of the three stimuli, which may encourage them to search for differences between the stimuli. The match-to-standards procedure may therefore encourage participants to search for matching features to a greater extent. We minimized this difference by explicitly asking participants to put together the stimuli that go together best and not to think of it as an odd-one-out task.

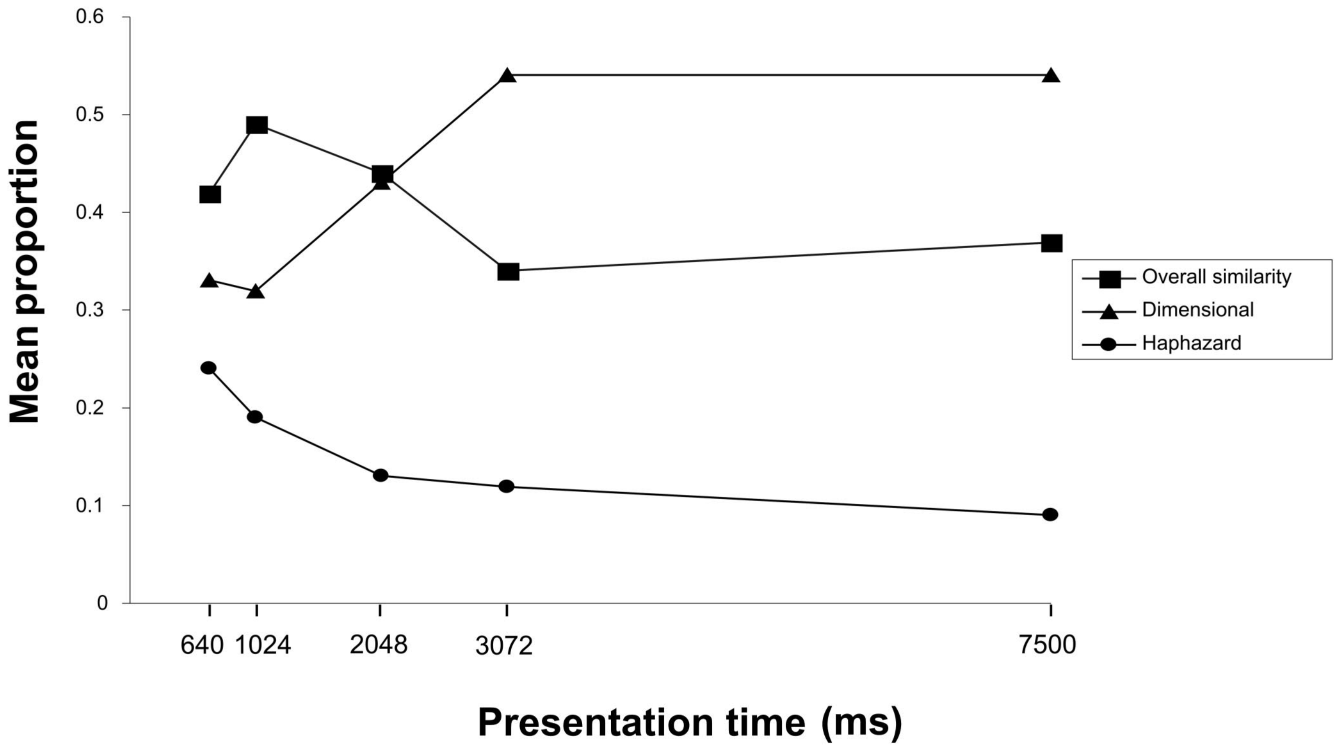

Figure 6. The proportion of overall similarity, dimensional, and haphazard responses for each condition in Experiment 5.

Results and Discussion

In a triad task, two of the stimuli (A and B; see Figure 1) are identical on one of the dimensions and very different on the other dimension. Choosing A and B as the stimuli that go together best is considered a dimensional response. Stimulus C is close but not identical to Stimulus B on both dimensions. Choosing B and C as the stimuli that go together best is considered an overall similarity response.[5] Grouping Stimuli A and C together is considered a haphazard response as it neither maximizes overall similarity nor makes use of the dimensional identity.

The mean proportions of overall similarity, dimensional, and haphazard responses are shown in Figure 6. A between-subjects ANOVA yielded a significant main effect of condition on overall similarity sorting, $F(4, 145) = 3.13$, $p < .02$, $\eta_p^2 = .08$. Trend analyses showed a significant linear trend, $F(1, 145) = 5.81$, $p < .02$; a significant cubic trend, $F(1, 145) = 5.53$, $p < .02$; but no significant quadratic trend ($p > .3$). Post hoc Fisher's LSD analyses revealed that there was a significant difference in overall similarity sorting between the 640-ms and the 1,024-ms conditions ($p < .01$), with participants in the 640-ms condition producing fewer overall similarity sorts than did those in the 1,024-ms condition. There was no significant difference between the 1,024-ms and 2,048-ms conditions in terms of overall similarity sorting, although the trend was for a decrease in similarity responding. There was, however, a significant decrease in overall similarity sorting in the 3,072-ms condition compared with in the 2,048-ms condition ($p < .05$). The difference in overall similarity sorting between the 3,072-ms condition and the 7,500-ms condition did not approach significance ($p > .5$).

Turning to dimensional responding, a between-subjects ANOVA yielded a significant main effect of condition, $F(4, 145) = 6.80$, $p < .001$, $\eta_p^2 = .16$. Trend analyses revealed that there was a significant linear trend, $F(1, 145) = 24.12$, $p < .001$, and a near significant cubic trend, $F(1, 145) = 3.05$, $p < .09$. The quadratic trend did not approach significance ($p > .8$). Post hoc Fisher's LSD analyses revealed no significant difference in dimensional responding between the 640-ms and the 1,024-ms conditions ($p > .5$). There was a significant increase in dimensional sorting between the 1,024-ms and the 2,048-ms conditions ($p < .02$) and a near significant increase in dimensional sorting between the 2,048-ms and 3,072-ms conditions ($p < .08$). The difference in dimensional responding between the 3,072-ms and 7,500-ms conditions did not approach significance ($p > .5$).

In terms of haphazard sorting, a between-subject ANOVA revealed that there was a significant main effect of condition, $F(4, 145) = 12.89$, $p < .001$, $\eta_p^2 = .26$. Trend analyses showed a

[5] Previous work has highlighted the potential ambiguity of interpreting overall similarity sorts (e.g., Thompson, 1994). For example, if a participant is presented with Stimuli 1–3 in Figure 5 and considers only the hull dimension, Stimuli 1 and 3 will be grouped together as they are more similar on the hull dimension. This will be classified as an overall similarity sort, even though only one dimension has been considered. One technique that arguably allows better interpretation of overall similarity sorts is latent class analysis (e.g., Raijmakers, Jansen, & van der Maas, 2004); one major disadvantage with this method, however, is that it requires extremely large sample sizes. We therefore used the same analyses as in earlier speeded triad classification studies (e.g., Ward, 1983) to enable better comparison between our work and previous findings.

significant linear trend, $F(1, 145) = 48.42$, $p < .001$, indicating that the level of haphazard sorting reduced with increasing time. There were no significant quadratic or cubic trends (both ps $>$ .15).

In general, the levels of overall similarity and dimensional responding are similar to those reported in previous triad studies (e.g., Ward, 1983; Ward et al., 1986), despite the somewhat different stimuli used in the current experiment. Admittedly, the haphazard level is, in general, slightly higher in our experiment compared with previous studies. This may at least partly be due to the fact that in our experiment, the stimuli disappeared before participants made a response, whereas in previous studies the stimuli were visually available until the end of the decision process.

The significant increase in overall similarity sorting between 640 ms and 1,024 ms is, as far as we are aware, the first demonstration that overall similarity sorting can rise with reducing time pressure in a triad task procedure. Nevertheless, this finding is in line with the assumptions of stochastic sampling models, such as the extended generalized context model (EGCM; e.g., Lamberts, 2000), which assumes that at very high time pressure there is insufficient time to process all dimensions sufficiently to observe consistent overall similarity responding. Our 640-ms condition probably represents a higher level of time pressure than used in previous studies (e.g., Ward, 1983; Ward et al., 1986), and we suggest that this difference may account for the difference between the current result and the results of these previous investigations.

From 1,024 ms onwards, the current results are in line with previous triad time pressure studies (e.g., Ward, 1983; Ward et al., 1986). That is, as time pressure was reduced, the level of overall similarity sorting decreased and dimensional responding increased. In line with previous work, we suggest that the overall similarity sorting observed at 1,024 ms is due to the stimuli being classified nonanalytically as integral blobs: Although both dimensions have been perceptually sampled, the analytic decomposition of the stimulus has not yet begun. Later on in the time course, participants analyze the stimuli back down into their component dimensions, and this permits dimensional responding. At this stage, overall similarity sorting decreases and dimensional responding rises.

Unlike in the match-to-standards procedure used in Experiments 1, 2, and 4, we failed to observe a subsequent rise in overall similarity sorting. Indeed, our pattern of results at 7,500 ms was virtually identical to the low time-pressure condition in Experiment 2 of Ward (1983), which required participants to spend at least 5 s on a response. The difference between the match-to-standards procedure and the triad task at these lowest levels of time pressure may plausibly be attributed to the differing task demands of the two procedures. We hypothesize that the overall similarity sorting observed under low time pressure in a match-to-standards procedure involves people matching identical dimensions in a highly analytic fashion. In contrast, overall similarity sorting under the triad procedure requires one to group two stimuli that do not share identical values; dimensional responding, however, does require one to group stimuli that share identical values. The presence of shared values for dimensional responding in the absence of identical values for overall similarity sorting may incline highly analytic participants toward a dimensional response in the triad task at low time pressure.

It is also important to note that in this triad task, and unlike in the match-to-standards task of Experiments 1, 2, and 4, increases in time pressure never led to a significant increase in dimensional responding—the reduction in overall similarity responding as presentation time drops from 1,024 ms to 640 ms is accompanied by a rise in haphazard sorting rather than a rise in dimensional sorting. We hypothesize that this difference may also be due to task demand differences between the triad and the match-to-standards procedures. More specifically, it seems likely that dimensional responding is a quicker strategy to apply in the match-to-standards task than in the triad task, for at least two reasons: First, in the match-to-standards task, participants can continuously use the same dimension throughout the sort, whereas in the triad task, participants are required to continually switch between dimensions in order to maintain dimensional responding. Second, it may be harder to make the necessary feature value discriminations for dimensional responding in the triad task than in the match-to-standards task. In the triad task, it appears relatively easy to detect the dimension mismatch between Stimuli A and B on one dimension (see Figure 1), but it is likely to take longer to establish that A and B are identical on the other dimension while also confirming that Stimuli B and C are not identical on either dimension. In the match-to-standards task, in contrast, consistent dimensional responding can be achieved by discriminating between the two values of any one of the four dimensions—a relatively simple, and presumably rapid, discrimination.[6] Hence, we hypothesize that in the match-to-standards procedure, very high levels of time pressure lead to participants adopting single-dimension classification, whereas in the triad procedure, participants continue to adopt an overall similarity approach, but their effectiveness at implementing this approach is reduced because there is insufficient time to sample both dimensions of all three stimuli.

The pattern of results from 1,024 ms onwards is in line with previous work (e.g., Ward, 1983; Ward et al., 1986) and showed dimensional responding rising with reducing time pressure. This is consistent with the idea that at lower time pressure participants analyze the stimuli down into their component parts and select the two stimuli that have an identical match on one of their dimensions.

In summary, Experiment 5 confirms the basic result of Experiment 4 within a triad procedure. Specifically, Experiment 5 confirms that the relationship between time pressure and the prevalence of overall similarity sorting is nonmonotonic and, hence, more complex than what has previously been observed (e.g., Ward, 1983; Ward et al. 1986). Presumably, it is the use of a larger range of time pressure conditions in the current study than in previous studies that has allowed this additional complexity to be observed. Nevertheless, the results from the triad task of Experiment 5 are far from identical to the results of the match-to-standards task in Experiment 4. Specifically, an increase in dimensional responding at very high levels of time pressure and an increase of overall similarity sorting at very low levels of time pressure were observed in Experiment 4 but not Experiment 5. These differences are plausibly accounted for by the differing task demands of the two procedures.

[6] We thank Robert Nosofsky for this explanation.

General Discussion

Previous research has demonstrated that the prevalence of overall similarity sorting is affected by a number of factors, including mode of stimulus presentation (Regehr & Brooks, 1995), spatial integrality of the stimuli (Milton & Wills, 2004), background knowledge (e.g., Spalding & Murphy, 1996), and inductive inference (Lassaline & Murphy, 1996). In the current research, we focused on the well-established finding that increased time pressure can increase the likelihood of overall similarity sorting in a free classification procedure (J. D. Smith & Kemler Nelson, 1984; Ward, 1983; Ward et al., 1986). This effect is often cited in support of the idea that overall similarity sorting is the result of a quick, primitive, nonanalytic process (J. D. Smith & Kemler Nelson, 1984; Ward, 1983).

In contrast, Experiments 1 and 2 of the current research demonstrated, across two different stimulus sets and two different methods for applying time pressure, that increased time pressure can reliably decrease the likelihood of overall similarity sorting. Although these findings are novel within the free classification procedure, one might reasonably predict their existence on the basis of related work on categorization with feedback (Lamberts, 2000; Ward & Scott, 1987). The results of Experiments 1 and 2 were also directly predicted by Milton and Wills (2004) on the basis of their dimensional summation hypothesis. This theory proposes that, under certain conditions, overall similarity sorting can be the result of a time-consuming, verbal, analytic process in which people consider each dimension in turn and place the stimulus into the category for which it has more characteristic features.

This idea received converging support from the results of Experiment 3, which introduced a concurrent load manipulation. According to our hypothesis, the overall similarity sorting observed in Experiments 1 and 2 was due to the use of an effortful, verbal, analytic strategy, and hence an additional prediction of our account is that the introduction of a concurrent load should lead to a reduction in the prevalence of overall similarity sorting in the match-to-standards procedure (which is the opposite of the result previously observed in the triad procedure; J. D. Smith & Kemler Nelson, 1984). The results of Experiment 3 supported our prediction.

The results of Experiments 1–3 may appear somewhat contrary to what might have been expected according to the previous well-established findings of J. D. Smith and Kemler Nelson (1984) and of Ward and colleagues (e.g., Ward, 1983; Ward et al. 1986). We considered two candidate explanations for this apparent discrepancy: differences in task demands and differences in the level of time pressure applied. The task demands explanation points to the fact that although both procedures contrast overall similarity responding with "dimensional" responding of some form, the complexity of the behavior required to emit responses that will be classified as dimensional differs in the two procedures. In the match-to-standards procedure, a simple one-dimension rule is easily applied. In contrast, dimensional responding in a triad procedure (the procedure largely used in previous studies) involves repeated switching between dimensions, which is likely to be time consuming. One possibility, therefore, is that the dimension-switching dimensional response in a triad task is more time consuming than an overall similarity response, whereas the simple, nonswitching, dimensional response in a match-to-standards task is less time consuming than an overall similarity response. Hence, both procedures simply demonstrate that people opt for the less time-consuming strategy when time is limited, and this explains the apparently opposing results found.

In contrast, the levels-of-time-pressure explanation points to the fact that although both sets of studies manipulate time pressure, the level of time pressure is arguably higher in previous studies than in Experiments 1 and 2. Hence, it is possible that moderate time pressure (as we hypothesize was present in our Experiments 1 and 2) leads to a reduction in overall similarity sorting whereas more severe time pressure (as we hypothesize was present in previous studies) leads to an increase in overall similarity sorting. These two explanations were explored in Experiments 4 and 5.

The results of Experiment 4 support the levels-of-time-pressure explanation. They show that within a match-to-standards procedure, the prevalence of overall similarity sorting rises from 64 ms to 256 ms; it then decreases from 256 ms to 384 ms, before rising again between 384 ms and 640 ms. In contrast, the presence of a nonmonotonic relationship between stimulus presentation time and prevalence of overall similarity sorting does not seem to naturally emerge from a task-demands explanation.

The results of Experiment 4 raise the interesting possibility that analytic and nonanalytic theories of overall similarity sorting are both correct but apply to different parts of the time course. For example, the results of Experiment 4 are consistent with the notion that Ward's (1983) integral-to-separable model applies relatively early in the time course, whereas Milton and Wills's (2004) dimensional summation hypothesis applies somewhat later on. These two theories might be integrated in the following way: Early in processing, stimuli are treated nonanalytically as integral "blobs," and this leads to overall similarity sorting (Ward's integral phase). Somewhat later on, analytic processing begins. Stimuli are intentionally analyzed into their constituent dimensions, and this permits dimensional sorting (Ward's separable phase). Milton and Wills's (2004) dimensional summation hypothesis concurs that at this stage, sorting will be single dimensional because time pressure is still relatively high, and so, although there is sufficient time to apply a dimensional rule, there is insufficient time to apply the more complex and time-consuming dimensional summation rule. However, as the time available for processing continues to increase, it becomes possible to apply the dimensional summation rule, and hence, the prevalence of overall similarity sorting begins to rise again.

What such an account fails to explain is why the prevalence of overall similarity sorting increases between 64-ms and 256-ms presentation times. We hypothesize that this effect is due to the early perceptual processes that lead to the formation of Ward's integral "blob." There is good evidence from studies of categorization with feedback (e.g., Lamberts & Freeman, 1999) that stimulus dimensions are stochastically sampled through time. Hence, at very short presentation intervals, only the most salient dimension is likely to have been sampled, and therefore, single-dimension sorting is the only systematic option open to participants.

In summary, we are proposing that stimulus dimensions are sampled stochastically over time to build up a stimulus representation. This stimulus representation is initially acted on by quick, nonanalytic processes that facilitate overall similarity sorting. However, given more time, participants become increasingly likely to analyze the stimulus back down into its constituent dimensions. Where time is still relatively short, participants will tend to apply only simple, single-dimension sorting rules to this analyzed representation, but as the amount of time available increases, the sorting rules they apply become more complex and can include dimensional summation, which leads to overall similarity sorting.

One possible criticism of this account is to ask why overall similarity sorting in Experiment 4 is not more prevalent at the stimulus presentation times that are being assumed to elicit nonanalytic overall similarity sorting (256 ms) and analytic overall similarity sorting (640 ms). Dealing with analytic overall similarity sorting first, we suggest that 640 ms is arguably still a fairly short stimulus presentation time for the application of a complex rule such as dimensional summation. As presentation time increases beyond 640 ms, one would expect dimensional summation to be increasingly applied. Cross-experiment comparisons of Experiments 2 and 4 largely support this idea: Although there is no reliable change in the prevalence of overall similarity sorting from 640 ms to 1,024 ms, overall similarity sorting is significantly more prevalent at 4,096 ms in Experiment 2 than it is at 640 ms in Experiment 4.

The question of why overall similarity sorting is not more prevalent in the 256-ms condition still remains. One possibility is that the peak of nonanalytic overall similarity sorting resides somewhere between the 64- and 256-ms presentation times. Locating this peak may present some practical challenges as one needs to find conditions under which there is sufficient time for all dimensions to have been perceptually sampled but where participants do not engage in the analysis of the stimulus back into its constituent dimensions.

Experiment 5 used multiple time conditions to investigate the time course of overall similarity sorting under the triad procedure. This enabled us to better elucidate the impact that the differing task demands of the match-to-standards and triad procedures have on overall similarity sorting. The results of Experiment 5 revealed that overall similarity rises between 640 and 1,024 ms and that this is followed by a subsequent fall at lower levels of time pressure. These results support the idea that the time course of overall similarity sorting follows a nonmonotonic pattern while also highlighting important differences in the task demands between the match-to-standards and triad procedures. We again relate the initial rise in overall similarity sorting to the assumptions of stochastic sampling models (e.g., Lamberts, 2000): At 640 ms, there is insufficient time to sample both dimensions to sort by overall similarity consistently. At 1,024 ms, there is now sufficient time to process both dimensions in a (presumably) nonanalytic fashion, and this leads to an elevation in overall similarity responding. Later on, the dimensions can be analyzed back down into their component dimensions, and this leads to a reduction in overall similarity sorting. Unlike in Experiment 4 (and also Experiments 1 and 2), however, we failed to observe a subsequent rise in overall similarity sorting at very low time pressure (7,500 ms). We suggest that this difference is due to differences in task demands between the match-to-standards and triad tasks (see the *Results and Discussion* of Experiment 5 for more details).

There are, however, important qualitative differences in the results across Experiment 4 and Experiment 5 that do not directly emerge from our account. In Experiment 4, single-dimension sorting was more prevalent than overall similarity sorting at the shorter time intervals, whereas at the longest time interval (4,096 ms), overall similarity sorting was more common. In Experiment 5, the opposite effect was observed: Overall similarity sorting was higher than single-dimension sorting at the shorter time intervals, and single-dimension sorting was more prevalent than overall similarity sorting at longer time intervals.

We believe that our account requires two additional assumptions in order to better capture this qualitative difference. The first assumption is that when dimension switching is required, dimensional responding will be more time consuming. This implies that under high time pressure, dimensional responding will be lower when dimension switching is necessary (as in the triad task) than when it is not (as in the match-to-standards task). This appears a plausible explanation for the greater prevalence of single-dimension sorting at high time pressure in the match-to-standards procedure than in the triad procedure.

The second assumption is the notion of valued similarity introduced by L. B. Smith (1989) in her influential model of perceptual classification. According to this theory, identity is treated as special, and identical dimensions are weighted higher in the classification decision. This suggests that when participants have sufficient time to process the dimensions analytically, the number of identity matches should influence sorting behavior. This is consistent with our findings. In the match-to-standards procedure, there are more identity matches for overall similarity sorting than for single-dimension sorting, and there is a corresponding greater prevalence of overall similarity sorting than of single-dimension sorting under low time pressure. For the triad task, there is one identity match for the dimensional response but none for the overall similarity response, and this leads to a greater level of dimensional sorting than of overall similarity sorting under low time pressure (where there is sufficient time to switch between dimensions).

Although the addition of these two assumptions should better account for the qualitative differences between the two procedures, this awaits verification through formal modeling. The current results present a challenge to existing models of categorization, and this makes formalizing our account an important next step in this research program.

Concluding Remarks

Multiprocess accounts of categorization have become increasingly popular in recent years (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; McLaren, Green, & Mackintosh, 1994). Our results seem to lend further support to this general contention, insofar as it seems unlikely that a single-process account would be fully able to account for the results we have presented. Additionally, multiprocess accounts of categorization often involve the notion of competition between a rule-based (possibly verbal) system and an associative (possibly implicit) system. Under certain assumptions, our account of the current experiments is also consistent with this idea. The first assumption (previously proposed by McLaren et al., 1994) is that the associative system is faster than the rule-based system and will therefore tend to win when time is limited. The second assumption is that the complexity of the rules produced by the rule system can be affected by the time available to it. Although this second assumption is perhaps not unreasonable, it is not generally included in formal expressions of multiprocess models.

As an alternative, it is possible that a modified version of a model such as EGCM may be able to account for our results. It is plausible, for instance, that people use different attentional or decisional weighting strategies depending on the processing time available, which a model such as EGCM could capture by varying such parameters with changes in processing time. Although such an account is speculative and requires verification through formal modeling, it may be that with such modifications a model such as EGCM may be able to account for our complex pattern of results.

In summary, our experiments provide support for the idea that overall similarity sorting can result from both nonanalytic and analytic processes depending on the particular task demands. As such, this work provides further evidence of the subtle factors that can influence the way people sort categories in the absence of feedback. In this general sense, our results add to previous work that has shown that stimulus presentation technique (Regehr & Brooks, 1995), spatial integration (Milton & Wills, 2004), background knowledge (e.g., Spalding & Murphy, 1996), and inductive inference (Lassaline & Murphy, 1996) all have a significant influence on the prevalence of overall similarity sorting.

References

Ahn, W. K. (1990). Effects of background knowledge on family resemblance sorting. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 149–156). Hillsdale, NJ: Erlbaum.

Ahn, W. K. (1991). Effects of background knowledge on family resemblance sorting and missing features. In *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 203–208). Hillsdale, NJ: Erlbaum.

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

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

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

Bradley, D. R., Bradley, T. D., McGrath, S. E., & Cutcomb, S. D. (1979). Type I error rate of the chi-square test of independence in R × C tables that have small expected frequencies. *Psychological Bulletin, 86,* 1290–1297.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

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

Cohen, A. L., & Nosofsky, R. M. (2003). An extension of the exemplar-based random walk model to separable-dimension stimuli. *Journal of Mathematical Psychology, 47,* 150–165.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics, 12,* 108–116.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury Press.

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

Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition, 27,* 699–712.

Kemler Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23,* 734–759.

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

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

Lamberts, K. (2002). Feature sampling in categorization and recognition of objects. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 141–154.

Lamberts, K., & Freeman, R. P. J. (1999). Categorization of briefly presented objects. *Psychological Research, 62,* 107–117.

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

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review, 79,* 410–419.

McLaren, I. P. L., Green, R., & Mackintosh, N. J. (1994). Animal learning and the explicit/implicit distinction: Or why what we think of as explicit for us can be implicit for them. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 313–332). London: Academic Press.

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

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

Overall, J. E. (1980). Power of chi-square tests for 2 × 2 contingency tables with small expected frequencies. *Psychological Bulletin, 87,* 132–135.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26,* 303–343.

Proctor, R. W., & Fisicaro, S. A. (1977). Time, capacity, and selection between perceptual attributes. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 337–348.

Psychological Software Tools. (2002). E-prime [Computer software]. Retrieved from http://www.pstnet.com

Raijmakers, M. E. J., Jansen, B. R. J., & van der Maas, H. L. J. (2004). Rules and development in triad classification task performance. *Developmental Review, 24,* 289–321.

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.

Ryle, G. (1951). Thinking and language. *Proceedings of the Aristotelian Society, 25*(Suppl.), 65–82.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

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

Smith, J. D., & Shapiro, J. H. (1989). The occurrence of holistic categorization. *Journal of Memory and Language, 28,* 386–399.

Smith, L. B. (1979). Perceptual development and category generalization. *Child Development, 50,* 705–715.

Smith, L. B. (1989). A model of perceptual classification in children and adults. *Psychological Review, 96,* 125–144.

Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology, 10,* 502–532.

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.

Thompson, L. A. (1994). Dimensional strategies dominate perceptual classification. *Child Development, 65,* 1627–1645.

Thorpe, S. J., & Imbert, M. (1989). Biological constraints on connectionist modeling. In R. Pfeifer, Z. Schreter, F. Fogelman-Soulie, & L. Steels (Eds.), *Connectionism in perspective* (pp. 63–92). Amsterdam: Elsevier.

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.

Ward, T. B., & Scott, J. (1987). Analytic and holistic modes of learning family-resemblance concepts. *Memory & Cognition, 15,* 42–54.

Wittgenstein, L. (1958). *Philosophical investigations* (2nd ed.). Oxford, United Kingdom: Blackwell.

Zeithamova, D., & Maddox, W. T. (2006). Dual task interference in perceptual category learning. *Memory & Cognition, 34,* 387–398.

Received June 2, 2006
Revision received August 23, 2007
Accepted August 27, 2007 ■