Is Overall Similarity Classification Less Effortful Than Single-Dimension Classification?

Andy J. Wills, Fraser Milton, Christopher A. Longmore, Sarah Hester and Jo Robinson

University of Exeter

Author Note

Andy J. Wills, Fraser Milton, Christopher A. Longmore, Sarah Hester and Jo Robinson, Psychology, University of Exeter.

Andy J. Wills and Christopher A. Longmore are now at the School of Psychology, Plymouth University.

This research was supported by ESRC grant RES-000-22-1779 awarded to Andy J. Wills and ESRC grant PTA-026-27-1256 and a Great Western Research Fellowship awarded to Fraser Milton.

Andy J. Wills, School of Psychology, Plymouth University, Drake Circus, Plymouth. PL4 8AA. United Kingdom. Email: andy@willslab.co.uk

Abstract

It is sometimes argued that the implementation of an overall similarity classification is less effortful than the implementation of a single-dimension classification. In the current article, we argue that the evidence securely in support of this view is limited, and report additional evidence in support of the opposite proposition – overall similarity classification is *more* effortful than single-dimension classification. Using a match-to-standards procedure, Experiments 1A, 1B and 2 demonstrate that concurrent load reduces the prevalence of overall similarity classification, and that this effect is robust to changes in the concurrent load task employed, the level of time pressure experienced, and the short-term memory requirements of the classification task. Experiment 3 demonstrates that participants who produced overall similarity classifications from the outset have larger working memory capacities than those who produced single-dimension classifications initially, and Experiment 4 demonstrates that instructions to respond meticulously increase the prevalence of overall similarity classification.

Is Overall Similarity Classification Less Effortful Than Single-Dimension Classification?

In a seminal article, Brooks (1978) argued for two different processes of categorization. In *analytic* categorization, the participant separates aspects of the stimulus and evaluates their ability to predict category membership. This process of analysis, Brooks assumed, will typically lead to a subset of the stimulus attributes controlling responding. In contrast, *nonanalytic* categorization is the process of predicting category membership on the basis of overall similarity to known examples – a process that results in all stimulus attributes having some control over responding. Brooks hypothesized that nonanalytic categorization would be more likely to occur where cognitive resources were limited. This is a striking hypothesis because it assumes that a categorization process that employs all the information in the stimulus (overall similarity) is less effortful than a categorization process that employs a subset of that information (analytic, or “rule-based”, categorization). Another way of phrasing this hypothesis is to say that overall similarity classification is considered to be less deliberative than, for example, single-dimension classification. For brevity, we will describe this as the *less-is-more* hypothesis – for example, less time spent categorizing objects results in more information from those objects having control over responding. We will contrast this with the *more-is-more* hypothesis – for example, more time spent categorizing objects results in more information from those objects having control over responding. Although categorization is the focus of the current article, we note in passing that less-is-more hypotheses are to be found in a number of areas of psychology, including reasoning (Sloman, 1996) and complex decision making (Dijksterhuis, Bos, Nordgren & van Baaren, 2006).

Following the publication of Brooks (1978), a body of experimental evidence accumulated that seemed to support a *less-is-more* hypothesis in the context of categorization. For example, Ward (1983) demonstrated that reducing the time available for categorization increased the prevalence of spontaneous overall similarity classification, and reduced the prevalence of spontaneous dimensionally-based classification. Smith and Kemler Nelson (1984) replicated this finding, and further demonstrated that concurrent load, and instructions to respond impressionistically, increased overall similarity responding. Kemler Nelson (1984), using a category-learning task, demonstrated that those who learned the category structure incidentally produced more overall similarity classifications than those who learned the category structure intentionally. Smith and Shapiro (1989) replicated the Kemler Nelson (1984) findings, and also demonstrated that concurrent load increased overall similarity classification in this procedure.

Of course, not all evidence pointed in the same direction. In particular, Ward and Scott (1987) reported that classifying by overall similarity took longer than classifying by a single-attribute-plus-exception strategy. A few cases were also reported where time pressure failed to have a significant effect on the prevalence of overall similarity classification (Smith & Kemler Nelson, 1984, Experiment 4; Smith & Shapiro, 1989, Experiments 2-3). Smith and Shapiro (1989), acknowledging the results of Ward and Scott (1987), and the absence of any time pressure effect in their own studies, hypothesized, “speed may promote extremely narrow attention as adults choose any attribute in a cognitive storm” (p. 394-395). Nevertheless, by the end of the 1990s the *less-is-more* hypothesis seemed to become an accepted part of the theoretical landscape in categorization. For example, Goldstone and Barsalou (1998) stated, “evidence suggests that in many situations, it is easier for people to base similarity and categorization judgments on more, rather than fewer, properties” (p.239-

240). In the same year, Ashby and colleagues published the influential COVIS model (Ashby, Alfonso-Reese, Turken & Waldron, 1998). At the heart of this model is the assumption that there are two competing categorization processes. In a similar vein to Brooks (1978), the verbal process in COVIS generates and evaluates verbalisable hypotheses, whilst the implicit process responds on the basis of overall similarity. The verbal process is considered to be disrupted by a concurrent verbal task, whilst the implicit process is not. In the paper that introduced COVIS, the results of Ward (1983), and of Smith and Kemler Nelson (1984), were presented as evidence in support of the COVIS framework (Ashby et al., 1998, p. 457).

Over the next decade, new research provided further evidence that seemed in support of the *less-is-more* hypothesis. Waldron and Ashby (2001) demonstrated that concurrent load retarded the acquisition of a single-dimension category structure more than it retarded the acquisition of a multidimensional category structure. Zeithamova and Maddox (2006) provided a similar demonstration with a different stimulus set. Tracy et al. (2003) reported more frontal lobe activation in single-dimension categorization than in overall similarity categorization (see also Patalano, Smith, Jonides & Koeppe, 2001). Couchman, Coutinho and Smith (2010) reported that whilst adult humans favoured single-dimension responding in the Kemler Nelson (1984) procedure, monkeys favoured overall similarity responding.

As with the earlier research, not all the evidence pointed in the same direction. Milton, Longmore, and Wills (2008) reported multiple experiments in which increased time pressure resulted in a *reduction* in overall similarity classification in a spontaneous classification task. Tharp and Pickering (2009) reported that 44% of their participants reached criterion on the Waldron and Ashby (2001) multidimensional category structure by using a single-dimensional rule. Newell, Dunn and Kalish (2010) failed to replicate the results of

Zeithamova and Maddox (2006) across three experiments. Milton, Wills and Hodgson (2009) reported that overall similarity classification in a spontaneous classification task led to more frontal lobe activity than did single-dimension responding. Wills et al. (2009) compared pigeons, squirrels and adult humans under closely matched conditions but failed to find any species differences in the relative prevalence of overall similarity versus single-dimension classification.

Despite these further anomalies, the *less-is-more* hypothesis seems to have largely survived in contemporary research. For example, Couchman and colleagues recently concluded, "family resemblance categorization is the default mode of approach to tasks when full executive functioning is underdeveloped, inhibited, or lacking for any reason" (Couchman et al., 2010, p.180). Family resemblance is a term approximately synonymous with the term overall similarity. In the current article, we briefly review the three main methodologies that have been most widely considered to support the *less-is-more* hypothesis that overall similarity classification requires fewer cognitive resources than dimensional classification. Our conclusion is that none of these three procedures is well suited to distinguishing between the *less-is-more* and the *more-is-more* hypothesis. We then present five new experiments, using a procedure better equipped to do this, that support the case that the implementation of a overall similarity classification requires more cognitive resources than the implementation of a single-dimension classification – consistent with a *more-is-more* hypothesis.

Triad procedure

In the triad procedure, used most notably by Ward (1983) and by Smith and Kemler Nelson (1984), stimulus triplets are presented. Three stimuli, whose relationship to each other is illustrated in Figure 1, are presented simultaneously and participants are asked to decide which two stimuli go together best. Stimuli B and C are similar on both stimulus dimensions, but not identical on either, while stimuli A and B are identical on one stimulus dimension but quite dissimilar on the other. Three responses are possible – an AB response ("A and B go together best"), a BC response or an AC response. Time pressure, concurrent load, impulsivity, and instructions to respond impressionistically, all increase BC responses and decrease AB responses (Smith and Kemler Nelson, 1984; Ward, 1983; Ward, Foley & Cole, 1986). AB responding is typically described as "dimensional" responding and BC responding is typically described as "overall similarity" responding, hence leading to the claim that overall similarity (BC) responding increases as cognitive resources decrease.

One reason that the triad procedure is ill suited to testing a *less-is-more* hypothesis is that consistent AB ("dimensional") responding requires that the participant consider both stimulus dimensions on every trial. This is because the dimension on which A and B are identical varies unpredictably from trial to trial (see Figure 1), and so a consistent AB responder cannot decide in advance of stimulus onset to only attend to one of the stimulus dimensions. Consistent BC responding also requires consideration of both stimulus dimensions on every trial (irrespective of whether one believes that consideration to take the form of an analytic strategy or direct access to similarity relations through holistic "blobs"). Hence both consistent "overall similarity" (BC) responding and consistent "dimensional" (AB) responding requires consideration of all the relevant stimulus information on every trial. It is therefore not the case that overall similarity responding requires consideration of more of the available stimulus information than dimensional responding in the triad task, and hence the triad task is not well suited to testing a less-is-more hypothesis.

Ashby-Maddox procedure

Ashby and colleagues report a range of evidence that single-dimension (rule-based) categories are learned by a verbal explicit system, whilst certain multidimensional (information integration) category structures are learned by a non-verbal procedural system (see Ashby & Maddox, 2005, for a review). As illustrated in Figure 2, the most common comparison is between a single-dimension category (Figure 2A) and a multidimensional category structure that is a 45-degree rotation of the single-dimension structure (Figure 2B). Zeithamova and Maddox (2006) reported that concurrent load impaired the acquisition of the single-dimensional structure more than the acquisition of the multidimensional structure (although see Newell et al., 2010). On the basis of such results, one might wish to argue that single-dimension classification is more effortful than multidimensional classification, a form of *less-is-more* hypothesis. However, recent work (Filoteo, Lauritzen & Maddox, 2010; Maddox, Pacheco, Reeves, Zhu & Schnyer, 2010l; although see Newell, Moore, Wills & Milton, submitted) suggests that concurrent load also impairs the acquisition of a conjunctive categorization (Figure 2C) more than it impairs the acquisition of an information-integration categorization (Figure 2B). The conjunctive and information-integration structures both require the participants to be sensitive to both stimulus dimensions. Hence, it is presumably some factor other than the dimensionality of the categorization problem that is crucial in these studies. The most likely candidate (and the one endorsed by Ashby and colleagues) is the verbalisability of the category structure. The structures in Figures 2A and 2C are quite easy to verbalize, the structure in Figure 2B, less so.

In summary, work employing the Ashby-Maddox procedure may elucidate an important difference between the processes underlying the acquisition of easy-to-verbalise versus hard-to-verbalise categories, but these results do not speak directly to the question of whether single-dimension classification is more or less effortful than overall similarity classification.

Criterial-attribute procedure

The criterial-attribute procedure (Kemler Nelson, 1984; Smith & Shapiro, 1989) involves supervised categorization of a set of stimuli (the abstract structure of which is illustrated in Figure 3). During the training phase of this procedure, it is possible to determine the category membership of an item with 100% accuracy in at least two ways. One way is to respond on the basis of the criterial attribute (Dimension 1 in Figure 3), as the criterial attribute is perfectly correlated with category membership. Another way to achieve 100% accuracy during training is to aggregate information from at least three of the stimulus dimensions. This second method of achieving perfect accuracy during training is typically referred to as a family resemblance (overall similarity) strategy. A subsequent test phase places the criterial-attribute strategy into competition with the overall similarity strategy - the critical test stimuli (shown in Figure 3) indicate one category on the basis of the criterial attribute, but the other category on the basis of overall similarity.

Kemler Nelson (1984) found that adults were more likely to respond to these test stimuli on the basis of overall similarity if they were trained under incidental conditions than if they were trained under standard intentional conditions. The term “incidental” here refers to the fact that, although category membership information was available on all trials, participants were not told that the purpose of the training phase was to learn categories. In Kemler Nelson (1984), participants were instead asked whether they had seen each of the presented stimuli before. Ward and Scott (1987) failed to replicate the effect, but given that their incidental training procedure directly drew attention to the presence of two categories, this is perhaps not unexpected (see also Kemler Nelson, 1988, for other possible reasons for the difference between these two studies). Smith and Shapiro (1989) replicated Kemler Nelson’s incidental learning effect, and further demonstrated that concurrent load during training also increased the incidence of overall similarity classification. Smith and Shapiro (1989) also manipulated time pressure, but found no significant effects of this manipulation. In all these experiments, the test phase was conducted under intentional conditions and in the absence of time pressure and concurrent load.

Although these results are generally taken to be in support of a less-is-more hypothesis, they are also consistent with a more-is-more account. What any theoretical account has to explain is (1) why incidental training conditions lead to more overall similarity classification in an intentional test than intentional training conditions, and (2) why concurrent load during training leads to more overall similarity classification in a full-attention test than full attention during training. In the former case (incidental training) one explanation is that the incidental training conditions directly encourage an even distribution of attention across the stimulus dimensions in a way that the intentional condition does not. Kemler Nelson (1984) used an old-new recognition task as her incidental training condition, and one cannot perform well on such a task by ignoring three out of the four stimulus dimensions. Smith and Shapiro (1989) used reading out the stimuli (which were pronounceable nonwords) as their incidental training condition. Again, this is a task that encourages participants to consider the whole stimulus, rather than focus their attention narrowly on one dimension (one letter) of it. In contrast, the task in the intentional training condition is to predict category membership, which of course can be done on the basis of a single dimension in the criterial-attribute procedure.

Turning to the effects of concurrent load during training, a sufficient explanation is that the search for the single-dimension hypothesis that allows the participant to meet the training criterion is in itself effortful, and that concurrent load disrupts this effortful process. Note that, in the criterial-attribute task, any one of the non-criterial dimensions can support 75% accuracy - above chance, but below the learning criterion of the task. This may make it particularly challenging to select the single one-dimensional hypothesis that permits 100% accuracy. Having been unable to identify the criterial attribute, participants may resort to alternative approaches such as classifying by overall similarity, despite the additional effort this requires to implement.

In summary, the *more-is-more* hypothesis can accommodate the results of criterial-attribute tasks by assuming that whilst the implementation of single-dimension classification is low effort, the discovery of the criterial attribute is high effort. Consistent with this hypothesis, Smith, Tracy and Murray (1993) found that depressed participants were impaired relative to controls on learning a criterial-attribute categorization, but not on learning an overall similarity categorization. They also found that depressed participants were more likely than controls to employ single-attribute responding in an overall similarity categorization task. Under the hypothesis that depression reduces the effort expended on categorization tasks, these data are consistent with the idea that discovering the appropriate single-dimension hypothesis is effortful in a criterial-attribute task, but that implementing a single-dimension classification is less effortful than implementing an overall similarity classification. It is the implementation of single-dimension and overall similarity classifications that is the focus of the current article.

Match-to-standards procedure

The purpose of the current article is to further investigate whether the implementation of an overall similarity classification strategy is more or less requiring of cognitive resources than the implementation of a single-dimension classification strategy. We have argued that the triad procedure (Figure 1) is not ideally suited to this investigation because both consistent overall similarity (BC) responding and consistent dimensional (AB) responding require use of all the relevant stimulus information on every trial. We have further argued that the Ashby-Maddox procedure, at least in its standard form (Figures 2A and 2B), probably assesses a difference in the verbalisability of the optimal classification rule, which is an interesting question in itself, but different to the one we are currently asking. And we have also argued that the criterial-attribute procedure (Figure 3) may reflect, not so much the effort of implementing a single-dimension versus an overall similarity classification, but the effort involved in discovering the one single-dimension rule that meets the training criterion amongst four possibilities, all of which are at least 75% adequate.

What is required, then, is a procedure where both single-dimension and overall similarity classifications are potentially verbalisable (it seems hard to design a procedure in which a single-dimension classification would not be verbalisable), and where the difficulty of implementing a single-dimension classification is not confounded with the difficulty of finding the one single-dimension classification that perfectly predicts feedback. The latter requirement suggests the use of a spontaneous classification procedure (i.e. one where participants classify as they see fit, rather than in response to experimenter-provided feedback), but we have already argued that one of the most commonly employed spontaneous classification procedures – the triad task – is not ideally suited to such an investigation. We believe that the match-to-standards task provides a less interpretatively ambiguous alternative to the triad task.

The match-to-standards procedure was introduced by Regehr and Brooks (1995) as a means of increasing the prevalence of overall similarity classification of novel stimuli, relative to the more commonly employed array sort procedure, in which single-dimension classification dominates (Medin, Wattenmaker, & Hampson,1987). In line with the findings of Milton et al. (2008), Regehr and Brooks observed that reaction times were longer for overall similarity classification than single-dimension classification in this procedure, although they did not publish these observations (Brooks, personal communication, 20 October 2009).

In the match-to-standards procedure, participants sequentially free classify each of a series of target stimuli as belonging to one of two categories. The two categories are represented by two *standards* – that is, two stimuli that appear on each trial. The two standards differ from each other on all variable stimulus dimensions. For example, in Experiment 1, the two standards are as shown in Figure 4A. In the current experiments, there are 10 distinct to-be-classified stimuli, with the abstract structure shown in Table 1. In some respects, the match-to-standards procedure is similar to the triad procedure, because each trial involves deciding which two of the three stimuli go together best (although, unlike the triad task, the option of saying that the two least similar stimuli - the two standards in the match-to-standards procedure - go together is not available). Also, in the match-to-standards procedure, each participants' classification strategy for a particular block is determined over 10 trials (rather than independently for each trial, as in the triad task). Both single-dimension and overall similarity classifications are verbalisable in the match-to-standards procedure.

Indeed, in the standard version of the procedure, the participant is asked to describe their classification strategy immediately after classifying the 10 items, and there is a high correspondence between the objective classification produced and the participant's description of that classification (Milton & Wills, 2009).

The current studies

The experiments reported in the current article are a further investigation of the opposing predictions of the more-is-more and less-is-more hypotheses of overall similarity classification, using the match-to-standards procedure. Milton et al. (2008) reported that concurrent load decreased the prevalence of overall similarity sorting in the match-to-standards procedure, which is consistent with a more-is-more hypothesis. However, Milton et al. (2008) employed a concurrent load task with substantial short-term memory requirements but limited operational load (6-item digit load), whilst previous studies consistent with a less-is-more hypothesis have employed concurrent load tasks with substantial operational load requirements but minimal short-term memory requirements (e.g. counting backwards, Smith & Kemler Nelson, 1984, Smith & Shapiro, 1989). Perhaps operational load is particularly effective in inducing the sort of holistic, fall-back mode of cognition assumed by the less-is-more hypothesis? Experiment 1A investigates this possibility by employing an operational load task in the Milton et al. procedure.

Second, previous studies consistent with a less-is-more hypothesis tend to conduct concurrent load studies in a context where participants are expected to complete their classifications within a specified time limit (Smith & Kemler Nelson, 1984; Smith & Shapiro, 1989?). Perhaps it is the combination of time pressure and concurrent load that is particularly effective in inducing a holistic mindset? Experiment 1B investigates this possibility by introducing time pressure into the Experiment 1A procedure.

Third, some studies supportive of a less-is-more hypothesis use a simultaneous stimulus presentation procedure (e.g. Smith & Kemler Nelson, 1984), whilst Milton et al. used a sequential procedure. The Milton et al. procedure thus incorporates a memory component that studies consistent with the less-is-more hypothesis often lack (e.g., Smith & Kemler Nelson, 1984; Ward, 1983; Ward et al., 1986). Perhaps the memory requirement leads participants to approach the stimuli in the Milton et al. procedure in a particularly analytic, verbal manner? Experiment 2 modifies the Experiment 1A procedure such that the two standards and the to-be-classified stimulus are presented simultaneously. Experiment 2 also removes the requirement to describe one's classification strategy at the end of each 10-item block – another feature of the Milton et al. procedure that might be considered to encourage a particularly verbal, analytic mindset.

Experiments 3 and 4 seek convergent evidence using two further ways of manipulating the availability or utilization of cognitive resources –individual differences in a measure of working memory capacity (Experiment 3), and instructions to respond meticulously (Experiment 4).

Experiment 1A

Milton et al. (2008), whose results were consistent with a more-is-more hypothesis, employed a concurrent load task with substantial short-term memory requirements but limited operational load. In contrast, previous studies consistent with a less-is-more hypothesis have employed concurrent load tasks with substantial operational load requirements but minimal short-term memory requirements. Perhaps operational load is particularly effective in inducing the sort of holistic, fall-back mode of cognition assumed by the less-is-more hypothesis? In Experiments 1A-1B, we investigated the effect of an operational load on the prevalence of overall similarity classification in the match-to-standards procedure.

Method

Participants

Forty-two participants from the University of Exeter took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision.

Materials

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. These prototypes were constructed by taking all the positive values on the dimensions for one of the stimuli (1, 1, 1, 1) and all of the zero values on the dimensions (0, 0, 0, 0) for the other category. The rest of the stimuli (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 between-group differences. The stimuli were silhouette lamps, and are shown in Figure 4A. Each lamp was composed of four dimensions (top, lampshade, stand and base) that were all coloured black. Each dimension could contain straight or curvy edges. For all participants, one prototype was composed of all the curved features, while the other was composed of all the straight features.

Apparatus

The stimuli were presented on a 17" CRT monitor, set to a resolution of 800x600 pixels and a colour depth of 16-bits per pixel using the E-Prime testing software. Responses were made using a standard keyboard.

Procedure

Participants were randomly allocated to either the load or no load conditions and were tested individually in a quiet cubicle. The participants sat in front of the computer screen at a distance of approximately 50cms and were given written instructions on the screen before the experiment began.

We used a slight variation of the match-to-standards task developed by Regehr and Brooks (1995); the task is a computer-based version of the Milton and Wills (2004) task, and has previously been used by Milton et al. (2008). The trial procedure in the no-load condition was as follows. At the start of each trial, participants were presented with the prototypes of category A and category B and were allowed to examine the prototypes for as long as they wished. Upon pressing the spacebar the screen went blank for 2480ms (the same period as employed in Milton et al., 2008), after which a fixation cross was presented for 500ms followed by a blank screen for 500ms. The stimulus to be classified was then displayed for 1500ms and was immediately followed by a grey mask that remained on the screen until the participant made a response. Participants were required to press the "C" key (which was labelled "A") to indicate the stimulus belonged to category A or the "M" key (which was labelled "B") to indicate the stimulus belonged to category B. They then pressed the space bar to move on to the next trial. A total of 120 trials were presented (in 12 blocks of 10 trials). In each block, each stimulus in the set (Table 1) was presented once in a random order. At the end of each block, participants were asked to provide a written description of their categorization strategy in the preceding block.

The procedure for the load condition was identical to that in the no-load condition, except that on each trial numbers were presented over headphones from the moment the prototypes disappeared until the moment the participant made a categorization response. Hence, participants were able to study the prototypes in silence, but numbers were presented during the rest of the trial. The numbers were voice-synthesized, ranging between 11 and 99. Each utterance of a number was of approximately 1s duration, and there was a 200ms pause between the end of one utterance and the beginning of the next. Participants were instructed to mentally keep count of the number of even numbers they heard and to focus primarily upon this task. At the end of each block, participants were required to indicate how many even numbers they heard in that block. They were given immediate feedback on their performance on the numbers task, and were encouraged to perform well on this task. As in the no-load condition, participants were required to provide a written account of their sorting strategy in each block upon its completion.

Dependent measure

Each participant was classified as having produced one of the sort types described below in each of the 12 blocks of the experiment. These sort types are very similar to those employed by Regehr and Brooks (1995) and are identical to those employed by Milton and Wills (2004) and Milton et al. (2008).

A *uni-dimensional* sort is based on a single dimension of the stimulus. It does not matter which dimension is 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 has to describe their sort as being based on a single dimension.

Participants were considered to have produced a *one-away uni-dimensional* 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 the other item was placed into the wrong category.

An *overall similarity sort*, also commonly known as a "*family resemblance*" sort (Medin et al., 1987), has a structure identical to that shown in Table 1. In order to receive this classification, the participant 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 placing 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 uni-dimensional sort with the exception that the error occurred in a sort that was otherwise 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 sorts described above. The correspondence between the classification produced by a participant and their verbal description of the sort they have produced is high in this procedure (Milton & Wills, 2009). The verbal descriptions were classified by the authors. We also analysed the data ignoring the verbal descriptions, classifying solely on the basis of the objective sort pattern produced. These two analysis methods did not differ on the ordinal pattern of results, nor the pattern of significance, observed, and produced similar sort prevalence values. For brevity, only the analyses combining objective and subjective information are reported. The combined method was chosen for consistency with all previous reports of the match-to-standards procedure.

Results and Discussion

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

The proportion of overall similarity sorts was significantly lower for participants who completed the classification task with a load task (mean = .14; s.d. = .20) than for participants who sorted with no load task (mean = .63; s.d. = .41); $t(40) = -4.88$, $p < .0005$. Other sorts were significantly more prevalent in the presence of the load task (mean = .35; s.d. = .30) than in its absence (mean = .03, s.d. = .07); $t(40) = 4.645$, $p < .0005$. The prevalence of uni-dimensional sorts in the presence of load (mean = .52, s.d. = .41) and in its absence (mean = .34, s.d. = .41) did not differ significantly, $t(40) = 1.38$, $p = .18$.

The results of Experiment 1A are consistent with the idea that, under load conditions, participants have fewer cognitive resources available for categorization and thus have difficulty in implementing an effortful overall similarity strategy. We assume, however, that participants are still able to implement the less cognitively demanding uni-dimensional strategy. The significant increase in Other sorts seems consistent with the idea that the concurrent load task is interfering with the classification process to such an extent that participants classify more haphazardly. The fact that the increase in Other sorts is accompanied by a reliable decrease in overall similarity sorts but no reliable effect on uni-dimensional sorts is in line with the idea that overall similarity sorting is the more resource-demanding sort type, and hence the more prone to interference. The failure to detect a significant increase in uni-dimensional sorting might be taken to indicate that participants experienced difficulty in changing their preferred classification strategy in response to a concurrent load. Those who preferred a uni-dimensional strategy were able to perform such a strategy under concurrent load, whilst some of those who preferred an overall similarity strategy could not perform this consistently under a concurrent load and hence produced Other sorts. The question of what might underlie such individual differences is returned to in Experiment 3.

Experiment 1B

Previous studies consistent with a less-is-more hypothesis have conducted concurrent load experiments in a context where participants are expected to complete their classifications within a specified time limit. For example, Smith and Kemler Nelson's (1984) concurrent load experiment instructed participants to complete the experiment within 144s (for 32 responses). Participants in the control condition could comply with this instruction, whilst participants in the concurrent load condition exceeded the time limit (Smith & Kemler Nelson, 1984, Table 2), suggesting that participants in the concurrent load condition were under substantial time pressure. Perhaps the combination of time pressure and concurrent load is particularly effective in inducing an analytic, holistic mindset? In Experiment 1B, we substantially increased time pressure in our procedure by presenting the target stimulus for just 300ms. The results of Milton et al. (2008) indicate that a presentation time of around 300ms produces significant time pressure in a match-to-standards task.

Method

Participants, Materials and Apparatus

Forty-two participants from the University of Exeter took part in the experiment in return for course credit or payment. No participant contributed to more than one experiment in this paper. All participants had normal or corrected to normal vision. Materials and apparatus were identical to those used in Experiment 1A.

Procedure

The procedure was identical to that used in Experiment 1A except that, in both the no load and load conditions, each to-be-classified lamp was presented for 300ms.

Results and Discussion

The proportion of overall similarity sorts was significantly lower in the presence of a concurrent load (mean = .09, s.d. = .24) than in its absence (mean = .38, s.d. =.38), $t(40) = -2.99$, $p < .01$. Other sorts were significantly more prevalent in the presence of the load task (mean = .29, s.d. = .25) than in its absence (mean = .13, s.d. = .13), $t(40) = 2.68$, $p < .025$.

The prevalence of uni-dimensional sorting in the presence of load (mean = .62, s.d. = .34) and in its absence (mean = .50, s.d. = .36) did not differ significantly, $t(40) = 1.18$, $p = .25$.

The results of Experiment 1B, like those of Experiment 1A, are consistent with the idea that, under load conditions, participants have fewer cognitive resources available for categorization and thus have difficulty in implementing an effortful overall similarity strategy. Experiment 1B adds to the findings of Experiment 1A in that this conclusion still holds when participants are under substantial time pressure. The time pressure in Experiment 1B does seem to be substantial, as evidenced by the lower proportion of overall similarity sorts in the control condition of Experiment 1B, relative to Experiment 1A. This is the expected effect of time pressure in a match-to-standards procedure (Milton et al., 2008).

Experiment 2

Some previous studies supportive of a less-is-more hypothesis have used a simultaneous presentation procedure. For example, in the triad procedure (Smith & Kemler Nelson 1984), each trial involves the simultaneous presentation of the three stimuli upon which the participant is asked to make their decision. In contrast, the match-to-standards procedure, at least as implemented in Experiments 1A and 1B, presents the two standards, removes them, waits for about 2.5 seconds, and then presents the to-be-classified stimulus. Hence, Experiments 1A and 1B contain a memory component that the triad task lacks. The presence of a memory component might be argued to induce a verbal, analytic mindset in participants (cf. Wattenmaker, 1992 - overall similarity classification is less prevalent when stimuli have been memorized than when they are visually present). Experiment 2 addresses this possibility by removing the memory component of Experiment 1. It does this by presenting the two standards and the to-be-classified stimulus simultaneously on each trial. It also removes the request for participants to describe their sort strategy at the end of each block, another factor that might be considered to induce a verbal, analytic mindset in Experiments 1A and 1B. At the very end of the task, participants received an unheralded instruction to describe the sort strategy or strategies they had employed, in order to informally assess the correspondence between participants' sort strategies and their description of those strategies.

Method

Participants

Thirty-six participants from the University of Exeter took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision.

Materials and Apparatus

As in Experiments 1A and 1B, the abstract stimulus structure employed was as shown in Table 1. The stimuli were line drawings of boats modelled on Lamberts (1998) and are shown in Figure 4B. The four dimensions on which each boat could vary were; shape of the flag at the top of the mast (rectangular or triangular), size of the main sail (large or small), shape of the porthole (circle or diamond) and the length of the base of the hull (long or short). The apparatus was identical to that used in Experiments 1A and 1B.

Procedure

Each trial began with the presentation of three stimuli – the prototypes for category A and category B at the top of the screen, with the to-be-classified stimulus presented below the prototypes, mid-way between them. All three stimuli remained on the screen until the participant placed the to-be-classified stimulus into either category A (by pressing the 'c' key on the keyboard) or category B (by pressing the 'm' key). Each trial was followed by a 1500ms inter-trial interval (blank screen). As in Experiments 1A and 1B, participants classified 12 blocks of 10 stimuli, with the participant reporting the number of even numbers at the end of each block, and initiating the next block by pressing a key. At the end of the experiment, participants were asked to provide a written description of their categorization strategy. There was no forewarning that they would be expected to give a verbal description

The concurrent load task was identical to that employed in Experiments 1A and 1B, except that the presentation of digits started at the beginning of each block of 10 trials and only ceased at the end of each block (in Experiments 1A-1B, the digits were only presented on each trial from the onset of the to-be-classified stimulus to the detection of a response to that stimulus). This change in the concurrent load procedure was a consequence of the change in the classification procedure.

A further change to the procedure was that participants in the Load condition began the experiment with 10 trials of digit task practice, in which the participants were exposed to the digit task in the context of a simplified classification task. In the simplified classification task, the stimuli of the main classification task were replaced with the letters A and B. The classification task on these practice trials was simply to press the left-hand key if the to-be-classified stimulus was the letter A and to press the right-hand key if the to-be-classified stimulus was the letter B. These practice trials served to familiarize participants with the task requirements of the digit task and the basic response key requirements of the classification task. Participants in the No Load condition did not receive practice on the digit task but did receive 10 trials of the simplified classification task prior to starting the classification task proper.

Results and Discussion

Participants did not describe their classification at the end of each block; therefore, unlike in Experiments 1A and 1B, the following analyses are based solely on the sorts participants objectively produced. Inspection of the post-experiment questionnaire indicated that twenty-eight out of thirty-five participants (one questionnaire was lost) correctly described the classification strategy they had employed.

The proportion of overall similarity sorts produced by participants was significantly lower in the presence of a concurrent load (mean = .02, s.d. = .07) than in its absence (mean = .41, s.d. = .41), $t(34) = -4.01$, $p < .005$. The proportion of uni-dimensional sorts was significantly higher in the presence of a concurrent load (mean = .91, s.d. = .16) than in its absence (mean = .53, s.d. = .43), $t(34) = 3.50$, $p < .005$. The prevalence of Other sorts in the presence of load (mean = .07, s.d. = .11) and in its absence (mean = .06, s.d. = .10) did not differ significantly, $t(34) < 1$.

The results of Experiment 2, like those of Experiments 1A and 1B, are consistent with the idea that, under load conditions, participants have fewer cognitive resources available for categorization and thus have difficulty in implementing an effortful overall similarity strategy. Experiment 2 adds to the findings of Experiments 1A and 1B in that this conclusion still holds when the memory requirements of the task are minimal and when there is no requirement to produce a verbal description of the sort produced (other than an unheralded request after the experiment has been completed).

Experiment 3

If overall similarity classification is more demanding of cognitive resources than uni-dimensional classification, then those participants spontaneously producing overall similarity sorts might be predicted to have higher scores on individual difference measures of cognitive resources than participants producing uni-dimensional sorts. As Experiments 1A, 1B and 2 all implicate working memory as an important resource for overall similarity classification, Experiment 3 investigates the relationship between working memory capacity at an individual level and the prevalence of overall similarity classification, employing the simultaneous presentation version of the match-to-standards task used in Experiment 2.[1].

Method

Participants

Forty-six participants from the University of Exeter took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision.

Materials and apparatus

The stimuli were one of the lamp stimulus sets used by Milton and Wills (2004, Experiment 5), which had the same abstract structure as the stimuli used in the previous experiments. Each lamp had four variable features; lampshade (with either 5 or 10 dots), width of stand (wide or narrow), colour of bar (light or dark blue) and size of base (long or short). See Figure 4C for the prototypes of each category. The apparatus was identical to that used in the previous experiments.

Procedure

The procedure was split into two parts. In the first part, participants were presented with a total of 80 classification trials, in 8 blocks of 10 trials. Each trial began with a blank screen that was presented for 1000ms, followed by three lamps arranged as in Experiment 2. In other words, there were two lamps at the top of the screen that depicted the prototypes for category A and B with the to-be-classified lamp presented below the prototypes. Each stimulus array remained on the screen until the participant placed the to-be-classified lamp into either category A (by pressing the 'c' key on the keyboard) or category B (by pressing the 'm' key). After the participant made a response the next trial began. At the end of each block, participants were asked to write down the sorting strategy they used before moving on to the next block.

In the second part of the procedure, participants' working memory capacity was assessed using an operation span (OSPAN) task (e.g. Turner & Engle, 1989). In this task, participants were presented with a total of 60 trials split into 15 groups of trials. There were 3 groups each of 2, 3, 4, 5 and 6 trials and participants were presented with the groups of trials in a pseudo-random order. Each trial consisted of a simple mathematical equation (e.g. 2 x 3 + 1 = 7) presented simultaneously with a word (e.g. BED). The participant's task was to indicate whether the answer given to the equation was correct or incorrect. They were also required to remember each word. Each equation/word combination was response terminated with a timeout of 5 seconds. At the end of each group of trials, participants were asked to recall, in order, the words presented within that group. For example, in a block of size 3, participants would be presented with three equation/word pairs. After the presentation of the third pair, participants would be asked to recall the first word in the group followed by the second word and finally, the third word. Participants were not allowed to backtrack and change a previously given answer. There was no limit placed on time for recall.

Results and Discussion

Participants' working memory span was defined as the largest group size for which the participant could recall all of the words within a group for all 3 instances of that group size. For example, if participants could recall both words presented in a group of size 2 for all 3 instances of that group size, and all 3 words in each of the 3 instances of a group size of 3, but for the group size of 4, only two of the 3 instances had all 4 words recalled correctly, then that participant's working memory span was defined as 3 (e.g. Cianciolo & Sternberg, 2004).

For analysis purposes, participants were split into two groups depending upon the type of sort they produced in the first block of the categorization task. Data from the first block was used because, with continued practice with the stimuli over successive blocks, the role of working memory capacity with regards to the prevalence of overall similarity sorting might be expected to decrease as participants became more familiar with the task (cf. Milton & Pothos, 2011). The mean working memory span for participants who produced an overall similarity sort in block 1 was 3.73 (s.d. = 1.49, N = 15), whilst for participants who produced a uni-dimensional sort in block 1, the mean working memory span was 2.39 (s.d. = 1.80, N = 31). No participants produced an Other sort in block 1 (see also Milton & Wills, 2004, Experiment 5; Milton & Wills, 2009, both experiments). Participants who used an overall similarity sorting strategy in the first block of the categorization task had a significantly higher working memory span than those who used a uni-dimensional strategy; $t(44) = 2.507$, $p < .05$..This effect is specific to the first block, no other block approaches significance.

The results of Experiment 3, as with the results of our previous experiments, are consistent with the idea that overall similarity classification is more requiring of working memory resources than single-dimension classification. Participants who favour an overall similarity classification strategy from the outset tend to have larger working memory capacities than participants who initially produced a single-dimension classification strategy.

Experiment 4

Smith and Kemler Nelson (1984, Experiment 5) employed two instructional conditions; a *first-impressions* condition ("classify without thinking about it, just give your first impression, just let whatever happens happen", p. 150) and a *meticulous* condition ("give the exact opposite of your first impression, carefully decide which two squares go together best, be sure to place each card exactly where you want it, and be meticulous and careful, taking all the time you need", p. 150). Our Experiment 4 was conceptually similar in that we compared meticulous instructions to the standard, fairly neutral, instructions used in previous experiments (both sets of instructions are presented in the Appendix). In light of the results of Experiments 1A, 1B, 2 and 3, we predicted that meticulous instructions would increase the prevalence of overall similarity sorting in a match-to-standards procedure. Instructing participants to approach the task meticulously, if it has any effect, presumably increases the effort expended on the task. Under a more-is-more hypothesis, increased effort induced by instruction should lead to an increased prevalence of overall similarity responding.

Method

Participants, Materials and Apparatus

Twenty-eight participants from the University of Exeter took part in the experiment in return for course credit or payment. All participants had normal or corrected to normal vision. The stimuli and apparatus were identical to those used in Experiment 3.

Procedure

Participants were randomly allocated to the control (neutral instructions) and meticulous ('think carefully' instructions) conditions and were tested individually in a quiet cubicle. The procedure for both conditions was identical except the instructions given both before and during the experiment differed in the two conditions. The participants sat in front of the computer screen at a distance of approximately 50cm and were given written instructions on the screen before the experiment began.

Participants in the neutral instructions condition were instructed to sort the stimuli in the way they thought most appropriate (as in Milton & Wills, 2004, see the Appendix for the full instructions). In contrast, participants in the meticulous instruction condition were told that the stimuli they were about to see were complex and that they should be meticulous and careful when classifying each stimulus. In addition, participants in the meticulous instruction condition were asked to study each image in detail. Participants were presented with a total of 80 trials, in 8 blocks of 10 trials. Each trial began with a blank screen that was presented for 1000ms, followed by three lamps arranged as in Experiments 2 and 3. In other words, there were two lamps at the top of the screen that depicted the prototypes for category A and B with the to-be-classified lamp presented below the prototypes. Each stimulus array remained on the screen until the participant placed the to-be-classified lamp into either category A (by pressing the ‘c’ key on the keyboard) or category B (by pressing the ‘m’ key). After the participant made a response the next trial began. At the end of each block, participants were asked to write down the sorting strategy they used before moving onto the next block.

Results and Discussion

Each participant’s sorting strategy was classified in the same manner as in Experiments 1A, 1B and 3. Overall similarity sorting was significantly more prevalent in the meticulous-instructions condition (mean = .75, s.d. = .27) than in the neutral-instructions condition (mean = .36, s.d. = .40), $t(26) = 3.043$, $p < .01$. Uni-dimensional sorting was significantly less prevalent in the meticulous-instructions condition (mean = .23, s.d. = .26) than the neutral-instructions condition (mean = .58, s.d. = .39), $t(26) = -2.881$, $p < .01$. The prevalence of Other sorts did not differ between the meticulous-instructions condition (mean = .02, s.d. = .06) and the neutral-instruction condition (mean = .06, s.d. = .18), $t(26) = 0.7503$, $p = .49$.

The results of Experiment 4, like the results of the previous experiments, are consistent with the idea that the implementation of an overall similarity classification strategy is more effortful than the implementation of a uni-dimensional classification strategy. Where people are directed to take particular care and attention, the majority produce an overall similarity classification in this experiment. However, with more neutral instructions, the majority produce a uni-dimensional classification.

General Discussion

The results of the five experiments reported in this article may be straight forwardly summarized – implementing an overall similarity classification strategy is more effortful than implementing a single-dimension classification strategy, at least within the match-to-standard procedure on which our experiments are based. This conclusion finds support across three different stimulus sets (see Figure 4), two different versions of the match-to-standards procedure (simultaneous presentation – Experiments 2-4; sequential presentation – Experiments 1A and 1B)[2] and three different manipulations - concurrent load (Experiments 1A, 1B and 2), individual differences in working memory capacity (Experiment 3), instructions (Experiment 4).

We believe that these results provide a strong evidence base for the claim that overall similarity classification is more effortful than single-dimension classification in the match-to-standard procedure. As such, these results pose a challenge to the view that overall similarity classification is a 'fall back' mode of classification that is resorted to when cognitive resources are limited (e.g. Couchman et al., 2010) – a view we have described in the current article as the *less-is-more* hypothesis because it appears to suggest that as the availability of cognitive resources is reduced, more of the available information in the stimulus has control over the classification decision.[3]

Of course, the less-is-more hypothesis has not been suggested in the absence of evidence Two principal sources of evidence in support of a less-is-more hypothesis are those using (a) the triad procedure (Smith & Kemler Nelson, 1984; Ward, 1983), and (b) the criterial-attribute procedure (Kemler Nelson, 1984; Smith & Shapiro, 1989). A third set of evidence that one might take to be supportive of a less-is-more hypothesis – work using the

Ashby-Maddox procedure (e.g. Zeithamova & Maddox, 2006) – is perhaps best considered as being more about the verbalisability of category structures (see the Introduction for further details).

We see no necessary contradiction between the results of the current studies, and the results of previous work employing the criterial-attribute procedure. What the criterial-attribute procedure demonstrates is that the search for the one perfectly predictive single-dimension rule amongst four possibilities, all of which are at least 75% predictive, can be highly effortful. So, whilst the criterial-attribute procedure tells us something about the difficulty of *discovering* non-obvious one-dimensional rules, the match-to-standards procedure tells us about difficulty of *implementing* one-dimensional classifications. In some cases, single-dimension rules may be simultaneously hard to discover and easy to implement once discovered.

There is clearly more work to be done in reconciling the results of the current article with results produced through use of the triad procedure. Nevertheless, the results of Experiments 1A, 1B, 2 and 4, which substantially increase the evidence base in support of a more-is-more hypothesis within the match-to-standards procedure, raise the possibility that it is the inherent nature of the triad procedure, rather than other factors (e.g. differences in time pressure; differences in type of load task; sequential vs. simultaneous presentation) that lead to the difference in results.

The idea that overall similarity classification is requiring of substantial cognitive resources may seem paradoxical, given that natural categories appear to have an overall similarity structure (Mervis & Rosch, 1981; Rosch & Mervis, 1975), and are classified accurately and very rapidly (Thorpe & Imbert, 1989). However, the exposure we receive to natural objects, and the practice we have in classifying them, presumably both contribute to the speed and accuracy with which everyday overall similarity categories are employed. By analogy with studies of automaticity of other cognitive skills (e.g. Logan, 1988) it seems likely that an initial effortful process of deriving overall similarity classifications would later be overtaken by fast retrievals from a store of previous exemplars.

In the remainder of this paper, we discuss the extent to which the results of the current article may be specific to the stimuli we chose to employ. We also discuss some other issues pertaining to the particular procedure we employed.

Stimulus factors

All the stimuli employed in the current paper seem likely to be separable by Garner's (1974) definition, and it is possible that our results would not generalize to more integral/holistic stimuli. On the other hand, the speeded categorization of faces (generally considered to be fairly holistic stimuli) seems to be primarily based on a single feature (Goshen-Gottstein & Ganel, 2000), with the domination of single features reducing as presentation time increases (MacRae & Martin, 2007). This result seems broadly consistent with the idea that overall similarity classification is more requiring of cognitive resources than single-feature classification.

Relatedly, categories in the natural world typically have features whose instantiation differs between categories. For example, both birds and people sing, but the instantiation of singing is different in these two categories. In contrast, the features in the stimuli of the current experiments appear in identical form in both categories. It is possible that our results would not generalize to categories that have features whose instantiation differs between categories. On the other hand, Brooks, Squire-Graydon and Wood (2007), using instantiated features of this form, found that diverting attention in a supervised overall similarity classification task led to an increased likelihood that participants would state that there was a perfect single-attribute rule that defined the classification. This result seems broadly consistent with the idea that overall similarity classification is more requiring of cognitive resources than single-attribute classification, although the procedure employed is quite different to the one employed in the current paper, and further research is needed.

Procedural factors

One commonly employed unsupervised classification procedure is an *array* sort (e.g. Medin et al., 1987). An array sort typically involves presenting all to-be-classified stimuli simultaneously. As Regehr and Brooks (1995) have demonstrated, the array sort procedure produces a lower prevalence of overall similarity classification than a match-to-standards classification of the same stimuli. Such a result is consistent with the idea that overall similarity classification is more effortful than single-dimension classification. An array sort procedure presents the participant with a very large number of potential stimulus component comparison pairs, with no clues from the experimenter as to how to reduce this to a more manageable level (e.g. an array sort of the stimuli in the current paper presents the participant with 180 potential pairwise comparisons – 45 for each of the four stimulus dimensions). Even under untimed conditions, participants are presumably unable or unwilling to use all this information. By picking on a single dimension the entire classification task can be reduced to just a few comparisons. In a match-to-standards procedure, the experimenter produces a similar level of simplification by splitting the task into a set of discrete trials and by providing prototype stimuli. On any given trial, an overall similarity classification of the presented item can be achieved by making a small number of comparisons (no more than eight in the experiments of the current paper). Of course, a single-dimension sort can be achieved even more economically but for some participants at least - possibly the ones with larger working memories - overall similarity classification remains a viable option, and has the advantage over a single-dimension classification that it does not ignore 75% of the presented information.

Conclusion

It is sometimes argued that the implementation of an overall similarity classification is less effortful than the implementation of a single-dimension classification. In the current article, we have argued that the evidence securely in support of this view is limited, and have presented a series of experiments that support the opposite conclusion - implementing an overall similarity classification strategy is *more* effortful than implementing a single-dimension classification strategy.

References

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

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56*, 149-178.

Bott, L., Brock, J., Brockdorff, N., Boucher, J., & Lamberts, K. (2006). Perceptual similarity in autism. *Quarterly Journal of Experimental Psychology*, *59*(7), 1237-54.

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

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory and Cognition, 35*, 1-14.

Cianciolo, A.T. & Sternberg, R.J. (2004). *Intelligence: A brief history*. Malden, MA: Blackwell Publishing

Couchman, J. J., Coutinho, M. V. C., & Smith, J. D. (2010). Rules and Resemblance: Their Changing Balance in the Category Learning of Humans (Homo sapiens) and Monkeys (Macaca mulatta). *Journal of Experimental Psychology-Animal Behavior Processes, 36*, 172-183.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science, 311*, 1005-1007.

Edwards, D. J., Perlman, A., & Reed, P. (2012). Unsupervised Categorization in a sample of children with autism spectrum disorders. *Research in Developmental Disabilities*, *33*, 1264-1269.

Filoteo, J. V., Lauritzen, S., & Maddox, W. T. (2010). Removing the frontal lobes: The effects of engaging executive functions on perceptual category learning. *Psychological Science, 21*, 415-423.

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

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*, 231-62.

Goshen-Gottstein, Y., & Ganel, T. (2000). Repetition priming for familiar and unfamiliar faces in a sex-judgment task: Evidence for a common route for the processing of sex and identity. *Journal of Experimental Psychology: Learning, Memory & Cognition, 26*, 1198-1214.

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

Kemler Nelson, D. G. (1988). When category learning is holistic: a reply to Ward and Scott. *Memory & Cognition*, *16*, 79-89.

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

Logan, G. D. (1988). Towards an instance theory of automatization. *Psychological Review, 95*, 492-527.

MacRae, C. N., & Martin, D. (2007). A boy primed Sue: Feature-based processing and person construal. *European Journal of Social Psychology, 37*, 793-805.

Maddox, W.T., Pacheco, J., Reeves, M., Zhu, B., Schnyer, D.M. (2010). Rule-based and information-integration category learning in normal aging, *Neuropsychologia, 48,* 2998-3008.

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

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89 – 115.

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

Milton, F., & Pothos, E. M. (2011). Category structure and the two learning systems of COVIS. *European Journal Of Neuroscience*, *34*, 1326-36.

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.

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

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

Newell, B. R., Dunn, J. C., & Kalish, M. (2010). The dimensionality of perceptual category learning: a state-trace analysis. *Memory & Cognition*, *38*, 563-81.

Newell, B. R., Moore, C. P., Wills, A. J. & Milton, F. (submitted). Replacing the frontal lobes? Having more time to think improves “implicit” perceptual categorization. A comment on Filoteo, Lauritzen & Maddox, 2010. *Submitted to Psychological Science.*

Patalano, A. L., Smith, E. E., Jonides, J., & Koeppe, R. a. (2001). PET evidence for multiple strategies of categorization. *Cognitive, Affective & Behavioral Neuroscience*, *1*, 360-70.

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.

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin, 119*, 3-22.

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

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

Tharp, I. J., & Pickering, A. D. (2009). A note on DeCaro, Thomas, and Beilock (2008): further data demonstrate complexities in the assessment of information-integration category learning. *Cognition*, *111*, 411-5.

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.

Tracy, J., Mohamed, F., Faro, S., Pinus, A, Tiver, R., Harvan, J., Bloomer, C., et al. (2003). Differential brain responses when applying criterion attribute versus family resemblance rule learning. *Brain and Cognition*, *51*, 276-286.

Turner, M. L. & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127 – 154.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*, 168-76.

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.

Wattenmaker, W.D. (1992). Relational properties and memory-based category construction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1125-1138.

Wills, A. J., Lea, S. E. G., Leaver, L. A., Osthaus, B., Ryan, C. M. E., Suret, M. B., et al. (2009). A Comparative Analysis of the Categorization of Multidimensional Stimuli: I. Unidimensional Classification Does not Necessarily Imply Analytic Processing; Evidence From Pigeons (*Columba livia*), Squirrels (*Scurius carolinensis*), and Humans (*Homo sapiens*). *Journal of Comparative Psychology*, *123,* 391-405.

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

Footnotes

1. One might argue that the sequential procedure would be a better choice, given that it contains a short-term memory component. However comparison of Experiments 1A and 2 indicates that concurrent load affects the prevalence of overall similarity classification to a similar extent in the simultaneous and sequential procedures.

2. The reduction in overall similarity classification is associated with an increase in single-dimension classification in the simultaneous version of the procedure, but with an increase in Other classifications in the sequential version. However, this difference is not maintained across other uses of the match-to-standards procedure (see Milton et al., 2008).

3. Participants with Autistic Spectrum Disorder (ASD) are more likely than controls to spontaneously produce single-dimension classifications (Edwards, Perlman & Reed, 2012), and they also learn multidimensional category structures more slowly than controls (Bott, Brock, Brockdorff, Boucher & Lamberts, 2006). These results are also consistent with a *more-is-more* hypothesis.

Appendix

The instructions given to participants in Experiment 4 were as follows; passages of text that differ between the two instruction sets are shown in bold text.

Instructions given in the control condition

Thank you for taking part in this task. Please read the following instructions carefully.

Two pictures will be displayed at the top of the screen. One of these (on the left) will be characteristic of Category A and the other (on the right) will be characteristic of category B. These two characteristic pictures will be present throughout the experiment.

Directly under these two characteristic pictures another will be presented. Your task is to put this lower picture into either category A (by pressing the button marked A) or into category B (by pressing the key marked B).

There are many ways in which these pictures can be split and there is no correct answer. **We are just interested in what you think is the most appropriate way to sort these pictures.** There is no time limit and you are encouraged to take as much time as you need to complete the task.

In total there will be 10 pictures to categorize in a "block" and there will be 6 blocks in total. There will be an opportunity to rest at the end of each block if you so wish.

Instructions given in the meticulous condition

Thank you for taking part in this task. Please read the following instructions carefully.

The pictures you are about to see are quite complex. You should take particular care in your evaluation of how they differ. Study each image in detail.

BE METICULOUS! BE CAREFUL!!

Two pictures will be displayed at the top of the screen. One of these (on the left) will be characteristic of Category A and the other (on the right) will be characteristic of category B. These two characteristic pictures will be present throughout the experiment.

Directly under these two characteristic pictures another will be presented. Your task is to put this lower picture into either category A (by pressing the button marked A) or into category B (by pressing the key marked B).

There are many ways in which these pictures can be split and there is no correct answer. **Carefully decide which category each picture should go in and be sure to place each picture exactly where you want it.** There is no time limit and you are encouraged to take as much time as you need to complete the task.

In total there will be 10 pictures to categorize in a "block" and there will be 6 blocks in total. There will be an opportunity to rest at the end of each block if you so wish.

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

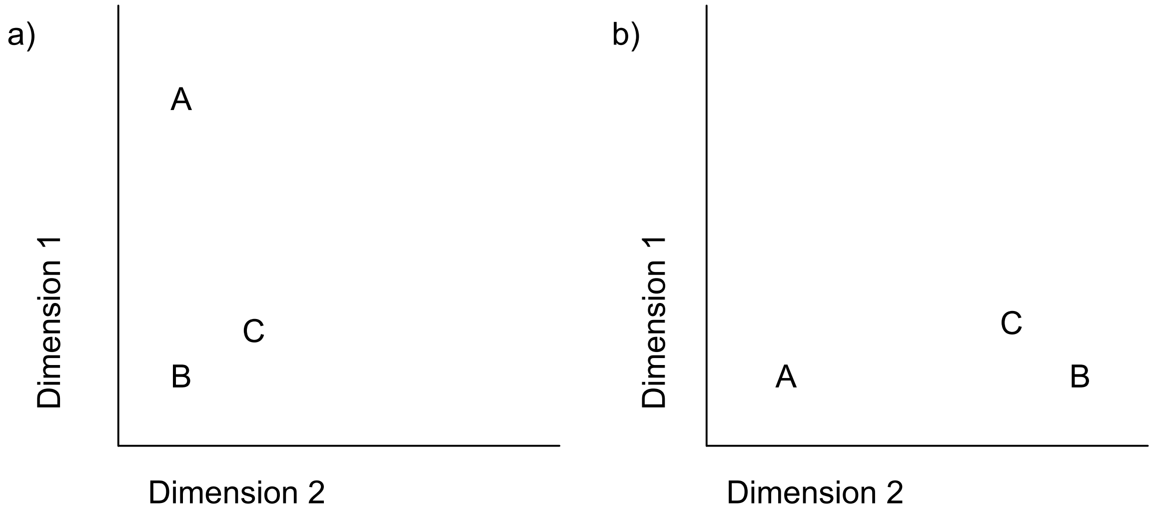

Figure 1. Abstract structure of the triad task. Typically, on half of trials A and B are identical on dimension 1 (Panel b), and on the other half of trials A and B are identical on dimension 2 (Panel a).

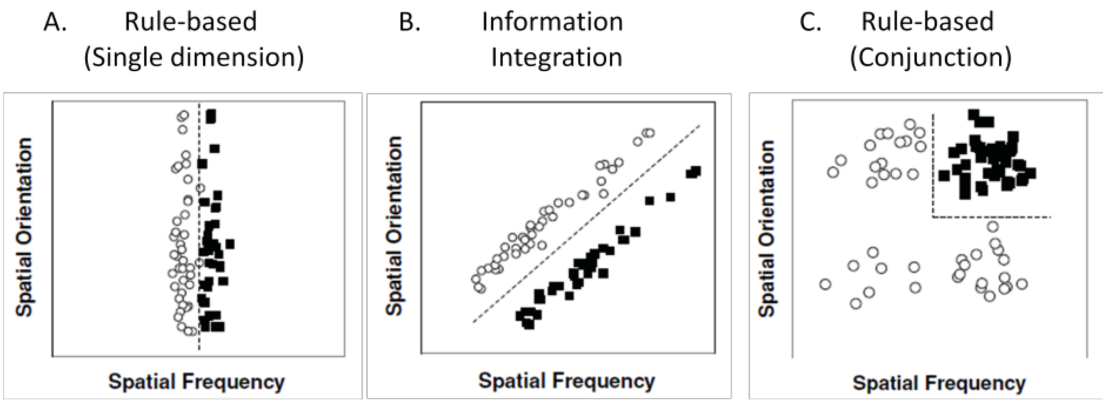

Figure 2. Examples of rule-based (single dimension), information integration, and rule-based (conjunction) category structures. Each open circle represents one member of category 1; each filled square represents one member of category 2. Figure adapted from Zeithamova and Maddox (2006).

Category A	Category B
0 0 0 0	1 1 1 1
0 1 0 0	1 0 1 1
0 0 1 0	1 1 0 1
0 0 0 1	1 1 1 0

Test
0 1 1 1
1 0 0 0

Figure 3: Abstract structure of the stimuli used in the criterial-attribute procedure.

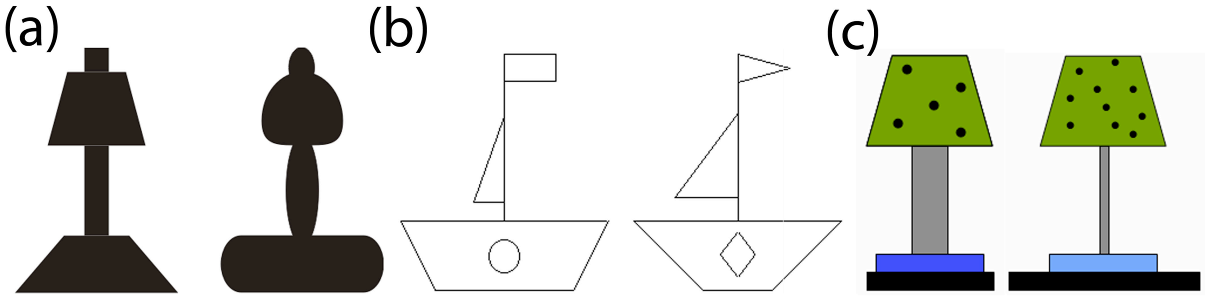

Figure 4. Examples of the stimuli used in (a) Experiments 1A and 1B, (b) Experiment 2, and (c) Experiments 3 and 4. The prototypes for each category are shown.