

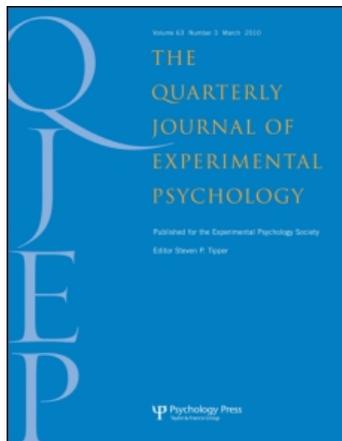
This article was downloaded by: [Bradmetz, Joel]

On: 6 January 2011

Access details: Access Details: [subscription number 925010947]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Quarterly Journal of Experimental Psychology

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t716100704>

An extended study of the nonindependence of stimulus properties in human classification learning

Fabien Mathy^a; Joël Bradmetz^b

^a Université de Franche-Comté, Besançon, France ^b Université de Reims, Reims, France

First published on: 02 August 2010

To cite this Article Mathy, Fabien and Bradmetz, Joël(2011) 'An extended study of the nonindependence of stimulus properties in human classification learning', The Quarterly Journal of Experimental Psychology, 64: 1, 41 – 64, First published on: 02 August 2010 (iFirst)

To link to this Article: DOI: 10.1080/17470218.2010.493972

URL: <http://dx.doi.org/10.1080/17470218.2010.493972>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

An extended study of the nonindependence of stimulus properties in human classification learning

Fabien Mathy

Université de Franche-Comté, Besançon, France

Joël Bradmetz

Université de Reims, Reims, France

Categorization researchers have tried to verify their models through laboratory experiments with simplified stimulus sets, a requirement that can rarely be met in real-world situations in which properties are often connected. Still, the targeted simplification of the material might be illusory. We replicate and extend Love and Markman's (2003) study of the nonindependence of canonical stimulus properties such as *size*, *colour*, and *shape* in human classification learning, in which the authors concluded that shape takes precedence over other dimensions. To support their hypothesis, Love and Markman showed that certain classifications are more difficult for participants when shape is combined to one of its putative subordinate features, size or colour, than when shape is irrelevant to the task. A data set of 290 + 50 adult participants completing one or more classification tasks was collected. The results confirm that certain combinations of shape, size, and colour can hinder or facilitate classification learning, but not necessarily in the form expected by the nonindependence postulated by Love and Markman, especially in Experiment 2 where a totally reverse pattern of difficulty is observed (shape does not take precedence over other dimensions). Also, we show that simple similarity effects in clustering retain considerable intuitive appeal and can offer an alternative account to the nonindependence of stimulus properties, especially because slight variations in the dimensions chosen make the observations of Love and Markman unstable.

Keywords: Concept learning; Categorization; Boolean; Rules; Similarity.

Does the logical structure of categories only determine category learning performance?

The stimuli of perception are multidimensional. A problem of fundamental importance is to determine how features combine when they are

processed in a particular task (Ashby & Townsend, 1986). Many psychologists or programmers in machine learning (but few philosophers—see Fodor, 1998; Ryle, 1951; Wittgenstein, 1953) are sympathetic to the

Correspondence should be addressed to Fabien Mathy, Université de Franche-Comté, Département de Psychologie, 30–32 rue Mégevand, 25030 Besançon Cedex, France. E-mail: fabien.mathy@univ-fcomte.fr

This research was partly supported by a postdoctoral research grant from the Fyssen Foundation awarded to Fabien Mathy in 2005. We thank the students of Rutgers University and the Université de Reims Champagne-Ardenne who kindly volunteered to participate in this study. The authors wish to thank Cordelia Aitkin, Erica Briscoe, David Fass, and Jacob Feldman from the Visual Cognition Lab for their many helpful comments.

so-called classical view that concepts can be created using conjunctions and disjunctions of features (e.g., a play is a recreational activity, a puzzle, or a game such as a competition with rules to determine a winner, etc.). The focus of this paper is the use of such logical operations based on the *and* and *or* operators, two major building blocks for human conceptualization. In the case of natural concepts, the conjunctions of features or components are rarely the result of independent associations (e.g., first-degree murder, red cherry). Unfortunately, such systematic associations preclude experimental studies on concept learning, in the same way as words can prevent psychologists from evaluating memory span because their meaning facilitates chunking processes. To alleviate this problem, psychologists in the 1950s began to devise “cleaner” artificial classification tasks by combining alleged independent dimensions made up of simple values such as a square, triangle, or circle, and so on. In these tasks, participants are required to learn some arbitrary rules that separate a set of objects into two categories on the basis of feedback given by the experimenter (Bourne, 1970; Bruner, Goodnow, & Austin, 1956; Hovland, 1966; Shepard, Hovland, & Jenkins, 1961). The objective was to measure the effect of the logical structure of the categories on performance. Since then, prototype theories (Hampton, 1993; Osherson & Smith, 1981; Rosch & Mervis, 1975; Smith & Medin, 1981) and exemplar theories (Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984, 1986; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994) have provided a good fit to many results in classification studies using similarity metrics. Bypassing many clustering or hybrid models, a rule-based approach has resurfaced since 2000 with the emergence of models using compressibility metrics instead of the language-based classical approach in order to account for the complexity of the logical structure of the categories (Bradmetz & Mathy, 2008; Feldman, 2000, 2006; Lafond, Lacouture, & Mineau, 2007; Mathy & Bradmetz, 2004; Vigo, 2006). Most of these recent studies still use stimulus sets made of basic dimensions, sometimes

called canonical dimensions. Love and Markman (2003; hereafter L&M) recently offered a critical examination of the postulate of independence between features such as *colour*, *shape*, and *size*, which experimenters frequently choose in artificial concept studies as canonical dimensions. L&M concluded that these dimensions are not independent by showing that Type II concepts (see later explanation) are simpler to learn when shape is not relevant to the task. Let us introduce some primitive concepts and the requisite notation to elaborate on L&M’s study.

In classification tasks, a set of stimulus objects are presented sequentially to participants. For each stimulus, participants are asked to group the stimuli in two mutually exclusive and exhaustive classes. From the feedback participants receive from the experimenter who knows the target classification rule, participants are progressively able to learn the classification rule, by trial and error. Classification learning is analogous to natural concept learning in which a class of objects is associated to one label. For instance, children progressively learn to recognize the positive instances of a feline from the negative instances (i.e., other animals), thereby forming an abstract concept of feline from the extensive list of felines they have been shown.

The three Type II classification tasks (i.e., *a*, *b*, and *c*) studied by L&M are illustrated in Figure 1. Type II classification tasks as well as other classification problems were originally studied by Shepard et al. (1961) and since then have been widely used in the literature on categorization. For Type II three-dimensional classification problems, two dimensions among three are relevant; information about the third dimension is irrelevant to solving the problem. The three tasks (*a*, *b*, and *c*) are, respectively, *size–shape* relevant, *size–colour* relevant, and *colour–shape* relevant. For instance, Type II *a* is *size–shape* relevant in that the rule “large squares *or* small circles”, which structures the problem, is based on the shape and size dimensions only; colour is not diagnostic for categorization in Type II *a*. The cube on the left in Figure 1 represents a set of stimuli constructed from a combination of three Boolean

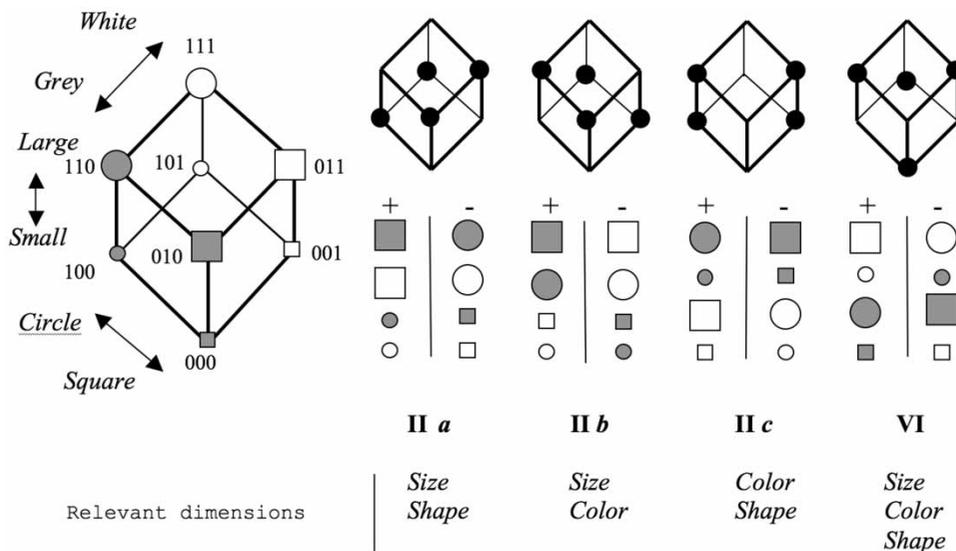


Figure 1. Type II concepts, labelled as in Shepard, Hovland, and Jenkins (1961). The cube on the left represents a training sample of stimuli generated from the combination of three Boolean dimensions. The cubes on the right define target conceptual structures. Positive examples of a concept are shown as black dots on the cubes and are listed below the cube in the “+” column; negative examples are vertices without dots and are listed below the cube to the right of the bar. There are three Type II concepts, depending on which pair of dimensions is relevant, but their structural complexity is assumed to be equivalent—that is, “(a and b) or (a' and b')”. For the sake of comparison, a much more difficult concept called Type VI—“(a and b and c) or (a and b' and c') or (a' and b and c) or (a' and b and c')”—is indicated in the right-hand column.

dimensions. Note that a rule operates on a stimulus set.¹ In Type II classification problems, four stimuli belong to one category and the other four stimuli to another category, using a specific structure (i.e., a rule), which can be depicted by black dots on cubes in which the stimuli are no longer represented (but the stimuli are still virtually attached to the vertices). The black dots represent the stimuli assigned to the first category whereas the empty vertices represent the stimuli assigned to the second category (sometimes called the positive and negative categories).

For instance, in Type II *a*, the black dots cover the large squares and the small circles. These two groups form two subcategories within the positive category. Note that the two subcategories, also called clusters, are very dissimilar, which makes the task quite difficult for participants. Another

difficulty is associated with the presence of the irrelevant dimension: There are objects of different colours (grey and white) in each of the clusters. Focusing on colour only would delay learning of the correct classification rule. Our study focuses on the effects of the dissimilarities between objects within clusters and on the dissimilarities between clusters on learning.

Despite the idea that performance is mainly dictated by the structure of the classification tasks in the classical studies, L&M showed that an important variance in performance was elicited by the choice of the relevant dimensions. They showed that the “large and grey, or small and white are positive; others are negative” rule (a Type II *b* classification) is less difficult for participants than the other two classifications implying a rule of similar structural complexity but assuming

¹ In Figure 1, each stimulus is attached to one vertex of a cube. The cube represents the whole stimulus set. The number of edges separating two stimuli represents the distance between the stimuli. For instance, the three differences between a large grey circle and a small white square are adequately represented by a distance of three edges (this type of distance is called city-block, by opposition to the Euclidean distance, which would compute distance using diagonals).

that shapes are relevant features (Type II *a* and Type II *c*). That a Type VI classification structure (Figure 1) is more difficult for participants than a Type II can easily be accounted for, because the identification of positive stimuli requires a more complex combination of features: “If *grey*, then [if *big circle* or *small square*, then positive]; If *white*, then [. . .]”. The question of why it is less difficult for participants to combine size and colour than other combinations implying shape is more puzzling. Their explanation is that shape components (which can be nouns in a language) and size or colour (more often adjectives) are treated hierarchically. They posit that shape takes precedence over size and colour when shape is relevant to categorization, because of a relational dependence between shape, colour, and size. They propose that a stimulus such as a large red triangle is represented as:

```
triangle
colour(triangle) = red
size(triangle) = large
```

In this representation, the value of the shape dimension serves as the argument for the colour and size predicates.² The authors give the reason for which Type II concepts are more difficult to learn when the shape dimension is involved as follows: If size and colour are properties of a superordinate class shape, this organization requires participants to break the hierarchy when shape is combined with one of its properties (i.e., colour or size), whereas there is no conflict when colour and size have to be combined (because colour and size are processed at the same level). This problem is of great importance, as L&M claim: “General principles that govern ease of category learning . . . cannot be defined without consideration of the principles that govern how representations are formed. . . . In fact, the latter set of

principles may play a larger role in determining category learning performance than does the logical structure of categories” (p. 798). Despite the elegance of the theory, the investigations of L&M had some limitations. For instance, because the irrelevant dimension was not constant across the Type II classification tasks in L&M’s study (i.e., size in the colour–shape concept, colour in the size–shape concept, and shape in the colour–size concept), the difficulty encountered by participants in grouping objects within clusters and their difficulty in dealing with the precedence of shape were confounded. To test the reliability and the generalizability of their results, it is important to conduct a similar study with more extensive manipulations.

Relational dependence between features or simple influence of perception on categorization?

Our first objective was to test several implications stemming from L&M’s theory in a series of new experiments. For instance, do their conclusions hold when the irrelevant dimension is controlled? If yes, this would confirm their theory. Or, can the same patterns of results be observed when only the irrelevant dimension varies? If yes, this would discredit their theory. Our second objective was to compare their theory to simpler accounts based on the salience of dimensions. We based our development on Rosch’s suggestion that people tend to categorize in a way that maximizes within-category similarity and minimizes between-category similarity (Rosch, 1975; see also Homa, Rhoads, & Chambliss, 1979, who quantify the structure of categories by computing the ratio between within- and between-category distance). In the same vein, Goldstone (1994)

² Although not pointed out by the authors, an object-oriented language would also have been appropriate for describing these dependencies:

```
object.shape = triangle
object.shape.colour = red
object.shape.size = large
```

In programming, then, shape, size, and colour are not taken as properties that are encapsulated at the same level in the object.

reports that subsequent to category learning, participants show acquired distinctiveness (increased perceptual sensitivity for items that belong to different categories) and acquired equivalence (decreased perceptual sensitivity for items that are categorized together). Gureckis and Goldstone (2008) also report a within-category effect whereby stimuli belonging to different clusters of a given category are better discriminated than when they belong to the same cluster. Here, we do not investigate the influence of prior categorization on perception but rather the simple influence of prior perception on categorization. For instance, we posit that it is difficult for participants who a priori consider two shapes as very distinct to categorize these shapes into a same category or into a same cluster.³ Perceptual sensitivity to dimensions might account for the results observed by L&M, but not in a simple way, as explained in more detail below.

First let us make a few comments on L&M's study:

1. *L&M's theory focuses on the conjunction of relevant features and implicitly considers the irrelevant dimension an inert dimension.* However, in the Type II concepts they investigated, inhibition of the irrelevant dimension might be achieved at different costs. Inhibition plays a major role in rule learning, especially in the Wisconsin Card Sorting Test in which participants must shift attention across various possible dimensions (Berg, 1948; Heaton, Chelune, Talley, Kay, & Curtiss, 1993). L&M concluded that colour-size-relevant

classifications are the easiest kind. If participants have difficulties inhibiting shape in such classifications, we hypothesize that the easiness of these tasks can be modulated using other irrelevant dimensions. For instance, if the rule is "*big grey or small white*", participants are required to cluster together some circle and square stimulus objects. The clustering process might be facilitated by less salient irrelevant dimensions.

2. *L&M's theory does not focus on the disjunctive aspects of the rules.* However, Type II rules always entail two clusters per category (for instance, the "grey circles" and the "white squares" for the positive category). The salience of shapes might also be responsible for the difficulty participants have in putting together two clusters differing in shape into a given category. An important Gestalt principle of perceptual organization is that similar things tend to be grouped together. Conversely, dissimilar things tend not to be clustered together. For instance, in a colour-shape concept, it might be difficult for participants to grasp that both squares and circles belong to the same category (as is the case, for instance, for the rule "*grey circles or white squares*" in Table 1), whereas it might be less difficult to put some large grey objects and small white objects together in the same class in a size-colour concept. Therefore, shape might help participants separate clusters between categories, but it might prevent them from grouping clusters within categories as well. Given that a rule is based on a single category (i.e.,

³ The idea that shapes, sizes, and colours are not treated equally has been confirmed by research in cognitive development. In several studies where stimuli varying in shape and colour were presented to 4-month-old infants, shape overrode colour as the basis for preferential choice when the stimuli represented combinations of preferred and nonpreferred colours and shapes (Spears, 1964). More recently, Tremoulet, Leslie, and Hall (2000) showed that for 12-month-olds, a difference in shape had a large effect on identification, whereas colour difference did not. Inhelder and Piaget (1959) also noted that children tended to prefer shapes in free classification tasks (although preferences vary with age; Brian & Goodenough, 1929). When the triad shape/colour/size was considered in preferential-matching tasks (with children around 6 years of age), shape was distinctively preferred over colour, and colour was preferred over size (Kagan & Lemkin, 1961). Lee (1965) showed that preschool children have progressively a greater ease in utilizing form over colour and size in concept-identification tasks, with the youngest preschool children making fewer errors with color and size than with form, and with the oldest preschool children making fewer errors with form than with color and size. In another study, children over 5 years have also been shown to prefer colour over size in preferential-matching tasks (Pitchford & Mullen, 2001). Such biases can still be observed in adults when appropriate measures are made. On the dimensional-change card-sort task, for instance, Diamond and Kirkham (2005) showed that response times were longer when participants had to sort cards by colour than when they had to sort them by shape.

Table 1. *A series of Type II concepts in different contexts*

| Series | N | Concept number | | |
|----------------------|-----|------------------|------------------|------------------|
| | | 1 | 2 | 3 |
| 200 | 77 | SiSh | SiCo | CoSh |
| 300 | 144 | SiSh/Co' | SiCo/Sh' | CoSh/Si' |
| 310 | 79 | SiSh/Cst' | SiCo/Cst' | CoSh/Cst' |
| 320 | 80 | CstCst/Co' | CstCst/Sh' | CstCst/Si' |
| 400 | 66 | SiSh/ Co'Cst' | SiCo/ Sh'Cst' | CoSh/ Si'Cst' |
| Exp. 2. Shape-max | 50 | SiSh/Co' | SiCo/Sh' | CoSh/Si' |
| Exp. 2. Regular | 50 | SiSh/Co' | SiCo/Sh' | CoSh/Si' |
| Exp. 2. Shape-min | 50 | SiSh/Co' | SiCo/Sh' | CoSh/Si' |

Note: N = number of participants who learned a series of classification tasks of one kind. An abbreviation such as SiSh/Co' (i.e., the concept kind) means that the concept is size–shape relevant and colour irrelevant. The concept number (1, 2, or 3) added to the series number identifies each concept. For instance, 201 = SiSh (size and shape relevant), 301 = SiSh/Co' (size and shape relevant, but colour irrelevant), 311 = SiSh/Cst' (size and shape relevant, but a third dimension, which is constant across the series, is irrelevant), 321 = CstCst/Co' (the irrelevant dimension is colour, but the two relevant dimensions are constant across the series), 401 = SiSh/Co'Cst' (size and shape relevant, but colour irrelevant, as well as a fourth dimension, which is constant across the series).

participants classify the negative instances by negation of the rule defined on the positive ones), there is a possibility that participants are affected by these within-category structures. This would explain the difficulties encountered in shape-relevant concepts. L&M concluded that shape-relevant concepts were more difficult to learn despite the fact that shape is more salient than colour or size (p. 794). We hypothesize that the potential salience of shape might hinder the clustering of objects of different shapes into subcategories

(clusters). This hypothesis is not directly testable, because within-category dissimilarities and between-category dissimilarities are somewhat confounded in Type II. For instance, if two clusters within categories differ in size and shape, so it is for the clusters between categories.⁴ Because our hypothesis cannot be tested directly, the only way to give some credit to this hypothesis is to find some consistent patterns of saliency in various Type II tasks.

To summarize and to anticipate, L&M did mention that perceptual sensitivities (that can translate into preferential matching judgements) should help participants discriminate clusters from different categories in shape-relevant classifications, but they considered neither the potential difficulty that participants encounter in clustering together objects of different shapes within clusters within categories nor the possibility that participants encounter difficulties in grouping different clusters within categories. We hypothesize that saliency effects can modulate or exceed the nonindependence effect found by L&M, without mentioning that saliency effects might simply offer an alternative explanation to their results and our results. For instance, participants might have encountered difficulties in learning shape-relevant concepts in L&M because dissimilar shapes were found to be difficult to integrate into one category (giving that shape was a salient dimension), or participants could have found the size–colour concepts even less difficult if categorization had not been hindered by the need to inhibit the shape dimension. In brief, perceptual sensitivities should offer a systematic explanation of diverse patterns of categorization results, including the result of L&M. Our results show that perceptual sensitivity offers an appealing alternative to nonindependence effects in order to account for the different patterns of difficulties that we observe in our study. The combination of shape, size,

⁴ In other words, in the size–shape relevant concepts in Figure 1, the large squares are different in shape from the large circles and different in size from the small squares (idem for the small circles). The only difference is that the clusters within categories (e.g., the large squares and the small circles) differ in terms of both shape and size, which place them further apart than the clusters of the opposite category; the greater distance between the clusters within categories is noticeable in the cube, in which the positive clusters are opposed by a diagonal distance, whereas clusters between categories are only opposed by an edge.

and colour features effectively hinders or facilitates the generation of category representation, but not necessarily in the form expected by the relational conceptual organization postulated by L&M, especially in Experiment 2 where a totally reverse pattern of difficulty is observed. We show that slight variations in the dimensions chosen make the observations of L&M unstable.

In our first experiment, participants were presented with four versions of the size–shape, size–colour, and colour–shape relevant concepts. In a fifth condition, participants were presented with a series of Type II concepts in which the relevant dimensions were constant, but in which only the irrelevant dimension was manipulated (colour, shape, or size). In the sixth condition (Experiment 2), the classification structures were typical of L&M's study, except that several sets of stimuli were built by varying dimension saliency. Also, a different protocol was used in the last condition in order to avoid strict sequential learning. The last condition was devised to test whether the use of complex shapes could more radically distort the pattern observed by L&M. The first five conditions are presented and analysed simultaneously in a section called Experiment 1, whereas the sixth condition is described in the Experiment 2 section.

EXPERIMENT 1

We sought to differentiate the effects of the hierarchical organization of dimensions hypothesized by L&M from those resulting from the sensitivity to dimensions. If L&M's hypothesis is correct, the relationships between shape, colour, and size should apply in all the situations in which these dimensions are relevant conjunctive features, no matter what the irrelevant dimension. In addition, the alternative approach should account for the variation of performance due to the manipulation of the irrelevant dimension. Our goal is to find consistent patterns of saliency across the conditions in order to give some credit to the hypothesis that participants encounter difficulties both in grouping clusters within categories and in grouping dissimilar objects within clusters. Saliency effects are not

denied by L&M and could potentially combine with the nonindependence of features, but we aim at showing that the salience of dimensions is powerful enough to account for a full range of results. Note that similarity ratings are not measured beforehand as we expect similarity effects to be indirectly measurable in the tasks.

Method

Participants

The participants were 290 university students, 230 attending the University of Reims (France) and 60 attending Rutgers University (New Jersey, USA). Participants received course credit in exchange for their participation. The participants were divided into several groups and were given different tasks as described below. There is only one condition (Series 300, as described later) in which 60 American and 84 French students were administered the same three classification tasks, simply because one of the authors was working at Rutgers at that time. In all other conditions, the participants were native French speakers.

Series of concepts

In the present section and below, “classification tasks” are sometimes replaced by “concepts” in order to facilitate reading. As shown in Table 1, several Type II concepts were administered to participants in a variety of forms called series (200, 300, 310, 320, and 400). The exact procedure is described in the next section. Each series was composed of three versions labelled by adding 1, 2, or 3 to the series number (for instance, the 200 series can be divided into three classification tasks called 201, 202, and 203). This resulted in 15 different Type II concepts. The three versions per series were: (a) size and shape relevant (SiSh), size and colour relevant (SiCo), and colour and shape relevant (CoSh), in the 200 and 310 series; (b) colour irrelevant (Co'; note that the prime denotes a negation), shape irrelevant (Sh'), and size irrelevant (Si'), in the 320 series; or (c) a conjunction of both—that is, SiSh + Co', SiCo + Sh', and CoSh + Si', in the 300 and 400 series. In a series number, the three

digits hence represented, respectively, the dimensionality, an increment for distinguishing the series having identical dimensionalities, and the concept version. When a participant was assigned to one series, he or she was systematically administered the three versions.

In the series 200, the Type II concepts were generated using a subnormal structure since no third irrelevant dimension was used. The concepts were thus reduced to an “exclusive or” or “XOR” classification type. Concepts 201, 202, and 203 were SiSh, SiCo, and CoSh, respectively. The series 300 was merely a replication of L&M’s study materials, except that circles were used instead of triangles and that the manipulation was within subjects: Concepts 301, 302, and 303 were, respectively, SiSh + Co’, SiCo + Sh’, and CoSh + Si’. The series 310 was designed to control potential differences in inhibiting the irrelevant dimension: Concepts 311, 312, and 313 were SiCo + Cst’ (Cst standing for constant), CoSh + Cst’, and CoSh + Cst’, respectively. In other words, the irrelevant dimension did not vary across the 311, 312, and 313 conditions. The irrelevant dimension was hatched versus gridded. In the series 320, only the irrelevant dimension was manipulated, and the relevant conjunction did not vary (the rule was “*gridded and with a frame*, or *hatched and with a hat*”): The concepts 321, 322, 323 were CstCst + Co’, CstCst + Sh’, and CstCst + Si’, respectively. The series 400 was generated by adding a supplementary constant irrelevant dimension to the series 300: The three concepts were SiCo + (Co’Cst’), CoSh + (Sh’Cst’), and CoSh + (Si’Cst’). The second irrelevant dimension was *with a frame* versus *without a frame*. The 400 series was devised to enhance inhibition difficulties.

Procedure and stimuli

Each participant learned either three (201, 202, 203; 311, 312, 313; 321, 322, 323, $N = 80$) or one (301, 302, 303, $N = 144$; or 401, 402, 403, $N = 66$) series of Type II concepts on the basis of trial and error with corrective feedback, in less than a one-hour single session. The order of the classification tasks was randomized between

subjects. Testing was preceded by a brief explanation about how to sort stimuli (by pressing 1 or 0) and how to complete a classification (by filling up the entire progress bar). Feedback displayed at the bottom of the screen indicated whether a response was right or wrong. The feedback was provided for 2 s. One point was added to the progress bar for each correct response. A point was represented by an empty box that was filled in when the answer was correct. The number of boxes in the progress bar was equal to four times the length of the training sample—that is, 4×2^N ($N =$ number of dimensions). Participants had to correctly categorize stimuli on four consecutive blocks of 2^N stimuli—that is, they had to fill up a progress bar of 4×2^N points in a row, without knowing that reaching 2×2^N correct responses was considered the learning criterion. Participants were only instructed that the classification task would end once the entire progress bar filled up. The response times were measured during the last 2×2^N responses to determine whether differences in learning persisted with concept use. Participants were considered as using the concept rather than learning it during the last 2×2^N responses because this phase followed a period in which 2×2^N responses had just been correctly given. Responses had to be made in 8 s or less, otherwise participants lost 3 points on the progress bar. When wrong responses were given, all points scored so far were lost (the progress bar went back to 2×2^N points in cases where the participant succeeded in the learning phase). Success (4×2^N points scored) was rewarded with a digital image (animals, fractals, etc.). The criterion of 4×2^N was identical to the one used in the pioneer study of Shepard et al. (1961) in their first experiment.

Stimulus objects were presented one at a time in the upper part of the computer screen (the lower part of the screen was reserved for feedback). Blocks of 2^N stimuli were successively presented to participants with each stimulus appearing once per block. The first stimulus in each block was different from the last one in the previous block, although participants had no idea of where the blocks began. The positive and negative stimuli

were randomly ordered within blocks, and each new block was newly randomized.

The stimuli were geometric figures that could vary along five dimensions, depending on the series chosen, and each dimension used two values only: colour (any two of the following: yellow, orange, red, blue, green, or pink), shape (circles, squares), size (large, 5 cm × 5 cm, or small, 1 cm × 1 cm), filling (hatched or gridded, with lines separated by 2 mm), frame (white, with a hat of 0.5 cm × 0.5 cm wide, or outlined, 0.5 cm wide). The stimuli were presented sequentially, centred in a black window of 8 cm × 8 cm. The feedback was given in a white horizontal rectangular window of 8 cm × 2 cm. The rest of the screen was grey.

The stimuli are shown in the Appendix, Tables A1 and A2, and in Table 1 in a reduced format. For the 200 series, the stimuli were built from a combination of shapes and colours (the size was set to large). For the 300 series, the stimuli were built from a combination of shapes, colours, and sizes. For the 310 series, the stimuli were built from a combination of shapes, colours, sizes, and fillings. For the 320 series, the stimuli were built from a combination of fillings and frames, and colours, shapes, or sizes. The 400 series was made of a combination of shapes, colours, sizes, and fillings. The assignment of the physical dimensions was randomized for each concept and each participant, but constrained to obey the desired logical structure. For instance, for a colour–shape/size' concept in the 300 series (i.e., the 303 concept), in which only shape and colour were relevant to the classification, the computer generated a set of eight stimulus objects each made of a combination of two colours (e.g., red or blue), two shapes (square or circle), and two sizes (large or small)—that is, a large red square, a large blue square, and so on. One of the two possible assignments of shapes and colours to the classification rule was then randomly chosen (e.g., when the red squares and the blue circles are positive, the red circles and the blue squares are negative, and vice versa for another classification task). In this case, the participants had to induce the rule “the red squares or the blue

circles are positive” by trial and error. We made sure the colours were at least different from one classification task to another in order to make all stimulus objects appear different.

Results

Replication of L&M's study (Series 300)

A significant difference was found between the concepts 301 (SiSh), 302 (SiCo), and 303 (CoSh), $F(2, 286) = 5.29$, $p = .006$, $\eta_p^2 = 3.6\%$, in the number of blocks required to reach the learning criterion. The number of blocks is shown in Figure 2. When the number of blocks required to reach the learning criterion was calculated using L&M's method (i.e., averaging shape-relevant conditions), we found a significant difference between the shape-relevant conditions ($M = 13.7$, $SD = 8.4$) and the two shape-irrelevant conditions taken together ($M = 11.8$, $SD = 7.2$), $t(143) = 2.24$, $p = .027$. As predicted by L&M, participants were quicker to learn shape-irrelevant concepts. However, learning times differed between conditions SiCo and CoSh, $t(143) = 2.81$, $p = .006$, but not between conditions SiCo and SiSh, $t(143) = 0.65$, *ns*, which does not follow the hypothesis of a hierarchical conceptual organization between shape, colour, and size. However, there was a risk of not obtaining significant results because the standard deviations were high (the distributions were positively skewed because a few participants learned the concepts quite slowly), so we transformed the distributions by taking the natural logarithm of the data. Nevertheless, we again found no significant difference between the SiCo and the SiSh conditions. Such pairwise comparisons were not detailed in L&M's study, but we think they provide valuable information that is also used in the following analyses.

All series

The number of blocks necessary for participants to reach the learning criterion in all of the concept series is shown in Figure 2. The omnibus repeated measures analysis of variance (ANOVA) was significant in all series, $F(2, 156) = 4.6$, $p = .011$,

Table 2. Mean response times measured on the last two blocks for Series 200, 300, 310, 320, and 400 in Experiment 1

| Stimulus set | Mean | SE |
|--------------|------|------|
| 201 | 1.29 | 0.06 |
| 202 | 1.28 | 0.05 |
| 203 | 1.36 | 0.06 |
| 301 | 1.38 | 0.04 |
| 302 | 1.42 | 0.04 |
| 303 | 1.49 | 0.05 |
| 311 | 1.30 | 0.05 |
| 312 | 1.18 | 0.04 |
| 313 | 1.32 | 0.06 |
| 321 | 1.42 | 0.05 |
| 322 | 1.48 | 0.04 |
| 323 | 1.48 | 0.06 |
| 401 | 1.17 | 0.05 |
| 402 | 1.19 | 0.05 |
| 403 | 1.29 | 0.05 |

Note: SE = standard error. Response times in seconds.

$\eta_p^2 = 5.7\%$ for the 310 series, $F(2, 158) = 3.1$, $p = .049$, $\eta_p^2 = 3.7\%$ for the 320 series, $F(2, 130) = 7.7$, $p = .001$, $\eta_p^2 = 10.5\%$ for the 400 series, except for the 200 series for which there was no significant difference between the SiCo, SiSh, and CoSh conditions, $F(2, 152) = 0.4$, *ns*. Using Bonferroni's adjustment ($\alpha = .05/3 = .017$), we found no significant differences between any pair of concepts in the 200 series either. The 200 series might have been too easy to provide enough sensitivity to the putative effects (saliency or nonindependence), but this result might also simply reveal that learning conjunctions are facilitated in the absence of an irrelevant dimension.

When all series were analysed by pairs (within series), there was a significant difference between Concepts 302 and 303, $t(143) = -2.8$, $p = .006$, Concepts 311 and 312, $t(78) = 2.5$, $p = .016$,

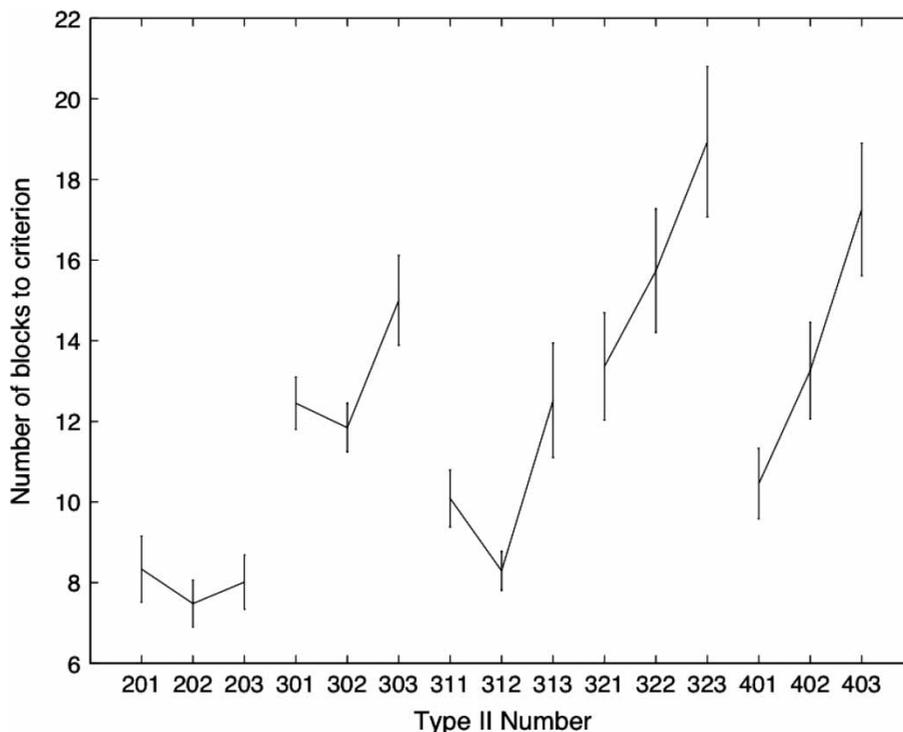


Figure 2. Number of blocks to reach the learning criterion for the five series. Error bars are \pm one standard error.

Concepts 312 and 313, $t(143) = -2.8$, $p = .007$, Concepts 321 and 323, $t(79) = 2.4$, $p = .017$, and Concepts 401 and 403, $t(65) = -4.2$, $p < .001$. Again, taking the natural logarithm of the data did not radically change the significance. Therefore, we observed at least a systematic significant difference in performance between the extreme means within each series (except Series 200). Also, the 303 condition (CoSh/Si'), and other derived versions such as 313, 323, and 403, produced the highest range of values, signalling a difficulty for participants to use a combination of colour and shape as relevant dimensions or a difficulty to inhibit the size dimension, or both.

Note that we observed systematic differences in learning a conjunction of two relevant dimensions in the 310 series, with the irrelevant dimension remaining constant. This is the only condition where the shape-relevant conditions are both more difficult than the size-colour condition. The 310 series therefore reflects the pure effect of the conjunctions of features, without any conditional variations in the irrelevant dimension. This conforms to the hypothesis of nonindependence as well as our hypothesis that participants have difficulties in putting together clusters of different shapes within categories. A completely different pattern emerges in the 320 series, where only the irrelevant dimension was manipulated. In this series, colour and size, respectively, advantaged and hindered learning. More interestingly, when combining the means observed in the 310 and 320 series, where only the relevant and irrelevant dimensions, respectively, were manipulated, we obtain 12, 12, and 16 (rounded to the nearest integer), which broadly corresponds to the means observed for the SiSh/Co', SiCo/Sh', and CoSh/Si' in the 300 series—that is, 12, 12, and 15. This tends to prove that the learning of Type II accumulates two types of difficulties (grouping objects within clusters and grouping clusters within categories), but these two types of difficulty are not necessarily tied to the same dimensions. Another possibility is that nonindependence

accumulates with saliency effects. In the 400 series, the adjunction of a second irrelevant dimension seemed to produce a pattern of results similar to the one observed for the 320 series, meaning that participants had greater difficulties inhibiting the irrelevant dimension than in the 300 series. This seems plausible because greater emphasis was put on the irrelevant dimensions.

The response times given in Table 2 (measured for the last two blocks) were quite similar to the number of blocks required to reach the learning criterion (these two dependent variables were positively correlated when looking at the means: $r = .545$, $R^2 = 30\%$, $p < .05$, $N = 15$; for the whole data points: $r = .08$, $R^2 = .5\%$, $p = .005$). Consistent with the fact that the number of blocks was higher when the series included an irrelevant dimension (the 300, 320, and 400 series, compared to the 200 and 310 series), the response times were higher for the 300 and 320 series; the response times were certainly lower than expected for the 400 series since the response times were averaged on twice the number of examples in the other series, so participants were certainly faster after classifying a greater number of examples.⁵ The important point is that the correlation observed between the two dependent variables indicates that the differences in difficulty are due not only to difficulty encountered by the participants in discovering the classification rules, but also to difficulty in applying these rules (although one could argue that the response times reflect the number of blocks to criterion because participants were more tired after categorizing more stimuli).

Contrary to what was targeted, no consistent patterns of saliency were observed across the conditions. For instance, the saliency that can be inferred from situations where the irrelevant dimension was difficult to inhibit (i.e., size in the 320 series) does not match the hypothesis that participants have difficulties gathering size-relevant clusters within categories in the 300 series (although the use of repeated measures might have introduced some slight variations, as we

⁵ Another possibility is that participants speeded up the classification when the sample of examples was larger (that is, when the progress bar was longer) because they were eager to complete the progress bar.

noted that size was the easiest to inhibit in the 320 series when the analyses were restricted to the first concept learned by participants, while the other patterns of results were similar). This absence of consistency—which does not totally discredit our hypothesis that similarity effects operate in these tasks—is discussed after Experiment 2.

EXPERIMENT 2

Experiment 2 tested whether larger differences in shapes could produce a pattern fundamentally different from the one observed by L&M and those observed in Experiment 1 for the 300 and 310 series (i.e., shape-relevant concepts are more difficult). For the sake of generalizability, we chose to rely on a procedure different from the one used in Experiment 1. The procedure is not based on the basis of sequential learning, in order to avoid any constraint from the presentation order (cf. Mathy & Feldman, 2009). Also, similarity judgements are known to be less pronounced in sequential than in simultaneous comparisons (Palmer, 1978). A simultaneous presentation of the stimuli will allow participants to more easily form clusters depending on the similarity between the examples that are presented together rather than sequentially. The experiment makes use of a restricted set of features in order to assess the effect of the particular aspects of the features. Our hypothesis is that the use of complex shapes can make the shape-relevant > shape-irrelevant pattern very unstable. We hypothesize that the use of complex shapes can be diagnostic of the difficulties participants have in grouping examples in clusters. More specifically, we hypothesize that a pattern opposite to L&M's (shape-relevant < shape-irrelevant) can occur if the difficulty for participants to group different shapes within clusters (in shape-irrelevant concepts) exceeds the difficulty for participants to gather clusters made of different shapes within categories (in shape-relevant concepts). In the case where the difficulty is reversed, the L&M pattern should be more pronounced (shape-relevant >> shape-irrelevant).

Method

Participants

The participants were 50 students at the University of Franche-Comté who received course credit in exchange for their participation.

Procedure and stimuli

A series of classification tasks was presented to participants using a procedure inspired by Feldman (2000), but different in certain aspects. In this procedure, the participants were briefly instructed that they would be required to recall a category of four stimuli in each classification task. The other four stimuli would belong to the concurrent category and should not be recalled. They were told that two categories would be arbitrarily chosen by the computer for each trial.

A set of stimuli was chosen randomly for each classification task (i.e., for each trial). Each set was constructed along three dimensions (shape, size and colour). The first set, the reference/regular set, was built from a combination of three dimensions: shape—vertical rectangles or horizontal rectangles (the shapes were in fact only differing in orientation, so the actual difference in shape was nil); and colour—red or orange; size—half of the stimuli had a surface area twice the size of the others. In a second set of stimuli, the shapes were similar, but the role of colours and size was maximized by increasing their salience (colours—red or blue; size—half of the stimuli had a surface area four times the size of the others). In a third set, the colours and the ratio of the surface areas were similar to those in the first set, but the differences between the shapes was increased by using two complex figures of a different nature (spirals versus blobs, cf. Figure 3). The stimuli were depicted using a black background in a square with a side length of 5 cm. These three stimulus sets are henceforth called regular, shape-min (i.e., shape minimized), and shape-max (shape maximized), respectively.

For each trial, a random Type II concept was generated (by randomly permuting the assignment of diagnostic features) and was applied to one of the three stimulus sets. The SiSh/Co', SiCo/

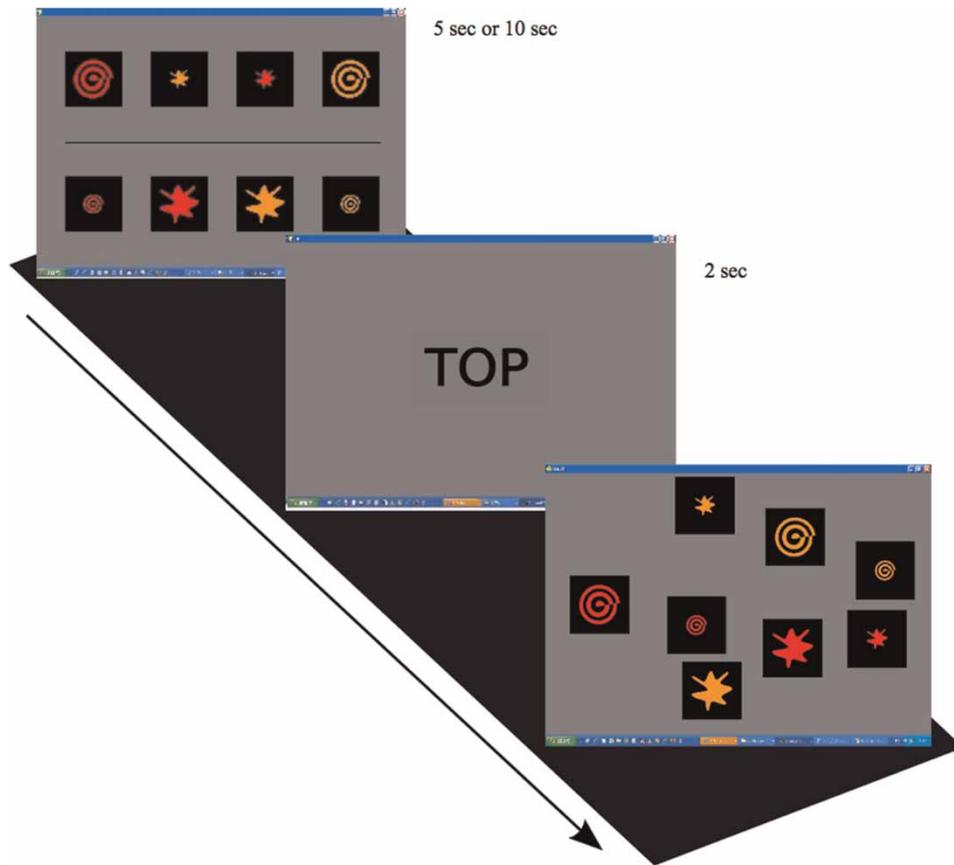


Figure 3. Procedure used in Experiment 2: Training window, target window, and classification window. The white and light grey shapes were red and orange, respectively. To view a colour version of this figure, please see the online issue of the Journal.

Sh', and CoSh/Si' conditions were presented with equal frequency for each stimulus set. Each trial was then divided into three phases (training, cue, categorization). In the training screen (cf. Figure 3), a horizontal line divided the screen into two parts. The positive examples appeared half of the time above and below the line for each condition. For instance, when the positive examples were randomly ordered from left to right and appeared in the upper half of the screen above the horizontal line, then the negative examples were randomly displayed in the lower half (and vice versa).

The stimuli were displayed for 10 s for half of the participants and for 5 s for the other half. The display time was reduced in comparison to

Feldman's (2000) procedure (i.e., 20 s), especially because the task was simplified by having participants induce an abstract rule of a similar kind, whereas the concept type varied from one task to another in Feldman's.

A given stimulus set resulted in 12 different classification tasks. For instance, for a SiCo/Sh' condition applied to the first set of stimuli (the red and orange rectangles of different sizes and orientations), the participants saw the two large rectangles and the two small rectangles on the top of the line in two different classification tasks. However, in one of the classification tasks, participants were asked to recall the stimuli below the line; they also saw the two large

rectangles and the two small rectangles below the line in two other classification tasks, and again participants were asked to recall the stimuli above the line only once. As a result, four different conditions were applied to the three L&M conditions. Therefore, the notions of positive and negative examples were irrelevant for the participants as they were simply told to recall one of two categories. The participants were therefore confronted by 36 classification tasks, randomly ordered for each participant.

Then, a second window, which was shown for 2 s (the target window/the cue), indicated to the participant which category to recall (top or bottom). Half of the time, participants had to recall the bottom category. For that reason, participants were instructed in the tutorial to pay attention to both sets of stimuli (above and below the line).

In a third phase, all the stimuli were displayed randomly on the screen (not sequentially as in Feldman, 2000, but simultaneously, in order to speed up the experiment). The experiment was then self-paced. The participants were asked to use the mouse to click on the four stimuli belonging to the category previously targeted on the screen (top or bottom). The participants were instructed that the category to be recalled would constantly appear in the title bar of the classification window. Each time the participant clicked on a stimulus, a white frame was added to the stimulus. Participants had the possibility of annulling a click by clicking again on the stimulus (in this case, the white frame was deleted). Participants had to press the space bar to validate their selection, but participants could only validate their selection if four figures were selected. The computer then proceeded to the next training screen. Participants were given feedback indicating the number of correct responses on a green background when 100% correct, or on a red background in case of an incorrect classification. The following analyses focus on the number of errors per classification and the response times (the time required to perform the selection of the four stimuli until the space bar was pressed). Because the number of trials is

quite large, we hypothesize that the participants will rapidly formulate an abstract Type II rule and will classify the drawings by processing the visual patterns of the stimuli. Therefore, performance is expected to be due mainly to salience or nonindependence.

Results

Because the display time (5 s vs. 10 s) had a small effect on the results (0.5 more errors in the 5-s condition on average), $t(48) = 2.05$, $p = .046$; $M_5 = 1.42$, $SD_5 = 0.85$; $M_{10} = 0.92$, $SD_{10} = 0.87$, and no significant effect on the response times, the two conditions were aggregated in the following analyses. Figure 4 shows the mean number of errors per condition given the L&M conditions (SiSh/Co', SiCo/Sh', and CoSh/Si') and stimulus set (regular, shape-min, and shape-max); there was an effect of the L&M conditions on the number of errors, $F(2, 98) = 11.37$, $p < .001$, $\eta_p^2 = 18.8\%$, no effect of the stimulus sets, but an interaction between the L&M conditions and stimulus set, $F(4, 196) = 2.75$, $p = .029$, $\eta_p^2 = 5\%$. The inversion of the pattern observed for the shape-max condition is of particular interest. Contrary to the pattern observed by L&M and in our first experiment where shapes were simple, the shape-relevant conditions appear to be simpler than the shape-irrelevant conditions. This tends to indicate that the difficulty encountered by participants in grouping different shapes within clusters exceeds the difficulty encountered by participants in gathering clusters of different shapes within categories.

Also, because the number of errors in the shape-max condition is not significantly lower for the two shape-relevant conditions than for the shape-min and regular conditions, this tends to show that there was no facility for participants to separate the salient shapes into different clusters. Note that difficulty in learning within the SiSh/Co' classifications and within the CoSh/Si' classifications was broadly equivalent between stimulus sets. Differences in colour did not affect the clustering of objects, nor did differences in size or shape affect the formation of clusters

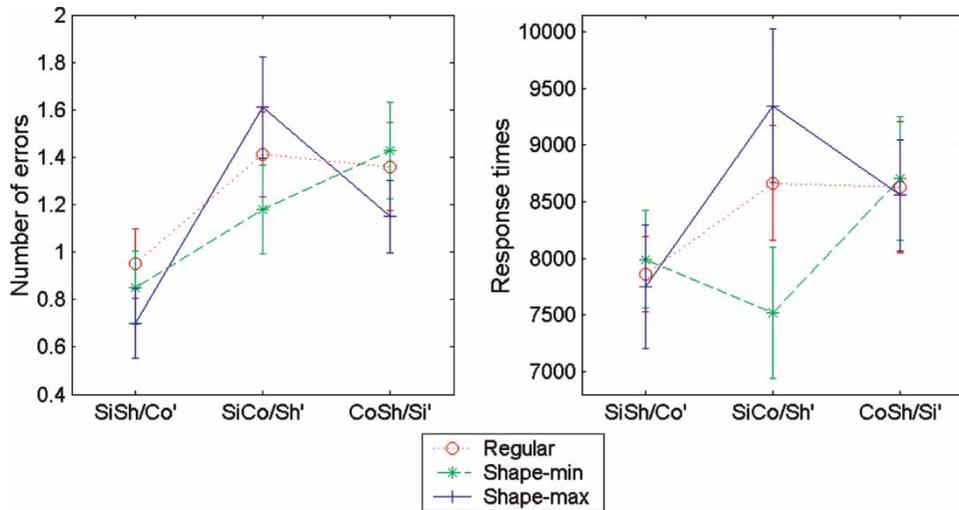


Figure 4. Mean number of errors and mean response times observed in Experiment 2. Regular set: a combination of vertical or horizontal rectangles, red or orange, with half of the stimuli twice as large as the others. Shape-min set: a combination of vertical or horizontal rectangles, red or blue, with half of the stimuli four times as large as the others. Shape-max set: a combination of spirals or blob shapes, red or orange, with half of the stimuli twice as large as the others. To view a colour version of this figure, please see the online issue of the Journal.

within categories in the size–shape-relevant concepts. Similarly, differences in size did not affect the clustering of objects within clusters, nor did differences in shape or colour affect the formation of clusters within categories in the colour–shape-relevant concepts.

Again, the critical differences in the *SiCo/Sh'* condition is of particular interest here. We observed a significant shape-min versus shape-max difference, $t(49) = -2.04$, $p = .046$, meaning that the participants apparently demonstrated a relative difficulty in grouping objects of very different shapes within clusters (the blobs and spirals in the shape-max condition), compared to a facility in grouping objects of similar shapes in the clusters within categories (the horizontal and vertical bars in the shape-min conditions, where the actual difference between the shapes were reduced to nil as they differed only by orientation). The difficulty of clustering within categories does not stand out here, because the clusters in the shape-min condition are more opposed (larger differences in size and in colour—that is, 4 times bigger and red vs. blue) than in the shape-max condition (2 times bigger and red vs. orange), whereas

performance is better. This tends to confirm that difficulties in inhibiting the nonrelevant dimensions prevail.

The graphic on the right in Figure 4 shows a similar pattern for the response times, with a quite high correlation between the mean response times and the mean number of errors, $r = .81$, $p = .008$, $N = 9$, although the conditions were more contrasted (probably because the measure was less coarse than the strict number of errors). Again there was an effect of the L&M conditions on the number of errors, $F(2, 98) = 3.23$, $p < .05$, $\eta_p^2 = 6.2\%$, no effect of the stimulus sets, but an interaction between the L&M conditions and the stimulus set, $F(4, 196) = 3.97$, $p = .004$, $\eta_p^2 = 7.5\%$. We again observed similar difficulties within the *SiSh/Co'* and *CoSh/Si'* conditions and some significant differences within the *SiCo/Sh'* condition. We observed a significant shape-min versus shape-max difference, $t(49) = -2.56$, $p = .014$, as well as a significant difference between the shape-min and regular, $t(49) = -2.28$, $p = .027$. Note that when taking the log of the response times in order to limit positive skewness, the differences were much

more apparent, $t(49) = -3.91$, $p < .001$, and $t(49) = -2.83$, $p = .007$, respectively, which allowed us to reach significance using a Bonferroni correction for three multiple comparisons for the SiCo/Sh' condition ($\alpha = .017$). The difference between the shape-min and regular difference simply means that there was also a gain in identifying the clusters within categories when the salience of colours and surface areas were both increased (since the shapes were identical in the shape-min and regular conditions).

As in Experiment 1, in which the CoSh/Si' was often found more difficult, we again found a significant difference between the SiSh/Co' and the CoSh/Si' conditions for the number of errors, $F(1, 49) = 16.53$, $p < .001$, $\eta_p^2 = 25.2\%$, and for the natural logarithm of the response times, $F(1, 49) = 5.47$, $p = .023$, $\eta_p^2 = 10\%$. Given the low effect of salience on clustering within categories for the whole experiment, this probably denotes a difficulty for participants to inhibit size in CoSh/Si' classifications.

Multidimensional scaling (MDS) was used to study the perceptual dissimilarities between the stimuli within each stimulus set used in Experiment 2. A total of 7 participants different from those in Experiment 1 made similarity judgements about each possible pairing of stimuli within each stimulus set. Participants were presented with pairs of stimuli and were asked to respond with a numerical rating of the degree of similarity between the stimuli (between 1 and 9). These ratings were then used to produce a geometric representation of each stimulus set in which the stimuli were identified with points in a three-dimensional space. For the regular stimulus set, R^2 values ranged from .35 to .84 for the 7 participants, with an averaged R^2 equal to .65 (with $R^2 > .60$ considered acceptable fit). Stimulus coordinates showed that the stimuli differed principally in size and colour, followed by shape, with an overall importance of each dimension equal to .26, .20, and .19, respectively. Results were similar for the shape-min stimulus set (individual

R^2 ranging from .36 to .89; averaged $R^2 = .60$; overall importance of each dimension = .31, .15, .15). For the shape-max stimulus set, participants principally found that stimuli differed by their shape. The shape dimension accounted this time for the larger proportion of variance, followed by colour and size (individual R^2 ranging from .82 to .99; averaged $R^2 = .88$; overall importance of each dimension = .74, .09, .06). Therefore, the differences in shapes (nil vs. maximized) that we targeted in Experiment 2 seem to be confirmed in these independent judgements.

GENERAL DISCUSSION

The dimensions shape, size, and colour are the building blocks of complex object representations. It therefore seems crucial to investigate the degree of independence pertaining to these basic dimensions. On the one hand, a hierarchical organization of these dimensions can be suspected because shapes are most often used as nouns, whereas size and colour are generally used as adjectives. Following this observation, L&M hypothesized that shape is not independent of size and colour. They predicted and observed that it is more difficult to learn Type II classification tasks that combine shape and one of its properties (colour or size) than to learn Type II concepts in which colour and size are combined. To support this interpretation, the authors indicated that their observation is counterintuitive since it contradicts the fact that the higher salience of shapes should help participants separate the stimulus objects in the different categories when shape is relevant. We agree with this argument, which we can develop: For instance, the optimal attention weights for a basic exemplar model such as the General Context Model (GCM), based on the Minkowski metric (Nosofsky, 1984) using city-block distances, are, respectively [.5 .5 0] (considering that the two first dimensions are relevant⁶). Any increase in saliency on the

⁶ The idea is that the participants are inclined to focus on the dimensions that are relevant and to ignore the ones that are irrelevant. A greater attention weight indicates that participants focus more on the dimension and that dimensional values are better

irrelevant dimension makes the Type II concept less learnable. For that reason, when fitting GCM to the mean data points of L&M, GCM states that the shape dimension is the less salient one (incorrectly as L&M's MDS analyses stated otherwise), because the attention weight for that dimension is set to a low value (a high value on the shape dimension would not conform to L&M's results).

On the other hand, similarity is an intuitively compelling explanatory construct (Medin, Goldstone, & Gentner, 1993) that can have more perverse effects in Type II. For that reason, we hypothesized here that similarity can determine how clustering operates within categories and that clustering difficulties can subsequently determine categorization performance. GCM cannot handle such a hypothesis because any increase on the relevant dimensions stretches the distances between all clusters (within and between categories). For any increase in the relevant dimensions, the clusters between categories are better separated (which is supposed to facilitate learning) as well as the clusters within categories (which according to our hypothesis is supposed to make the formation of the classification rule more difficult, as participants have a tendency to put objects that are different into different categories). Unfortunately, because the hypothesis of difficulty for participants to form clusters of different shapes within categories is not directly testable, the present study focused on the possible effect of saliency on the modulation of L&M's observation, especially on the more testable hypothesis that saliency can have an effect on the irrelevant dimension (which conforms to GCM predictions for that matter).

Our objective was to replicate and develop L&M's study to gather an extended set of data in Type II related classification tasks. A specific goal was to control the systematic variation in the irrelevant dimension in the Type II classification tasks studied by L&M (i.e., shape in the colour-size conjunction, colour in shape-size,

and size in shape-colour). In each of the experiments devised in our study, participants were asked to learn concepts in which shape, size, and colour were relevant dimensions and/or irrelevant dimensions. This extension targeted a better distinction between effects due to the hierarchical dependence of these canonical dimensions and effects due to similarity judgements. We particularly hypothesized that the saliency of dimensions could explain both the difficulty for participants to group two clusters of different shapes into a single category in shape-relevant dimensions and that inhibition of the irrelevant dimension could explain or modulate the effects observed by Love and Markman (2003).

Our results provide more direct evidence of the complexity of shape-relevant concepts, especially when the irrelevant dimension was controlled (Experiment 1, Series 310), which can confirm either L&M's hypothesis or our hypothesis that participants have difficulties integrating different shapes within categories.

The classification models in which explicit rules are abstracted from the sample of positive examples provides evidence for why integrating different clusters into the positive category might be more crucial than separating objects of different categories (Bradmetz & Mathy, 2008; Feldman, 2000, 2006; Lafond et al., 2007; Mathy & Bradmetz, 2004; Vigo, 2006). In such models, the minimal rule for a Type II concept is a disjunction of the clusters within categories—for instance (black and square) or (white and circle) for a shape-colour concept—which implies grouping clusters, pruning the irrelevant dimension, but which does not imply having a clear representation of the negative examples. Hybrid models (Sloman, 1996) might also be helpful to account for the complementary effects of rules and similarity computations that our data might reveal. Assuming that participants focus on positive examples when building rules, they are likely to be disturbed when required to consider two subclasses with strong perceptual differences as forming the class

discriminated. The attention weights are constrained to sum to one, which means that more focus on one dimension corresponds to less focus on another one.

of positive examples. The current study tends to demonstrate the greater salience of shape and colour (as suggested by the literature presented in the introduction and Footnote 3), consistent with the finding that the shape–colour concept is the most difficult kind in the 300, 310, and 400 series and in Experiment 2.

The difficulty of the shape-relevant conditions was more apparent in Series 310 than in L&M's experiment replicated here (Series 300). However, our results do not unilaterally confirm the presence of such a hierarchical organization between these dimensions or the single effect of clustering within categories. We do not observe the exact ordering expected by a hierarchical organization of the dimensions in the classical study (i.e., both shape-relevant Type II being more difficult than the shape-irrelevant Type II) in the series 300, despite a very large sample of 144 participants. The effect of the irrelevant dimension (manipulated in the series 320 and 400) might account for the difference between the results of Series 300 and those of Series 310 (the results for the series 310 better confirmed L&M's theory). We showed that combining the means of Series 320 (inhibition manipulated only) and 310 (conjunction manipulated only) produces a pattern of means similar to the one observed in 300 (where both inhibition and conjunction varied). The salience of the dimensions therefore strongly affects participants' ability to inhibit the irrelevant dimensions (which can be translated into difficulty in grouping different objects into clusters). Performance was clearly lowered in Experiment 1 when inhibition was manipulated (Series 300, 320, and 400), and Experiment 2 tended to show that important differences in shapes within clusters hinder the learning of categories. Our results do not invalidate the hypothesis of nonindependence, but the strong effect of salience on the irrelevant dimension tends to support our hypothesis that participants have difficulties in grouping different objects within clusters.

However, this does not totally explain why size alone, which in theory is less salient, was difficult to inhibit in Series 320 (a similar pattern was

also present in Series 400, where inhibition was complicated by a second irrelevant dimension). In Experiment 2, there was also global difficulty for size-irrelevant concepts. There is a possibility that dimensions have different effects depending on whether they are on the focus of attention or whether they need to be inhibited. This certainly relates to the problem of flexibility in similarity judgements observed in other experimental studies (Medin et al., 1993; Murphy & Medin, 1985, p. 296)—that is, the relative weighting of a dimension varies with the stimulus context and task. Similarity effects are perhaps different in forming clusters and separating clusters. To quote Medin et al. (1993), we would add that similarity has to be understood as a process. For instance, the data might appear quite noisy as size seemed more problematic for the participants when they had to inhibit it in Series 320, but this might simply be the results of specific comparison processes: Size might interact in an odd manner with the specific features (frame, hat, gridded, hatched) that we used in that condition. To make an analogy with an example used by Shannon (1988), who stated that the similarity between two nephews can be different depending on which aunts and uncles the nephews are compared with, our study could also reveal that the differences between sizes can be judged differently depending on which dimensions are manipulated during the classification.

Another reason for the apparent diversity of the results might be due to the repeated measures. To address worries about repeated measures, we conducted some supplementary analyses, which revealed that a similar pattern of results holds when only the first problem is included. For instance, when comparing the mean number of blocks to criterion given in Figure 2 and those obtained for the first problem only, we observed a correlation of .64 between the means. Note that the sample sizes were low for the series 200, 310, and 320 (fewer than 10 participants for each of the nine conditions, as the sample size was 80), which might not produce meaningful results. When considering the 300 and 400 series only, where the sample sizes were much more

satisfying, the correlation rises up to .97 and supports the idea that the patterns we observed are quite stable.

Our second experiment suggested a more drastic effect of salience on the learning of shape-relevant versus shape-irrelevant dimensions, in which we observed a completely reversed pattern: Shape-irrelevant classification tasks using more complex shapes turned out to be more difficult. The critic might properly note that there was a sort of "hidden contract" between research participants and experimenters, both of whom considered spirals and blobs very dissimilar (confirmed by our MDS analysis), but we believe that our experimental strategy is justified since the shapes we used, although not canonical, are not that extreme in complexity. Experiment 2 clearly showed a difficulty for participants to cluster together different shapes in shape-irrelevant tasks. L&M could argue that because the shapes used in Experiment 2 were functioning as dimension values, rather than canonical entities (contrary to those in Experiment 1), the shapes in Experiment 1 only could be perceived and processed hierarchically.

Because decision processes alter direct perception, there is a possibility that the Type II classification tasks are a stronger test of the subjective salience of dimensions than the classical similarity judgements most often used. In our experiments, similarity is simply measured indirectly. Participants can be seen as quite passive when performing preferential matching tasks or similarity judgements, compared to the more demanding classification tasks in which the salience of dimensions is subjacent. Consequently, some not critical differences in paired similarity judgements could have more drastic effects during categorization.

We believe that the intradimensional variations that we manipulated had a large impact on performance. Subsequent studies are necessary to better quantify this effect. For instance, size differences were more important in our study than in L&M's. Similarity judgements shall be measured before and after a single Type II classification tasks. Then, the similarity judgements should be submitted to multidimensional scaling so as to

principally measure the effect of prior judgements on categorization (and contingently to measure the effect of categorization on posterior perception). Intraindividual variations in perception and more extensive manipulation of intradimensional variations would certainly help clarify our results and those of L&M. Such similarity ratings are, however, difficult to study a posteriori with different participants and with such a large number of conditions as in the present study (for instance, the important number of colours in Experiment 1 makes the number of paired comparisons too important). We would like to add that trying to account for the results observed by L&M in terms of similarity (widely used in many exemplar models and hybrid models) needs to be taken seriously. The principle of Occam's razor invites researchers to invoke the simplest possible explanations. We believe that similarity is a simpler explanation than a hierarchical organization between the canonical dimensions.

Conclusion and limitations

Love and Markman (2003) predicted that the difficulty in mastering a classification rule could be predicted by the number of predicates that must be unbound in order to free rule-relevant stimulus dimensions. The authors claimed that the difficulty participants have in learning shape-relevant rules was due to the subordination of the colour and size features to shape. Our Experiment 1 replicates their results, especially when the irrelevant dimension was controlled (Series 310). However, the hypothesis of Love and Markman does not seem totally adequate to account for all the data gathered in our study. We tried to show that simple similarity effects can account for a more important portion of the variance in our results. For instance, our Experiment 2 shows that when stimulus dimensions show greater contrasts, the observation of Love and Markman can be totally reversed.

Another important question was whether the hypothesis of Love and Markman (2003) pertains both to rule discovery and category use. Our Experiment 1 shows that the difficulties

encountered by participants in learning certain rules subsist in category use (the number of blocks to criterion was correlated to the response times measured after the learning criterion). This observation is compatible with our hypothesis that the differences in performance are due to the dissimilarities between stimuli within and between the clusters forming the categories.

One can claim that no pattern stands out from our data with respect to the putative salience of the dimensions and that our massive use of repeated measures might have made our study vulnerable to carryover effects. Still, the idea was to free participants from searching for the correct rule and let them search for the relevant dimensions. Also, among the 10 patterns observed in Experiment 1 within each series for the number of blocks to criterion and the response times (for Series 200, 300, 310, 320, and 400), only two different patterns emerged from the data (Condition 2 easier than 1 and 3, or $1 < 2 < 3$). In Experiment 2, the patterns were voluntarily more distorted according to the manipulation of the dimension values.

We observed an odd opposition between an apparent salience of shape or colour when these dimensions were required to form conjunctive rules and an apparent salience of size when this dimension alone needed to be inhibited. We believe that the idiosyncratic aspect of performance in the classification tasks we devised would disconcert other categorization models. A major reason is inherent to Type II classifications: Within-category between-cluster variations is not independent of between-category between-cluster variations. Our ability to discriminate the effects of grouping clusters within categories and separating clusters between categories was therefore severely limited. Future studies need to find a way to distort this strict association in Type II or move on to other classification tasks. To the best of our knowledge, none of the categorization models are able to predict heightened discrimination for stimuli that belong to different categories and decreased discrimination for within-category/between-cluster stimuli when the same dimensions govern both between-category and within-category clusters. A final problem pertains

to the fact that what holds for a cluster of a given category may not hold for another one when the clusters contain opposite values (Love, Medin, & Gureckis, 2004, give the example that there is no characteristic weighting of dimensions for spoons that are composed of large wooden spoons and thinner spoons made of steel; such phenomenon makes modelling the abstraction of clusters a central problem in the categorization literature, cf. Vanpaemel & Storms, 2008).

Perhaps future experiments could investigate the specific relations between shape, size, and colour in situations involving less categorization. L&M's hypothesis is nevertheless very appealing. Features are rarely parcelled out because of causal or structural relations. However, our results do not establish a hierarchical organization of shape, size, and colour as the only possible explanation.

Original manuscript received 29 May 2008

Accepted revision received 29 January 2010

First published online 2 August 2010

REFERENCES

- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15–22.
- Bourne, L. E. J. (1970). Knowing and using concepts. *Psychological Review*, *77*, 546–556.
- Bradmetz, J., & Mathy, F. (2008). Response times seen as decompression times in Boolean concept use. *Psychological Research*, *72*, 211–234.
- Brian, C. R., & Goodenough, F. L. (1929). The relative potency of color and form perception at various ages. *Journal of Experimental Psychology*, *12*, 197–213.
- Bruner, J., Goodnow, J., & Austin, G. (1956). *A study of thinking*. New York: Wiley.
- Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychological Science*, *16*, 291–297.
- Feldman, J. (2000). Minimization of Boolean complexity in human concept learning. *Nature*, *407*, 630–633.

- Feldman, J. (2006). An algebra of human concept learning. *Journal of Mathematical Psychology*, 50, 339–368.
- Fodor, J. (1998). *Concepts: Where cognitive science went wrong*. New York: Oxford University Press.
- Goldstone, R. L. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, 123, 178–200.
- Gureckis, T. M., & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society* (pp. 1876–1881). Washington, DC: Cognitive Science Society.
- Hampton, J. (1993). Psychological models of concepts. In I. van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa, FL: Psychological Assessment Resources.
- Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 11–23.
- Hovland, C. (1966). A communication analysis of concept learning. *Psychological Review*, 59, 461–472.
- Inhelder, B., & Piaget, J. (1964). *The early growth of logic in the child: Classification and seriation* (E. A. Lunzer & D. Papert, Trans.). New York: Harper & Row. (Original work published in French 1959).
- Kagan, J., & Lemkin, J. (1961). Form, color, and size in children's conceptual behavior. *Child Development*, 32, 25–28.
- Kruschke, J. K. (1992). Alcov: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Lafond, D., Lacouture, Y., & Mineau, G. (2007). Complexity minimization in rule-based category learning: Revising the catalog of Boolean concepts and evidence for non-minimal rules. *Journal of Mathematical Psychology*, 51, 57–74.
- Lee, C. L. (1965). Concept utilization in preschool children. *Child Development*, 36, 221–227.
- Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition*, 31, 790–799.
- Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain, a network model of category learning. *Psychological Review*, 111, 309–332.
- Mathy, F., & Bradmetz, J. (2004). A theory of the graceful complexification of concepts and their learnability. *Current Psychology of Cognition*, 22, 41–82.
- Mathy, F., & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review*, 16, 1050–1057.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100, 254–278.
- Medin, D. L., & Schaffer, M. (1978). A context theory of classification learning. *Psychological Review*, 85, 207–238.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289–316.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104–114.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
- Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rules-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition*, 22, 352–369.
- Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, 9, 35–58.
- Palmer, S. E. (1978). Structural aspects of visual similarity. *Memory & Cognition*, 6, 91–97.
- Pitchford, N. J., & Mullen, K. T. (2001). Conceptualization of perceptual attributes: A special case for color? *Journal of Experimental Child Psychology*, 80, 289–314.
- Rosch, E. (1975). Universals and cultural specifics. In R. Brislin, S. Bochner, & W. Lonner (Eds.), *Cross-cultural perspectives on learning*. New York: Halstead Press.
- Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Ryle, G. (1951). Thinking and language. *Proceedings of the Aristotelian Society*, 25(Suppl.), 65–82.
- Shannon, B. (1988). On similarity of features. *New Ideas in Psychology*, 6, 307–321.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75 (13, Whole No. 517).
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22.

- Smith, E., & Medin, D. (1981). *Concepts and categories*. Cambridge, MA: Harvard University Press.
- Spears, W. C. (1964). Assessment of visual preference and discrimination in the four-month-old infant. *Journal of Comparative & Physiological Psychology*, *57*, 381–386.
- Tremoulet, P. D., Leslie, A. M., & Hall, D. G. (2000). Infant individuation and identification of objects. *Cognitive Development*, *15*, 499–522.
- Vanpaemel, W., & Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin and Review*, *15*, 732–749.
- Vigo, R. (2006). A note on the complexity of Boolean concepts. *Journal of Mathematical Psychology*, *50*, 501–510.
- Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.

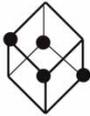
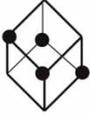
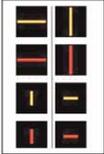
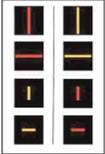
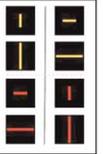
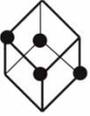
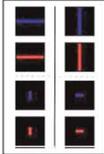
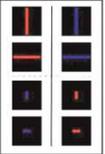
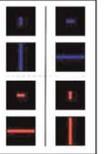
APPENDIX

A detailed version of Table 1: A series of Type II concepts in different contexts
 Table A1. Structures of positive stimuli and samples of stimuli employed during the experiments

| Series | Structure | Concept kind, stimulus set and concept number | | |
|--------|-----------|---|-----------------|-----------------|
| | | <i>SiSh/Co'</i> | <i>SiCo/Sh'</i> | <i>CoSh/Si'</i> |
| 200 | | 201 | 202 | 203 |
| 300 | | 301 | 302 | 303 |
| 310 | | 311 | 312 | 313 |
| 320 | | 321 | 322 | 323 |
| 400 | | 401 | 402 | 403 |

Note: For each sample of stimuli, the positive examples are listed on the left of the solid lines; negative examples are listed on the right of the solid lines; subcategories are separated by dashed lines. Note that from one classification task to the other, the assignment of stimuli to the negative and positive categories was randomly drawn. In Experiment 2, the difference between shapes was increased in the shape-max condition, compared to the regular stimulus set. In the shape-min condition, the role of shapes was minimized by increasing the differences between sizes and colours. To view a colour version of this table, please see the online issue of the Journal.

Table A2. Structures and samples of stimuli employed during Experiment 2

| Series | Structure | Concept kind, stimulus set and concept number | | |
|-------------------------|---|---|---|--|
| | | <i>SiSh/Co'</i> | <i>SiCo/Sh'</i> | <i>CoSh/Si'</i> |
| Exp. 2 Shape- Max |  |  |  |  |
| Exp. 2 Regular |  |  |  |  |
| Exp. 2 Shape- Min |  |  |  |  |

Note: For each sample of stimuli, the positive examples are listed on the left of the solid lines; negative examples are listed on the right of the solid lines; subcategories are separated by dashed lines. Note that from one classification task to the other, the assignment of stimuli to the negative and positive categories was randomly drawn. In Experiment 2, the difference between shapes was increased in the shape-max condition, compared to the regular stimulus set. In the shape-min condition, the role of shapes was minimized by increasing the differences between sizes and colours. To view a colour version of this table, please see the online issue of the Journal.