Université de Montréal

The Role of Exemplar Memory in Rule-Driven Categorization

par

Guy L. Lacroix

Département de psychologie, Université de Montréal

Faculté des Arts et des Sciences

Thèse présentée à la faculté des études supérieures
en vue de l'obtention du grade de
Philosophiae Doctor
en psychologie en sciences cognitives

Août, 2001

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée

The Role of Exemplar Memory in Rule-Driven Categorization

présentée par

Guy L. Lacroix

a été évaluée par un jury composé des personnes suivantes:

Président-rapporteur	ARGUIN, Martin
Directeur de recherche	LAROCHELLE, Serge
Membre du jury	BELLEVILLE, Sylvie
Examinateur externe	BROOKS, Lee R.
Représentant du doyen de la FES	HRIMECH, Mohammed

Thèse acceptée le: 15 février 2002

Summary

Rule-based and similarity-based views of categorization have been traditionally presented as opposing explanations of the same cognitive process. Yet, neither of these views offers a satisfying theory of categorization. This consideration has led to the hypothesis that the two views reflect separate cognitive mechanisms that jointly contribute to the categorization process. The goal of this dissertation is to explore this potential relationship using the rule paradigm devised by Allen and Brooks (1991; Regehr & Brooks, 1993). In this type of categorization experiment, participants first learn to classify exemplars from two categories using a perfectly predictive rule. In the transfer phase, participants are shown critical test items that are similar to training items, but that belong to the opposite category. Brooks and his colleagues showed that these "negative match items" produced higher error rates and longer response times than the other items. They proposed that these negative match effects were due to a conflict between a memory for the training exemplars and the application of the rule. However, this explanation required that participants learn exemplar attributes that were unrelated to category membership (called non-diagnostic) and to which they had paid little or no attention. This strong claim led to our re-evaluation of the role of exemplar-based learning on rule-based categorization. Experiment 1 shows that the non-diagnostic attributes' influence on the categorization of transfer stimuli was minimal. Experiment 2A shows that the time to test the rule attributes determined categorization response times. Experiment 2b shows that a genuine influence of exemplar memory on

categorization could be obtained in an induction task because participants were inferring rules that included the non-diagnostic attributes. Experiment 3 shows that the negative match effects obtained in Allen and Brooks' experiments were due to the use of a salient non-diagnostic attribute to which attention was given. Experiment 4 shows that memory for attributes unique to each individual exemplar can interfere with the application of the rule. Finally, Experiment 5 shows that even attributes that perfectly predict category membership are generally not learned if they are not the focus of the rule. Taken together, these results show that similarity-based effects on categorization are a direct consequence of the attention given to exemplars (and the attributes from which they are built) either because of the application of a given rule or of specific experimental instructions.

Résumé

Les approches portant sur la catégorisation basée sur les règles et celle basées sur la similarité sont traditionnellement perçues comme étant des explications opposées du même processus cognitif. Toutefois, aucune de ces approches n'offre une théorie satisfaisante de la catégorisation. Cette observation a fait naître l'hypothèse que les deux approches sous-tendent des mécanismes cognitifs distincts qui participent conjointement à la catégorisation (Erickson & Kruschke, 1998; Ashby, Alfonso-Reese, Turken & Waldron, 1998; Nosofsky, Palmeri, & McKinley, 1994). Le but de cette dissertation est d'explorer cette relation à l'aide d'un paradigme de classification avec règle qui fut élaboré par Allen et Brooks (1991). Dans ce type d'expérience, les participants apprennent à classer des exemplaires à l'aide d'une règle qui prédit parfaitement l'appartenance catégorielle. Les exemplaires sont bâtis à partir de cinq attributs binaires dont seulement trois sont mentionnés dans la règle. Dans la phase test, on présente aux participants des items critiques créés en modifiant la valeur d'un des attributs de la règle. La nature de la règle et l'existence des deux attributs qui ne sont pas mentionnés dans la règle font en sorte que ces items de transfert sont semblables aux items d'entraînement tout en appartenant à la catégorie opposée. Brooks et ses collègues ont montré que ces «items de transfert négatif» produisaient des taux d'erreur et des temps de réponse plus grands que les autres items. Ils ont donc proposé que ces effets de transfert négatif étaient occasionnés par un conflit entre la mémoire pour les items d'entraînement et le désir des participants d'appliquer correctement la règle. Cependant, cette explication

implique une importante capacité d'apprentissage incident. Pour qu'elle soit valide, il faut supposer que l'apprentissage des exemplaires inclut des attributs inutiles (dit non-diagnostiques) pour déterminer l'appartenance catégorielle et qui, par surcroît, n'ont pas été l'objet de l'attention des participants. Ces suppositions litigieuses ont conduit à notre ré-évaluation du rôle de l'apprentissage des exemplaires sur la catégorisation basée sur l'application d'une règle. L'expérience 1 a repris le paradigme de classification avec règle. Les résultats de cette expérience montrent que l'influence des attributs non-diagnostiques sur la catégorisation des stimuli au transfert est minime même lorsque les essais de pratique sont quatre fois plus nombreux que dans les expériences de Brooks et de ses collègues. Un test de mémoire explicite ajouté après la tâche de classification a montré que les participants étaient incapables de reconnaître les items de la phase d'entraînement. L'expérience 2A a reproduit les résultats de l'expérience 1. De plus, des analyses qui comparent les temps de réponse pour chaque item d'entraînement avec l'item de transfert correspondant ont été menées pour déceler des effets d'exemplaires n'apparaissant pas dans les données moyennées. Ces analyses n'ont produit aucune évidence en faveur de l'hypothèse de Brooks et de ses collègues. Plutôt, elles ont montré que les temps de catégorisation étaient directement liés au test d'attributs compris dans la règle. L'expérience 2B a montré que la mémoire des exemplaires pouvait véritablement influencer la catégorisation dans une tâche d'induction, car les participants infèrent des règles qui incluent des attributs non-diagnostiques.
L'expérience 3 a montré la raison pour laquelle les résultats obtenus dans les expériences 1 et 2A étaient si différents de ceux obtenus par Allen et Brooks. Ces

derniers, voulant rendre la tâche moins artificielle, présentaient leurs stimuli sur des paysages saillants dans le cadre d'une histoire de fond qu'ils racontaient au sujet des stimuli. Ces attributs étaient non-diagnostiques. Toutefois, les directives expérimentales faisaient en sorte que les participants devaient porter leur attention sur ces paysages autant à l'entraînement qu'au transfert. Les résultats ont montré que l'effet de transfert négatif qui a été obtenu par Allen et Brooks était dû à l'utilisation de cet attribut non-diagnostique additionnel. L'expérience 4 reprenait également le paradigme de classification par règle. Cependant, les attributs composant les stimuli étaient uniques à chaque exemplaire. Ce type de stimuli avait également permis à Regehr et Brooks de trouver des effets de transfert négatif. Toutefois, une condition supplémentaire dans laquelle les attributs non-diagnostiques avaient été modifiés au transfert avait fait disparaître ces effets. De nouveau, Regehr et Brooks ont affirmé que ces données supportaient la thèse que les attributs non-diagnostiques exclus de la règle influencent les décisions catégorielles. Cependant, Regehr et Brooks n'ont pas contre-balancé les items vus dans la phase d'entraînement et de transfert. Aussi, les ANOVAs utilisées pour déterminer l'existence d'effets de transfert négatif étaient moins appropriées que celles utilisées par Allen et Brooks. Le but de l'expérience 4 était d'évaluer l'importance de ces problèmes. Elle a montré que le type d'ANOVA utilisé par Regehr et Brooks avait fait apparaître des effets d'exemplaires qui n'étaient pas présents lorsque les analyses appropriées étaient menées. En soumettant les participants à une phase de pratique prolongée, des effets de transfert négatif ont été trouvés, même lorsque les attributs non-diagnostiques étaient modifiés à la phase de transfert. À nouveau, ce résultat s'oppose à la thèse de Brooks et de ses collègues et montre plutôt que la nature idiosyncrasique des attributs avait causé l'effet. Finalement, l'expérience 5 a utilisé un paradigme de classification avec règle. Cependant, la règle comprenait un seul attribut. Six autres attributs constituaient les exemplaires: un attribut qui co-variait parfaitement avec l'attribut mentionné dans la règle et cinq attributs qui co-variaient à 80% avec l'attribut de la règle. Au transfert, l'attribut ciblé par la règle était retiré. Les participants devaient donc tenter de catégoriser les exemplaires à partir des attributs restants. Les résultats ont montré que les participants n'associaient généralement pas ces attributs aux catégories appropriées. Il y avait cependant une exception. Lorsque la couleur co-variait parfaitement avec l'attribut mentionné dans la règle, les participants apprenaient l'association entre celle-ci et la bonne catégorie et ils pouvaient s'en servir pour catégoriser les items au transfert. Sommairement, ces résultats montrent que les effets basés sur la similarité entre les exemplaires sur la catégorisation sont directement liés à l'attention que portent les participants aux exemplaires (et les attributs qui les composent); que cette attention soit guidée par l'application d'une règle donnée ou par des directives expérimentales spécifiques.

Table of contents

Tables

Figures

Dédicace

Je dédie cette thèse à mes formidables et dévoués parents qui ont nourri ma passion pour la connaissance par grande bonté, bien au-delà de toute nécessité; à mon épouse, Marie-Douce, pour son amour et sa patience; et à mon fils, Louis, pour le bonheur qu'il apporte à ma vie.

Avant-propos

«*A wise man, therefore, proportions his belief to the evidence. In such conclusions as are founded on an infallible experience, he expects the event with the last degree of assurance, and regards his past experience as full proof of the future existence of that event. In other cases, he proceeds with more caution: He weighs the opposite experiments: He considers which side is supported by the greater number of experiments: to that side he inclines, with doubt and hesitation; and when at last he fixes his judgement, the evidence exceeds not what we properly call probability* (David Hume, *Of Miracles*, 1985/1748, Part 1, paragraph 4).»

Bien que Hume appliquait ce raisonnement aux croyances populaires et religieuses, je crois que le chercheur sage a tout intérêt à suivre ces conseils, particulièrement dans le cas d'expériences dont les résultats semblent trop fantastiques pour être vrais.

Introduction

When acquiring a new concept or classifying a novel object, it is possible to rely on similarities, general rules or both. For example, upon seeing a strange animal, one might attempt to categorize it by thinking of similar creatures that have been encountered or by recalling rules about animals such as: *if it has feathers, then it is a bird.* These ideas are hardly new. Nonetheless, these two potential ways of classifying objects have traditionally been presented as opposing explanations of the same cognitive process (Murphy & Medin, 1985; Medin, 1989; Komatsu, 1992).

1. Similarity-based views of categorization

On one side, probabilistic views (see Smith & Medin, 1981) have described the categorization problem as a matter of similarity computation. If an object to be classified has enough attributes in common with a given mental representation, then it is classified as a member of that category. For example, because it is generally true that if a creature has a beak and wings, then we may determine that it is a bird. According to some authors, these representations are prototypes (Posner & Keele, 1968; Rosch & Mervis, 1975; also see Smith & Minda, 1998); abstract mental entities created by averaging all encountered members of a given category. For others (Medin & Schaffer, 1978; Brooks, 1978; Hintzmann, 1986; Nosofsky, 1984; 1986), these representations are previously encountered exemplars. Although the debate is ongoing to determine which of these probabilistic approaches is to be preferred (Smith & Minda, 2000), recent favor has gone to exemplar-based models. One reason is that, while these models have been designed to account for the influence of memorized exemplars on categorization, they can also explain the facility that people show when categorizing prototypes (See Medin, 1989). In addition, the formal aspect of exemplar models (Nosofsky, 1984) as well as their kinship to episode-based models of memory (Tulving, 1972, 1983; Ratcliff and McKoon, 1988; Capaldi and Neath, 1995) and to connectionist models (Kruschke, 1992; Nosofsky, Kruschke, and McKinley, 1992) have made the exemplarist approach very attractive. We will at once present this approach through the description of two classic experiments.

Brooks (1978) sought to understand the impact of memorized individuals or exemplars on categorization independently of rule-driven behavior. To do so, he told subjects that they were participating in a paired-associate learning experiment. The first half of the associates were abstract letter strings created using two different miniature Markov grammars as had been done in Reber's classic implicit learning studies (1967, 1976; see Table 1). The second half was either city or animals names that evoked the *Old-World* (Europe) or the *New-World* (America). Naturally, these facts were unknown to the participants. When the criterion of having associated each string with the proper city or animal name once without any mistake was reached, participants were shown 30 new strings: 10 were generated with the first grammar, 10 from the second grammar, and 10 were new. Participants were told to decide to which of these three categories they believed the new string belonged. The results showed that participants were able to correctly identify old category items and new items between 60 and 65% of the time, which is well above the chance level of 33%. Furthermore, they were unable to explain how they had proceeded to successfully sort the test material. The fact that strings created from two different grammars were associated with two different responses suggests that an abstract representation of these grammars (prototypes) had been constructed. Rather, Brooks convincingly argued that participants had properly classified the new items in a non-analytic fashion by recalling previously learned examples of the paired-associate items.

Table 1

Examples of paired-associate learning training items used in Brooks (1978, p. 171).

Stings from grammar A	**Associated animal and city names**	**Stings from grammar B**	**Associated animal and city names**
VVTRXRR	Paris	MRMRTTV	Montreal
VVTRX	zebra	VVT	moose
XMVTTRX	baboon	VVTRTTV	Chicago
VT	Cairo	MRRMRVT	possum
VTRR	tiger	MRRRM	Halifax

Independently, Medin and Schaffer (1978) also set forth the exemplar-based view of categorization in their context theory of classification learning. They contended that "category judgments are based on the retrieval of specific item information; no categorical information is assumed to enter into the judgments independently of specific item information (p. 211)". However, Medin & Schaffer's model made formal predictions. We will present them in the following example.

Consider the abstract description of categories "A" and "B" given in Table 2. This is the notorious 5-4 category structure, which has been used often since then to support exemplar-based models of categorization (it has also been the focus of recent criticism, see Smith and Minda, 2000). Stimuli were built using four binary dimensions.

For instance, if the stimuli were geometric forms, "D1" could be made to refer to color ("0" = red and "1" = blue), "D2" might refer to shape ("0" = circle and "1" = square), etc. In an induction experiment, participants are shown these stimuli one at a time and are asked to assign them to category "A" or "B". At first, classification is random. However, with practice and feedback, performances improve and the categories are learned. To predict the relative difficulty of learning for each item, the context model computes the similarity of each individual training stimulus with every other training stimulus. This is first accomplished by a pair-wise comparison of attributes. A value of "1" is given when the two attributes are identical. Otherwise, a similarity parameter of *d1, d2, d3*, or *d4* used to indicate a difference on one of the attributes. Then, these values are multiplied. The classification difficulty of a particular stimulus is calculated by evaluating the likelihood that the stimulus will evoke a memorized exemplar of the category to which it belongs.

Table 2

The 5-4 categorical structure (Medin & Schaffer, 1978, p. 222).

Training Stimuli									
	"A" Stimuli					"B" Stimuli			
Stimulus number	Dimension values				Stimulus number	Dimension values			
	D1	D2	D3	D4		D1	D2	D3	D4
4	1	1	1	0	12	1	1	0	0
7	1	0	1	0	2	0	1	1	0
15	1	0	1	1	14	0	0	0	1
13	1	1	0	1	10	0	0	0	0
5	0	1	1	1					
Transfer stimuli									
Stimulus number	Dimension values				Stimulus number	Dimension values			
	D1	D2	D3	D4		D1	D2	D3	D4
1	1	0	0	1	9	0	1	0	1
3	1	0	0	0	11	0	0	1	1
6	1	1	1	1	16	0	1	0	0
8	0	0	1	0					

For example, to calculate the likelihood that stimulus 4 will be classified in category "A", we obtain the following equation:

$$\text{Stimulus 4 in "A"} = [(1*1*1*1) + (1*d2*1*1) + (1*d2*1*d4) + (1*1*d3*d4) + (d1*1*1*d4)] \,/\, [(1*1*1*1) + (1*d2*1*1) + (1*d2*1*d4) + (1*1*d3*d4) + (d1*1*1*d4) + (1*1*d3*1) + (d1*1*1*1) + (d1*d2*d3*d4) + (d1*d2*d3*1)]$$

or simply,

$$\text{Stimulus 4 in "A"} = \frac{1 + d2 + d2d4 + d3d4 + d1d4}{1 + d2 + d2d4 + d3d4 + d1d4 + d3 + d1 + d1d2d3d4 + d1d2d3}.$$

The context model predicts that the closer this overall similarity score is to 1, the greater chances are that it will classified in the correct category.

By applying this formula to the 5-4 category structure, one can predict that stimulus 4 will be harder to learn than stimulus 7. This was Medin & Schaffer's key prediction because prototype models suggest the opposite. By definition, prototype models assume that categorization proceeds from a representation derived by averaging encountered stimuli (see Reed, 1972). "1 1 1 1" and "0 0 0 0" are the prototypes for categories "A" and "B". Because stimulus 4 has three attributes in common with the prototype versus only two for stimulus 7, then prototype models must favor stimulus 4. However, the multiplicative rule used in the context model emphasizes inter-stimulus similarity. Therefore, because stimulus 7 has three attributes in common with two members of its category against only one for stimulus 4, we obtain opposing predictions. Medin & Schaffer's results showed that stimulus 4 did yield higher error rates during learning (Experiments 2 and 3), thus supporting their key prediction.

Note that a similar reasoning may also explain why prototypes usually generate low error rates and response times. Because prototypes are obtained by averaging the attributes of all members of a given category, they are the items that have the most attributes in common with the largest number of exemplars. Hence, when a prototype is presented, the probability that an exemplar of the correct category will be retrieved is greater than for any other item.

Lastly, the context model assumes that selective attention may be represented by a saliency parameter for each dimension. For instance, if a given participant pays more attention to color than shape, then the saliency parameter will increase the importance of the first dimension when calculating similarity scores. Medin and Schaffer add "...for tests that can be solved by attending to a single dimension, subjects may have only minimal information to distinguish the individual exemplars (p. 212)". Thus, exemplar similarity comparisons may be strongly influenced by the relative importance of each attribute. This factor is important in calculating quantitative classification predictions for the individual exemplars.

Medin & Schaffer derived such classification predictions. Saliency parameters for each dimension were adjusted to maximize the fit between the classification predictions made by the model and the categorization data. The correlation between the two (which included both learning and transfer stimuli) was $r = 0.81$. The corresponding correlation between classification predictions and the data for the prototype (independent cue) model was $r = .79$. Hence, the context model fit human categorization data as well as a prototype model while also explaining item-specific learning difficulties that eluded its rival. These results were instrumental in making the exemplar-based approach the preferred way of understanding similarity-driven categorization.

Medin and Schaffer's original experiments were followed-up by a series of articles, both theoretical and empirical, by Robert Nosofsky and colleagues (for example, see 1984; 1986; 1988; 2000) in which an expanded context model is developed. This Generalized Context Model (GCM) is based on two insights. First, the multiplicative similarity rule used by Medin and Schaffer can be related to multidimensional scaling theory (Shepard, 1957), which asserts that stimulus similarity is a monotonically decreasing function of the psychological distance between two stimuli. Nosofsky (1984) showed that multiplicative rule is a special case of multidimensional scaling. Second, the context model's response rule is a bias-free derivation of Luce's (1963) Choice theory (For a full discussion of these topics, see Nosofsky, 1984). Relating the context model to these two general theories effectively allowed the exemplar-based approach to be applied to any categorization problem, which was done with a good amount of success.

The generalized context model is the present reference for most work based on a similarity-based view of categorization (but see Ashby and Maddox, 1993). Although the GCM is now more sophisticated than the original context model (mainly through the addition of new parameters, see Nosofsky and Johansen, 2000), its assumptions remain fundamentally the same. Therefore, for the present purpose, it will not be necessary to detail the GCM any further.

2. Rule-based views of categorization

Notwithstanding the success enjoyed by the exemplar-based approach, there are major problems with the unconstrained use of similarity as the basis of categorization. This was put into focus in Murphy and Medin's (1985) seminal article on the role of theories in conceptual coherence. Their thesis is that, on its own, similarity is insufficient to produce a theory of categorization in which concepts will be coherent and meaningful.

All probabilistic views of categorization assume that similarity is the glue that holds concepts together. That is, if two items have a sufficient number of attributes in common, then they will be classified as belonging in the same category. However, it is legitimate to ask which attributes will be included in the similarity calculation and which will be ignored. Indeed, similarity is only useful if one knows which attributes to use when computing similarity. Otherwise, the number of categories to consider becomes limitless, because everything has the potential to be infinitely similar and dissimilar to everything else. For instance, it is not intuitively sensible to place both plums and lawnmowers in the same category (from Murphy and Medin, 1985, p. 292). Nonetheless, it is possible to make both objects highly similar if there are no principled limits on the number and kind of attributes that are counted. Both weight less than 10 000kg (and less than 10 001 kg, 10 002 kg,...), both did not exist 10 000 000 years ago (and 10 000 001 years ago, ...), and so on. Probabilistic views offer no mechanisms that explain why and how only pertinent attributes are included in similarity calculations.

A similar difficulty arises in stating how relevant attributes are correctly emphasized in different contexts. In a classic example (Barsalou, 1983), a set composed of children, jewelry, money, photograph albums and a portable computer may seem disparate until the label *things to take out of your home in case of fire* is applied. Hence, it can be argued that similarity is an empty and useless construct unless one specifies how the information to be entered in the categorization process is selected and weighted (Goodman, 1972). Notice that these difficulties also apply to the first concepts acquired during infancy (Carey, 1985; Keil, 1989; Gopnik, 1988).

Rips (1989) has proposed that there is an even more fundamental problem with similarity-based views by showing a dissociation between similarity and category judgments. Obviously, this statement is incompatible with the basic postulates of probabilistic models. To support his claim, Rips devised an experiment in which he succeeded in making his participants produce opposing similarity and categorization judgments about the same ambiguous objects. To achieve this goal, Rips asked participants to specify extreme values on a given dimension for a pair of categories (for example, the smallest diameter for a pizza and the largest diameter for a quarter). By design, each pair had one category with relatively variables members (i.e. pizzas) and the other with fixed members (i.e. quarters). Then, these two values were averaged to create the ambiguous objects. Hence, if the smallest pizza was 9 cm and the largest quarter was 3 cm, then the ambiguous object was given a 6 cm diameter. Participants thus built their own test material. Finally, Rips asked participants to make similarity and category judgments using questions such as: "Is the 6cm diameter object more similar to quarters or to pizza?" and "Is the 6cm diameter object a quarter or a pizza?" The results were the following. 69% of the participants deemed the object more similar to a quarter (the fixed category). However, 63% categorized it as a pizza (the variable category). From a common sense point of view, this is perfectly reasonable. Even though the object may look more like a quarter, it cannot be one because quarters do not vary in size. However, probabilistic models have no way of handling this problem, because they judge similarity and category decisions to be one and the same. Rips concluded that the dissociation between similarity and category judgments is obtained because participants apply their knowledge about the world to the problems posed.

In fairness, proponents of similarity-based models have acknowledged many of these problems (Goldstone, 1994; Nosofsky and Johansen, 2000) and have attempted to provide explanations by integrating categorization, attention, and learning models (Erikson and Kruschke, 1998; Kersten, Goldstone, and Scahffert, 1998, Kruscke, In press). Yet, Murphy and Medin's arguments retain their full force in the context of real world concept acquisition and categorization where they clearly are one variation of Quine's (1960) problem of induction (also see Macnamara, 1999).

Murphy and Medin (1985; and Medin, 1989) concluded that similarity-based models, by conception, could not provide an adequate theory of categorization. Their incapacity to determine which attributes to include and emphasize in similarity computation in order to explain everyday concepts is central to this failure. Murphy and Medin proposed that knowledge about the world is necessary to explain categorization. These criticisms have led some researchers to turn to knowledge-based views.

The knowledge-based view, also known as the theory theory (Murphy and Medin, 1985, Carey, 1985) asserts that categorization depends on the use of rich and deeply interconnected theories about the world. The analogy between this approach and scientific knowledge (inspired by Thomas Kuhn, 1962) has been made explicit. Scientific as well as naive theories contain rules, postulates, constructs that provide causal explanations about people and objects. The knowledge afforded in these theories dictate what categories are natural and which attributes are important in determining category membership. Hence, our theories about fruits tell us that plums go with apples, oranges, and bananas, and not with lawnmowers. Knowledge-based concepts have the additional advantage that they may be used to explain other high level cognitive processes such as inferential reasoning and the construction of goal-driven categories (see Johnson-Laird, 1983; Barsalou, 1983; Solomon, Medin, and Lynch, 1999).

Nevertheless, knowledge-based views also have problems. Ironically, here too we find a problem of constraints. As Medin (1989) states: “if we cannot identify constraints on theories, that is, say something about why we have the theories we have and not others, then we have not solved the problem of coherence: It simply has been shifted to another level (p. 1475)". Usually, theory theory proponents do not offer a description of their approach that goes beyond that given in the previous paragraph. As was intended by Murphy & Medin (1985), the heuristic form of this theory is a strong argument against similarity-based models. However, unspecified, it is not useful to define concepts (Fodor, 1998).

The question is what are these rule-based concepts? Ideally, the theories would follow the classical approach of concepts (Fodor, 1998; for references on the classical view, see Smith and Medin, 1981; Bruner, Goodnow, and Austin, 1956; Katz, 1972). According to this approach, concepts are abstract representations defined by a set of individually necessary and jointly sufficient attributes. For example, if we take the concept "square". The individually necessary attributes would be "closed figure", "four equal sides" and "four right angles" because they are essential properties of being a square. If one is removed, then the concept is lost and we might obtain concepts such as "triangle" or "rectangle". The three attributes are jointly sufficient because they are unique to squares. Any other attribute such as size, color, or texture is not. As mentioned above, the classical model is ideal. It would yield clear-cut, unambiguous definitions of all categories, as well as allow a powerful model of inference (see Collins and Quillian, 1969). However, it was rightly abandoned because it is implausible. There are many arguments (see Smith and Medin, 1981; Laurence and Margolis, 1998), but two are particularly compelling. First, except for mathematical entities and rare natural language instances (such as bachelor), it is not possible to find sets of individually necessary and jointly sufficient attributes for most categories (Wittgenstein, 1953). Secondly, many experiments have shown that typical attributes for a given category, while being non-necessary, influence categorization judgments (Rosch and Mervis, 1975; Rips, Shoben, and Smith, 1973).

Once the integral version of the classical view is discounted, none of the remaining options are satisfying. First, altered versions of the classical view may be proposed for which either attribute necessity or sufficiency is relaxed. These are the so-called neoclassical views (Laurence and Margolis, 1998; Smith and Medin, 1981). Unfortunately, this theoretical move is not very helpful. Either one emphasizes non-necessary, yet stereotypical attributes, which leads to the problems of the similarity-based view, or one attempts to maintain necessity or sufficiency which leads to the problems met by the original classical model.

Hence, neither similarity-based nor knowledge-based views seem to offer a complete account of the categorization process. Yet, both views appear to contribute separately to our understanding of the phenomenon. As we have seen, rule-governed theories cannot be ignored in understanding categorization and their importance in theories of cognition in general is well documented (Anderson, 1983; Fodor, 1975; Fodor and Pylyshyn, 1988). Research has also shown that similarity is a useful construct for understanding concept acquisition (Smith & Samuelson, 1998) and categorization (Goldstone, 1994; Smith & Sloman, 1994). These considerations have elicited a new look on the role of these theories. Rather than considering them as rivals, many researchers now hypothesize that the two views reflect separate cognitive mechanisms that contribute to the categorization process.

3. Combined views of categorization

Since the publication of Murphy and Medin's (1985) article, much attention has been given to the relationship between similarity-based and rule-based models of categorization (Waldron & Ashby, in press; Erickson & Kruschke, 1998; Ashby, Alfonso-Reese, Turken & Waldron, 1998; Smith, Palatano, & Jonides, 1998, Nosofsky, Palmeri, & McKinley, 1994; Smith and Sloman, 1994; Regehr & Brooks, 1993; Allen & Brooks, 1991; Rips, 1989). Broadly construed, the idea is that categorization involves two separate, but interacting, mechanisms. The first system relies on similarity and the second system employs more general classification rules. However, proposals as to how these systems interact are few and tentative, and empirical data showing a clear relationship between the two are limited.

Nosofsky, Palmeri and McKinley (1994) put forward the rule-plus-exception model (RULEX). They postulated that category learning is a search for single-dimension rules guided by diagnostic and salient attribute values. First, people search for a perfect one-dimension rule. Then, if failure occurs, imperfect one-dimension rules or conjunctive rules are sought, and the exceptions are memorized. Nosofsky et al.'s intent was to show that RULEX could account for empirical phenomena covered by exemplar-based models (in particular, the GCM) while explaining the propensity for idiosyncratic rule-based category learning behavior. Both goals were achieved. For many different category structures, the implemented version of RULEX predicted correct response probabilities that correlated with observed probabilities and with the response probabilities generated by the GCM. Also, RULEX accounted for 86% of the variance in the performance of the individual participants learning the 5-4 category structure of Medin and Schaffer (1978). Because the GCM faired much worse than RULEX for this type of analysis (with only 36% of the variance explained), Nosofsky et al. concluded that a complete model of category learning had to include rule-based and similarity-based components.

Erickson and Kruschke (1998) proposed a hybrid connectionist model named ATRIUM, which combines rule-based and exemplar-based modules. The rule module learns to draw categorical boundaries in the psychological space that evolves through contact with individual items and the exemplar module is a connectionist implementation of the GCM (see the ALCOVE model in Kruscke, 1992). ATRIUM predicts that the rule module will be used for categorization except in cases for which the to-be classified object is similar to an exemplar that violates the rule. To test their model, Erickson and Kruschke showed their participants rectangles that varied on two dimensions: height and location of a mark on the base. With two exceptions, all the stimuli could successfully be classified into two categories using a rule that considered the primary height dimension only. They are represented by the filled squares and circles in Figure 1. The outlined shapes represent the two exceptions. To successfully

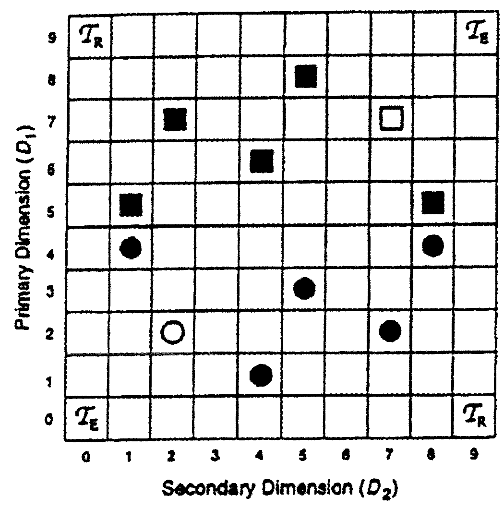


Figure 1. Schematic description of Erickson and Kruschke's (1998) Experiment 1. The filled squares and circles are instances of the two training categories described by a primary dimension (rectangle length) and a secondary dimension (location of a mark within the rectangle). The outlined figures are the two training instances that cannot be classified using only the primary dimension. The "T_E", "T_R" and empty cells were transfer items. Note-From "Rules and exemplars in category learning," by M. A. Erickson and J. K. Kruschke, 1998, *Journal of experimental psychology, 127*, p.110. Copyright 1998 by the American Psychological Association.

categorize these items, attention must be given to both the rectangle height and the base marks. Participants were trained to learn these categories through induction. In the transfer phase, test items (all the blank cells in Figure 1) plus two critical types were presented: new items that were similar to the training exception ("T_E" in Figure 1) and new items that simply followed the rule ("T_R" in Figure 1). Erickson and Kruschke hypothesized that following an exemplar-based account of category learning, T_E items should yield more incorrect responses due to similarity with an exception. However, a rule-based account predicts that participants will extrapolate category membership using height information and that there should be no difference in exception responses. In other words, when novel items are outside the range of previously seen exemplars, they may escape the effect of memorized exceptions and fall under the influence of the rule. This is what Erickson and Kruschke found. Furthermore, in agreement with exemplar-based models, they found that the probability of obtaining incorrect responses was greater for the other test items if they were similar to the exception on the primary dimension. Hence, Erickson and Kruschke argued that their results supported their two module account of category learning.

Ashby, Alfonso-Reese, Turken, and Waldron (1998) presented a neuropsychological category learning theory with multiple systems called COVIS (competition between verbal and implicit systems). Although Ashby et al. described Yheir model at the neurological and implementation levels, we will concentrate solely on the global behavior of the systems. The first system is verbal (rule-based). It implements the explicit, conscious effort to find rules that is particularly active early in category learning. This system learns which verbal (one-dimension) rule is the most accurate for classifying the stimuli and eventually, all category decisions come to focus on this dimension only. The second system is non-verbal, being based on procedural learning. It is supposed that with time and experience, this system learns to associate category responses with different regions of psychological space to create optimal decision boundaries (see Ashby et al. 1998, p. 445). In opposition with RULEX and ATRIUM, COVIS is not postulated to access exemplar memory during categorization. Finally, COVIS assumes that both systems compete to deliver categorization responses

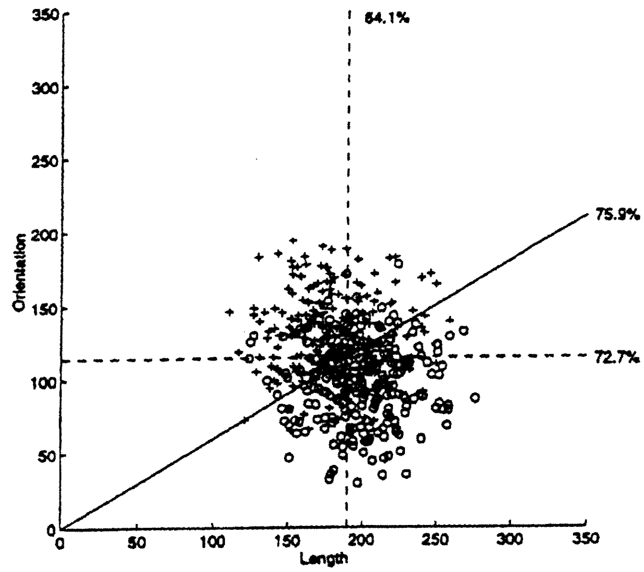


Figure 1. Schematic description of Erickson and Kruschke's (1998) Experiment 1. The filled squares and circles are instances of the two training categories described by a primary dimension (rectangle length) and a secondary dimension (location of a mark within the rectangle). The outlined figures are the two training instances that cannot be classified using only the primary dimension. The "T_E", "T_R" and empty cells were transfer items. Note-From "Rules and exemplars in category learning," by M. A. Erickson and J. K. Kruschke, 1998, *Journal of experimental psychology, 127*, p.110. Copyright 1998 by the American Psychological Association.

and the most accurate wins. Even though with practice, this competition favors the implicit system because of its greater accuracy, the verbal system still occasionally responds. Hence, COVIS predicts that the verbal system will bias responding and lead to sub-optimal performances. To show this interaction, Ashby et al. conducted an induction experiment in which participants categorized lines that varied on length and orientation on 2000 trials. These stimuli are represented by plus signs (category A) and circles (category B) in Figure 2. The dotted lines represent the optimal rule-based decision boundaries. Distribution of the stimuli made the optimal rule based yield a maximum categorization accuracy rate of 72.7% when the rule was based on length and of 64.1% when it was based on orientation. The full line shows the optimal boundary for the implicit system and it yields a categorization accuracy of 75.9%. COVIS predicts that learning should gradually bring a conflict between the verbal rule and the optimal implicit decision bound. Hence, accuracy should tend away from 75.9%, thus showing bias towards the rule. This was confirmed by the data. Because Ashby et al. believe that the GCM predicts performances that would be comparable to those generated by their implicit system, the authors conclude that the results are incompatible with single system accounts of category learning.

However, Nosofsky and Johansen (2000) have criticized all the models describe so far (including RULEX). They claim that all the results accounted by these multiple-system models can be predicted and explained by the latest version of the GCM. The data reported in support of ATRIUM (Erickson and Kruschke, 1998) was attacked on methodological grounds and GCM was shown to generate predictions as precise as RULEX and COVIS when care was taken to adequately simulate variations in the saliency parameters. Hence, the experimental data in support of a two systems account of category learning are at best tentative.

A series of categorization experiments which explore the same questions and that have escaped criticism are found in Allen & Brooks (1991) and Regehr & Brooks (1993). Somewhat like Ashby et al. (1998), Brooks and his colleagues postulate that category decisions involve two competing systems: one is rule-based and the other is

Table 3

Logical description of the stimuli used by Allen and Brooks (1991, Experiment 1).

	Old items (training phase)						Match items (test phase)						
	Item number	Body shape	Spots	Leg length	Neck length	Number of legs	Item number	Body shape	Spots	Leg length	Neck length	Number of legs	Background
Positive													
	1	1	1	1	0	0	10	1	0	1	0	0	4
	3	1	0	1	1	1	9	1	1	1	1	1	2
	6	0	1	0	1	1	15	0	0	0	1	1	1
	8	0	0	0	0	0	13	0	1	0	0	0	3
Negative													
	2	0	1	1	0	1	14	0	0	1	0	1	1
	4	1	1	0	1	0	16	1	0	0	1	0	3
	5	1	0	0	0	1	11	1	1	0	0	1	2
	7	0	0	1	1	0	12	0	1	1	1	0	4

Note. Each stimulus may be described by a combination of five binary attributes. The attributes "body shape", "spots", and "leg length" determined category membership while "neck length" and "number of legs" was non-diagnostic. Transfer items were created by changing the value of "spots". In the case of negative match items, this yielded stimuli that were highly similar to training phase items while belonging to the opposite category. All stimuli were presented on one of four backgrounds which were also non-diagnostic.

exemplar-based. To support this hypothesis, Brooks and his colleagues used two category learning paradigms. First, they used induction experiments to show that the categorization of transfer items may be influenced by previously classified exemplars. As discussed in the previous sections, this is a well known result. The original aspect of their research was a complementary condition in which participants learned to classify exemplars from two categories using a rule.

In the training phase, participants were given a perfectly predictive classification rule to distinguish two types of fictional animals. These animals were built using five binary attributes: body shape, spots, leg length, neck length and number of legs and a sixth attribute, background, which had four possible values (see Figure 3). Notice that attribute values could be implemented in only one way. For instance, all exemplars with a "curved body" had a physically identical curved body. The abstract description of the categorical structure is given in Table 3.

One of the rules stated that if an animal had two or three of the following attributes: long legs, angular body and spots, then it was a *builder* (all rules used are given in Table 4). Otherwise, it was a *digger.* These special names, along with the

Table 4

The rule types used by Allen and Brooks (1991, Experiment 1, p. 5 and 6).

Rule types
(a) Long legs, angular body, and spots
(b) Short legs, angular body, and no spots
(c) Short legs, curved body, and spots
(d) Long legs, curved body, and no spots

backgrounds, were part of a story given to the participants in order to give the task a certain amount of ecological validity. The three attributes that make up the rule (the first three in the abstract description in Table 3) are called diagnostic because they are logically necessary and sufficient to properly categorize the animals. The last three are non-diagnostic: they are not correlated with category membership.

Each trial was composed of three slides, in which the fictional animals were seen building or digging. On the presentation of the first slide, participants were told to categorize the creatures as quickly as possible without sacrificing accuracy. Then, they

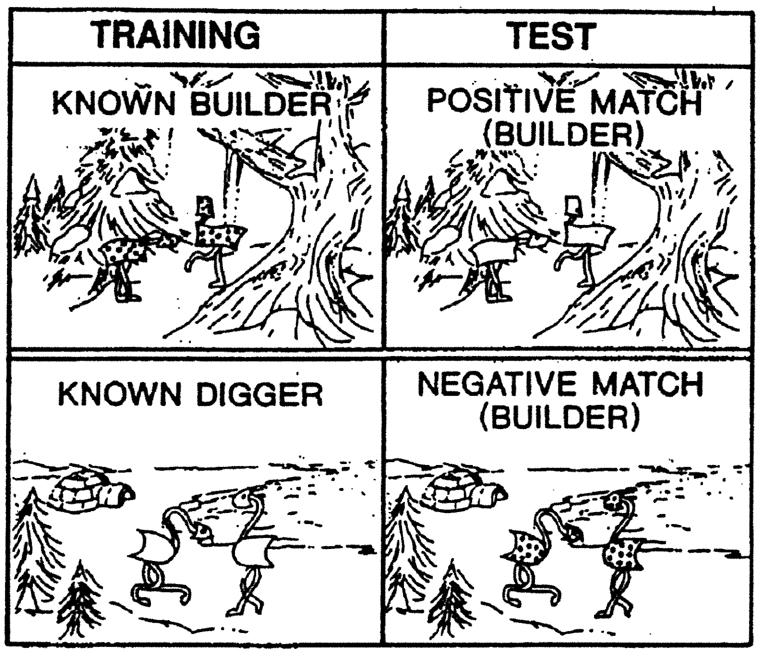


Figure 3. The stimuli used by Allen and Brooks (1991) in Experiment 1. . Note- From S. W. Allen, & L. R. Brooks, 1991, *Journal of experimental psychology: General, 120*, p. 4, Copyright by the American Psychological Association.

were told to observe how the creatures either built or dug on the last two slides. In all, the eight training creatures were presented randomly in five successive blocks. Hence, participants saw a total of 15 slides depicting each particular exemplar.

In the test phase, two types of transfer items were presented. All were highly similar to a corresponding training item. As can be seen in Table 3, each transfer item was obtained by reversing the value of the second attribute, the spots, while the other attributes remained the same. Notice that this manipulation makes it possible to have certain transfer items that are very similar to some training items and yet belong to the opposite category. These were called negative match items. By contrast, transfer items that belonged to the same category as their most similar training item were called positive match items. Allen and Brooks wanted to minimize the possibility that

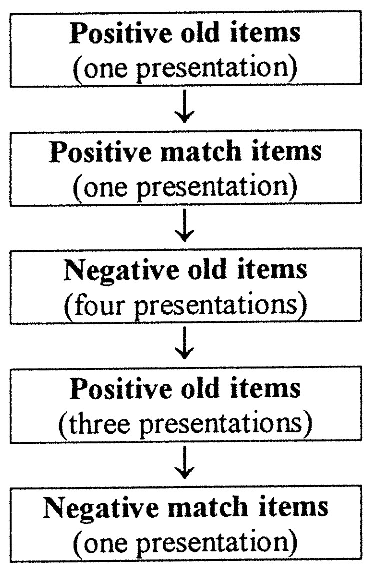


Figure 4. Schematic representation of the presentation order used by Allen and Brooks (1991) in the transfer phase of experiment 1. In order to avoid participants realizing that match items were being used, presentation order was not randomized in the transfer phase. Rather, items were presented in a specific sequence which separated positive and negative match items.

participants realize the presence of match items in the test phase. Therefore, they used a special presentation order in which match items were shown in separate phases with old items placed in between (see Figure 4).

Allen and Brooks hypothesized that there would be a conflict between the application of the rule and the memory of previously categorized exemplars in classifying the negative match items. In other words, for these items, the rule is saying:

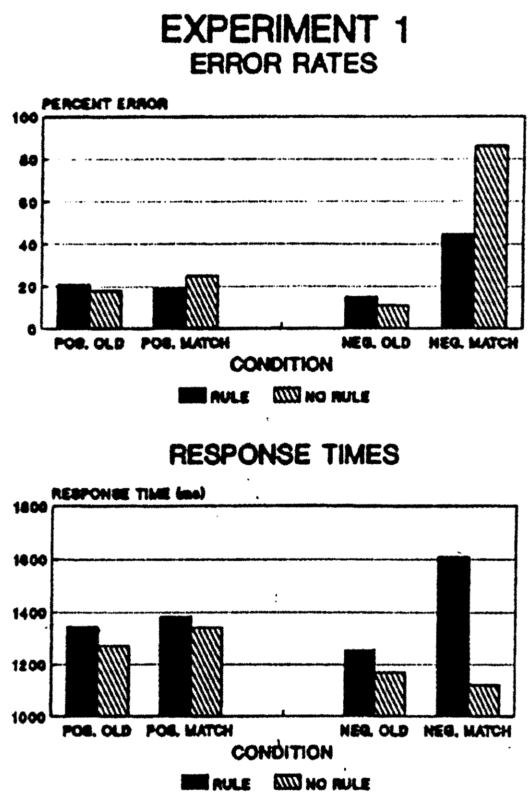


Figure 5. The results for Experiment 1 from Allen & Brooks (1991). The dark lines show the data in the rule condition. The high error rates and response times obtained for negative match items in comparison to negative old items were emphasized to support the hypothesized conflict between rule-based and similarity-based mechanisms of categorization. Note- From S. W. Allen, & L. R. Brooks, 1991, *Journal of experimental psychology: General, 120*, p. 7. Copyright by the American Psychological Association.

"The creature is from this category" and the similarity between the match item and its corresponding old item is saying: *"No, this looks like a creature from the other category"*. Hence, a conflict appears. Allen and Brooks predicted that this would

translate into more errors and longer response times for negative match items. This is exactly what they found. In their Experiment 1 (see Figure 5), error rates were 45% for negative match items compared with errors rates close to 20% for both training and positive match item. Similarly, response times were close to 1600 ms for negative match items compared to response times ranging from 1200 to 1400 ms for the other types of items. The authors' interpretation of this "negative match effect" was that the exemplars seen during the learning phase had been memorized and that the similarity between the test and the training items subsequently caused the memorized exemplars to be retrieved and to influence the categorization process even though a perfectly predictive rule had been mastered.

In a follow-up experiment, Regehr and Brooks (1993) varied the composition of the stimuli in order to explore further the relationship between memorized exemplars and the application of a categorization rule. The goal was to show that holistic individuation, which is the degree to which features cohere into an individuated whole (a Gestalt), had priority over feature uniqueness.

Using the same categorical structure as Allen & Brooks (1991), four sets of stimuli were made (see Figure 6). These sets varied on two dimensions. First, the individual features that composed the stimuli were either interchangeable or individuated. If the features were interchangeable, there was only one possible physical implementation for each value of a given logical attribute in the categorical structure as was the case in Allen and Brooks' Experiment 1. If the features were individuated, each exemplar had a unique physical implementation for each value of a given logical attribute. For instance, each creature that had a curved body type had its own unique kind of curved body type. Second, the impression of the whole item was either composite or individuated. If they were composite, all the stimuli looked alike, they were without personality. If they were individuated, the combinations of the features resulted in creatures that had a clear Gestalt, an individuality, a personality. Regehr & Brooks hypothesized that negative match effects (as found in Allen & Brooks) were driven by a conflict between the classification rule and the memorized exemplars as wholes and were not due to the conflicting effect of any given particular attribute.

To test this prediction, Regehr & Brooks (1993) used the Allen & Brooks (1993) paradigm with all four stimuli sets (Experiments 1D, 2A, and 2B). However, the backgrounds were eliminated and the transfer stimuli were presented in random order. All story slides that allowed the participants to observe the creatures in their ecological setting were dropped. Once more, participants were given a three attribute rule that was perfectly predictive of category membership to classify the creatures. After 40 learning trials, participants saw transfer items that were identical to a corresponding training item except for the "spots" attribute whose the value was reversed. Regehr and Brooks called these items "bad transfer items" instead of negative match items. Similarly, they called positive match items "good transfer item". However, for the sake of simplicity, Allen and Brooks' "positive match" and "negative match" labels will be used throughout the dissertation to refer to the transfer items.

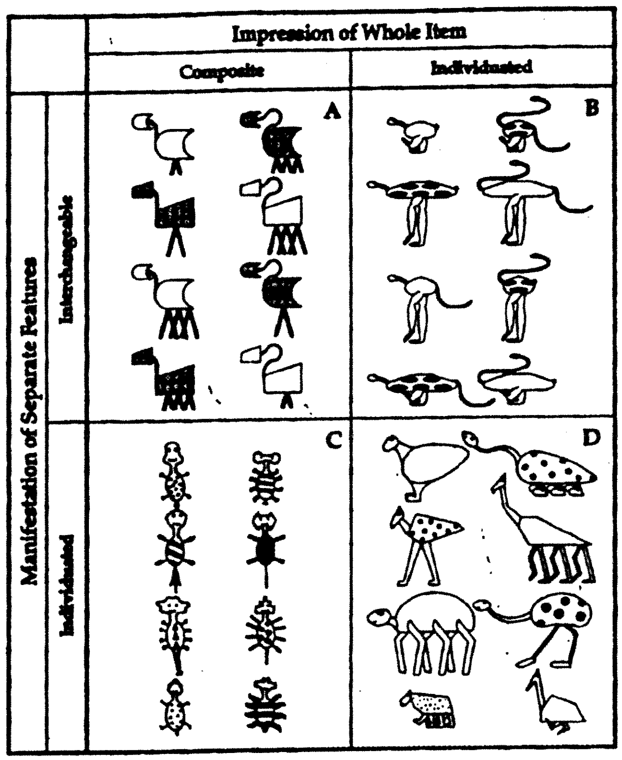


Figure 6. The stimuli used in Regehr and Brooks' (1993) Experiments 1D, 2A, and 2B. Creatures in the left panels are composite, that is they have no distinct personality, no gestalt whereas the creatures in the right panels do. The creatures in the top panels have interchangeable features, that is there is one physical implementation for each attribute value. Those in the bottom panels have idiosyncratic features, i.e. they are unique to each individual. Note- From G. Regehr, & L. R. Brooks, 1993, *Journal of experimental psychology: General, 122*, p. 94. Copyright by the American Psychological Association.

In general, the results supported Regehr and Brooks' hypothesis (see Table 5). For the individuated items with interchangeable feature condition, which replicate Allen & Brooks' (1991) Experiment 1, a negative match effect was found for median response times. For individuated items with individuated feature condition, a negative match effect was found for both median response times and the proportion of errors. These two first conditions contrast with the two last conditions in which exemplars were not individuated. In these conditions, negative match effects were not found. Hence, Regehr & Brooks concluded that holistic individuation, the Gestalt, was a necessary element to obtain exemplar-based transfer effects.

Table 5.

Pooled results of Experiments 1D, 2A, and 2B in Regehr and Brooks (1993, pp. 102,104 & 105) for error rates and response times.

	Stimuli types			
	Composite form		Individuated form	
Feature type	Interchangeable (A)	Individuated (C)	interchangeables (B)	Individuated (D)
Measure				
Proportion error				
Old	.031	.063	.052	.094
Positive match	.047	.078	.063	.109
Negative match	.047	.078	.083	.328
Median response times				
Old	863	903	909	894
Positive match	894	911	904	1018
Negative match	942	928	1136	1295

However, there is a problem with this last interpretation. Because the rule focused the participants' attention on center of the creature, then the negative match effect could have been due to the distinctiveness of the two unchanged attributes (the body and the legs) of the transfer items. To test for this possibility, Regehr and Brooks replicated their rule paradigm with creatures that had individuated features and that formed a Gestalt (Experiment 3A). More importantly, they added a condition

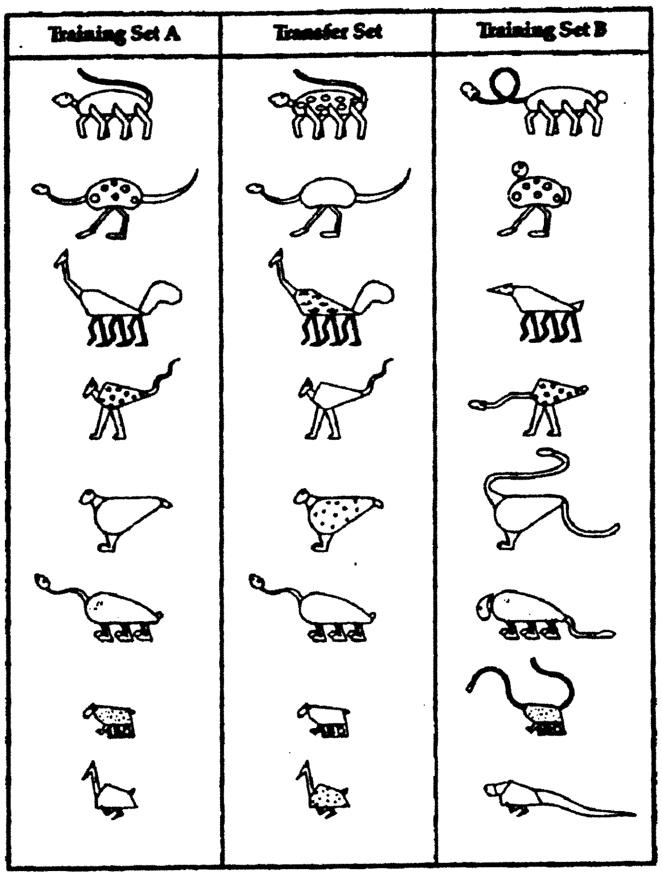


Figure 7. The stimuli used by Regehr and Brooks (1993) in Experiments 3A and 3B. The right and left columns show training stimuli and the central column show transfer stimuli. It can be seen that for training set A, the individuality of each stimulus is preserved when compared with corresponding match items. However, individuality is lost when training set B is used. Note- From G. Regehr, & L. R. Brooks, 1993, *Journal of experimental psychology: General, 122*, p. 107. Copyright by the American Psychological Association.

(Experiment 3B) in which the non-diagnostic attributes (the neck and the tail) were modified for transfer items, thus modifying the Gestalt (see Figure 7).

The results once more supported the importance of holistic individuation for exemplar-based transfer effect. As seen in Table 6, clear negative match effects were found for both error rates and response times when the Gestalt was conserved. However, when it was modified, effects for both these variables were lost.

4. A critical analysis of rule-based category learning and exemplar effects

Brooks and his colleagues' rule paradigm provides an insightful task to study the relationship between rule-based and exemplar-based mechanisms of categorization. The categorical structure used is nicely balanced and yields logically equivalent categories. Hence, any exemplar-based effect found cannot be due to a poor structure (in opposition, for instance, to the 5-4 category structure which may favor exemplar learning, Smith and Minda, 2000). Also, the rule paradigm appears to be a strong tool to investigate the possible interaction between rule-based and similarity-based categorization mechanisms. Because the categories are well defined by the classification instructions given at the beginning of the experiment, individual differences in the application of the rule are minimized and exemplar learning will be incidental by definition, if it occurs. Clearly, this is not the case with induction tasks which allow many category learning strategies, both rule and exemplar-based, to coexist within and between subjects (Nosofsky et al. 1994). Therefore, the rule paradigm presents a controlled way of studying a multiple-system hypothesis of categorization.

However, Brooks and his colleagues' results are surprising. Their main goal was to show that memorized exemplars could influence the application of a clear, well practiced, categorization rule. The higher error rates and longer response times they

Table 6

Pooled results of Experiments 3A and 3B in Regehr and Brooks (1993, p. 108) for error rates and response times

	Stimulus type	
	Form preserved	Form lost
Measure		
Proportion error		
Old	.102	.039
Positive match	.125	.047
Negative match	.344	.063
Median response times		
Old	811	1194
Positive match	875	1237
Negative match	1300	1271

found for negative match items appear to support this hypothesis. Yet, Allen and Brooks' explanation of the phenomenon requires that the exemplar-based mechanism of categorization have a very important capacity for incidental learning. It must also influence categorical decisions even when such decisions do not require any exemplar information. In the following, we discuss these strong pre-requisites.

The theoretical implications of Brooks and his colleagues' interpretation of the negative match effect.

To obtain a negative match effect, training phase items must first be memorized. This entails learning representations of the exemplars that include the three rule attributes and two non-diagnostic attributes. This is essential. Otherwise, maximal similarity between old and match items (all attributes identical except spots) is not obtained (see Table 3). Therefore, it is clear that the non-diagnostic attributes play a crucial role in generating the negative match effect. Furthermore, learning the non-diagnostic attributes must be incidental by definition. The experimental instructions were to use the rule attributes to classify the animals and to examine the animals' activities (whether they were building or digging). Nothing in the procedure required participants to actively try to memorize the animals, or to attempt to learn the non-diagnostic attributes which were not correlated with the categories. Note also that the non-diagnostic attributes used in Brooks and his colleagues' experiments were local features (i. e. they did not span the entire stimulus). For example, in Allen and Brooks' Experiment 1 the non-diagnostic attributes were neck length and number of legs. Hence, because the rule involved leg length, body shape and spots, it was possible to accomplish the categorization task without focusing any attention on at least one of the non-diagnostic attributes (i.e. number of legs). Finally, for any given exemplar, the learning had to be completed in just five trials. It is true that each trial was composed of three slides, in which the fictional animals were seen building or digging. Hence, participants actually saw 15 slides depicting each particular exemplar. Nonetheless, these complex stimuli had to be learned fairly quickly.

Hence, Brooks and his colleagues proposed a multiple-system categorization model with a very powerful exemplar-based mechanism. This system learns exemplars quickly, incidentally, and without regard to the diagnostic value of the features. Then, the memory trace of these exemplars is assumed to be so powerful that it can not only slow down decision times, but it can also lead people to wrongly apply a perfectly predictive and practiced categorization rule almost half the time. This is a strong claim. Yet, many authors have accepted Brooks and his colleagues' results and interpretations without reserve. Their study has been cited in different reviews on categorization (Goldstone, 1994; Hahn & Chater, 1996; Goldstone & Barsalou, 1998). The paradigm has been used to explore the neural correlates linked to rule application and exemplar similarity (Smith, Patalano & Jonides, 1998) and it has been used with children (Wagner & Allen, unpublished manuscript). We believe that such claims, and the experimental data upon which they are based, deserve a thorough analysis.

The assumptions underlying Brooks and his colleagues' multiple-system model of categorization are quite unique in the literature. We do not think that any of the categorization models reviewed would make the prediction that exemplar learning proceeds in the fashion described. Pure rule-based models and prototype models are irrelevant here because they do not postulate exemplars memorization. Exemplar models, such as the GCM (Nosofsky, 1986), would predict that non-diagnostic attributes to which little (or no) attention is given would not be given much weight in the similarity calculation because they have no cue validity. Hence, their influence on category decisions would be negligible. The multiple-system accounts lead to the same conclusion. RULEX (Nosofsky et al., 1994) would not predict any effects caused by memorized exemplars because the model predicts that this type of learning only occurs if any given item does not conform to the rule. ATRIUM (Erickson and Kruschke, 1998) would not predict any exemplar-based effects based on non-diagnostic attributes because the exemplar-based module, a connectionist implementation of the GCM, would not include these attributes in the representation of the exemplars. Finally, ALCOVE (Ashby et al., 1998) would not yield Brooks and his colleagues' prediction because the rule-based decision bounds and the optimal decision bound would both exclude the non-diagnostic attributes. Hence, Brooks and his colleagues' interpretation of the negative match effect does not fit in any of the prevailing exemplar or multiple-system models, contrary to the claims often made by the proponents of these models.

Brooks and his colleagues' assumptions may also be critically analyzed from an implicit learning perspective. As previously mentioned, the utilization of a categorization rule supposes incidental learning of the non-diagnostic attributes responsible for the negative match effect. Support for this view should therefore be found in the implicit learning literature. However, this is not the case.

Two facts about implicit learning are related to our present concern. First, implicit learning generally takes time. Second, some degree of attention is necessary for implicit learning to occur (Goschke, 1997; Cleeremans, Destrebecqz, & Boyer, 1998; but see Reber, 1989).

Implicit learning demands time. The methodology used in the classical implicit learning paradigms requires participants to actively observe and respond to each particular stimulus. For instance, in Reber's (1967, 1976) grammar learning experiments, the participants had to transcribe the 28 training sentences an average of five or six times each before making grammatical judgments about transfer sentences. In Lewicki, Hill & Bizot's (1988) sequence learning experiment, the participants responded to over 700 logical sequences of five target locations before the switch to random sequences that provided evidence for implicit learning. In experiments where the time to learn is more limited, the task difficulty is greatly reduced. For example, in Lewicki's (1986) Experiment 1, participants implicitly learned from a set of pictures concurrently presented with short descriptions that shorthaired women had one trait (e.g. kindness) and longhaired women had another trait. It this experiment, participants successfully learned the association by seeing each pictures only once for 16 to 17 seconds. However, there were only six training phase pictures, the decision required of the participants involved only two traits and the physical attribute (hair length) was perfectly predictive of a given trait. These studies suggest that implicit learning takes time especially when the relationship between the material and the response pattern is complex. Also, these studies required that participants focus their attention on the stimuli whether they be it Reber's sentences, Lewicki et al's logical sequences or Lewicki's pictures.

Experiments that directly assess the relationship between attention and implicit learning generally support the same conclusion. For example, studies using a divided attention paradigm (Reed & Johnson, 1994; Stadler, 1995) found that learning still took place, but that it was reduced. Other studies using partial report paradigms and sequence learning paradigms (Carlson & Dulany, 1985; Willingham, Nissen & Bullemer, 1989) have been unable to show any implicit learning of unattended attributes. However, in a sequence learning experiment in which geometric figures were used, Mayr (1996) succeeded in making participants sensitive to location sequences although the instructions only required naming the stimuli. Note, however, that this naming task still required the participants to attend to the location of the figures. Considering these results, we tend to agree with the conclusion that Goschke (1997) draws from his review of literature on the implicit learning of unattended features namely that: "implicit learning of unattended contingencies may be more probable when subjects have to respond to the critical stimulus feature (p. 278)" This is simply not the case in Brooks and his colleagues paradigm. Moreover, recent work has suggested that these same principles of learning should be applied to both implicit and explicit learning (see Wright and Whittlesea, 1998). Finally, note that implicit learning may occur without necessarily affecting decision processes. One can imagine some form of low-level perceptual sensitivity to previously encountered exemplars that would not interact with high-level rule driven category decisions.

Empirical evidence in support of the exemplar-based transfer effects

In their first experiment, Allen and Brooks (1991) obtained a negative match effect that they interpreted as supporting a multiple-system view of categorization. However, as just discussed, this finding is at odds with theories of implicit learning and with models of categorization. In such troubled theoretical waters, replications are critical to determine whether the phenomenon is robust enough to stay afloat. Because the negative match effect has been taken for granted, one must turn to the work of Brooks and his colleagues to find such replication attempts. They did two studies, which are sufficiently comparable to the original experiment to allow for an evaluation of robustness. First, Allen and Brooks (1991) replicated the rule paradigm of Experiment 1 in their Experiment 3. Three experimental conditions were created by using training phase instructions that either emphasized speed (perfectly replicating Experiment 1), accuracy (participant were told that responding correctly was the main concern), or alertness (participant were told about negative match item and they were told to respond correctly). The results obtained are shown in Table 7. As can be seen, none of the conditions succeeded in fully replicating the results of Experiment 1. An effect for error rates was found in all conditions. However, they were smaller than in Experiment 1.

Furthermore, no differences in response times were found between negative old and negative match items.

The other replication is found in Regehr and Brooks (1993, Experiment 2). Although slightly different stimuli were used (see Figure 6), they were constructed using the same abstract categorical description and with only one possible physical implementation for all logical values of the attributes. As is shown in Table 5, success at replicating the results was again limited. First, error rates did not show any evidence of a negative match effect. Also, they were excessively small compared to the error rates in Allen and Brooks' Experiment 1 (8% vs. 45%). Although Regehr & Brooks (1993) acknowledge that this was an exemplar effect of a "weaker form (p. 104)", they did not clarify why the effect for error rates in the Allen & Brooks study had disappeared.

Table 7

The results of Experiment 3 Allen and Brooks (1991, p. 11) for error rates and response times.

	Positive items			Negative items		
Condition	Old	Match	Difference	Old	Match	Difference
			Percent errors			
Speed	13	13	0	6	28	22
Accuracy	5	19	14	8	19	11
Alert	9	8	-1	4	18	14
			Response times (ms)			
Speed	1404	1609	205	1345	1348	3
Accuracy	1469	1478	9	1352	1519	167
Alert	2012	2157	145	2003	1936	-67

Regehr and Brooks did find an effect for response times. However, this was accomplished using a less conservative type of analysis than that used by Allen and Brooks. Regehr and Brooks compared the results obtained with negative match items to those obtained with all old items, both positive and negative. Hence, finding a difference between negative match and old items might be due to the fact that positive old items yielded faster average response times (or fewer errors) which, in turn, diminished the average response times (or error rates) for old items, thus producing the difference. Finding such a difference does not provide clear evidence in support of Brooks and his colleagues' hypothesis. By contrast, Allen and Brooks analyzed the differences between negative match and negative old items separately from the differences between positive old and positive match items. Although this procedure does not compare positive and negative items in the same analysis, it at least provides a stronger evaluation of the negative match effect because it directly assesses the difference between negative match and negative old items. Although mainly statistical in nature, this argument warrants caution in drawing inferences from the results reported by Regehr and Brooks.

In short, we do not find Brooks and his colleagues' replications to be very convincing. First, the very high 45% error rate for negative match items in Allen and Brooks' Experiment 1 has never been replicated. Also, Brooks and his colleagues have never been able to show a negative match effect for both error rates and response times in any of the replications. These replication difficulties, in conjunction with the theoretical problems raised by Brooks and his colleagues' interpretation of the negative match effect set the stage for our re-investigation of the rule paradigm.

5. Goals

The goal of this thesis is to further explore the potential relationship between rule-based and similarity-based mechanisms of categorization. In agreement with Ashby et al. (1998) and Allen & Brooks (1991; and Regehr & Brooks, 1993), we propose that two systems operate independently during category learning and both compete to provide categorical decisions. We will follow Smith and Shoben (1994) in defining the role of the rule-based categorization system as "[deciding] whether an [...] object belongs in a category by selecting out certain special features and determining whether the object satisfies a rule suggested by these features (p. 377)". Moreover, we propose that similarity-based system is exemplar-based and operates in accordance with the principles of the Context Model (Medin & Schaffer, 1978; Nosofsky, 1986). Our thesis is that rule-driven categorization precedes and heavily constrains exemplar learning. In opposition to Brooks and his colleagues, we will show that exemplar learning is limited to those attributes singled-out by the rule and that consequently, their influence on subsequent categorizations is also limited to these rule-based attributes.

The dissertation is mainly empirical. Specifically, we will investigate the acquisition of the attributes that compose exemplars using Brooks and his colleagues' categorization paradigm which states the classification rule at the beginning of the experiment. Using Experiment 1 will show that non-diagnostic attributes do not influence the categorization of transfer stimuli. In Experiments 2A and 2B, we will submit the data obtained in the rule paradigm data to an item analysis. Specifically, Experiment 2A replicates Experiment 1 and shows that the response times are determined by the time to apply the rule. By contrast, Experiment 2b shows a genuine influence of exemplar memory on categorization, using both subject and item analyses in an induction paradigm in which the rule is not explicitly given. Experiment 3 shows that the negative match effects obtained in Allen and Brooks' original Experiment 1 were due to the use of salient backgrounds to which attention was given. Using Regehr and Brooks's (1993) stimuli, Experiment 4 shows that exemplar effects are not due to holistic individuation, but rather to the distinctiveness of the attributes specified in the rule. Finally, Experiment 5 uses a one-dimension rule paradigm to show that memory for attributes, which are not focused upon by the application of the rule, is very limited even when these attributes perfectly predict category membership. Taken together, these results show that similarity-based effects on categorization are a direct consequence of the attention given to the exemplars, and the attributes from which they are built, through the application of a given rule and the observance of the experimental instructions.

Experiment 1

In Experiment 1, we will re-evaluate the role of exemplar-based learning on rule-based categorization using Allen & Brooks' (1991) rule paradigm. To conduct a strong test of Brooks and his colleagues' hypothesis that incidentally learned non-diagnostic attributes influence the application of a practiced, perfectly predictive rule, we followed Regehr and Brooks' (1993) variation of the paradigm. First, backgrounds were eliminated. This variation ensures that any negative match effect be due to an exemplar effect based on the two non-diagnostic attributes composing the creatures. Also, transfer phase items were presented in random order to avoid any order effect. Moreover, great care was taken to avoid possible confounding factors stemming from the verbal specification of the rule attributes or from their physical appearance. All exemplars served as both training and transfer items; all possible values of the three diagnostic attributes were used in the rules assigned to the participants; and all non-biased presentations order of the three rules attribute were presented. This methodological caution goes beyond all efforts at counter-balancing originally taken by Brooks and his colleagues.

Furthermore, conditions and tests were added to adequately evaluate Brooks and his colleagues' hypothesis. First, participants either received five training blocks (as was the case in the standard version of the paradigm) or 20 blocks. This last condition was added to give the participants an increased opportunity to learn the exemplars and thus, increase the probability that a negative match effect be found. Another innovation was to add a *naming condition*. In our variation of the rule paradigm, participants categorized the stimuli using "family names" for the creatures. In the naming condition, they were also told that each stimulus had a "first name" and their secondary task was to learn these names. These instructions were intended to encourage the participants to memorize the individual creatures, which according to Brooks (1978) should increase the chances of finding exemplar-based effects. Finally, an explicit recognition test was performed after the categorization phase of the experiment to determine if participants could remember the training phase items.

In opposition to Brooks and his colleagues, we will show that non-diagnostic attributes' influence on rule-based categorization is minimal, even when increased practice is given along with directives that helped to individualize the stimuli.

Method

Participants.

Sixty-four students at the Université de Montréal participated in the study. All received 3$ as compensation for their time. This experiment, and all the experiments reported throughout this dissertation, was conducted in French.

Materials.

The stimuli were drawings of fictional animals (similar to those used in Allen and Brooks, 1991) built from five binary attributes: tail type (cane-shaped or stair-shaped), back pattern (striped or spotted), head shape (parabolic or oval), body type (oval or parallelogram), and color (yellow or green). The logical description of the categorical structure is given in the Table 8.

Table 8

Logical description of the stimuli used in Experiment 1.

	Old items (training phase)						Match items (transfer phase)					
	Item number	Tail type	Back pattern	Head shape	Body type	Color	Item number	Tail type	Back pattern	Head shape	Body type	Color
Positive												
	1	1	1	1	0	0	10	1	0	1	0	0
	3	1	0	1	1	1	9	1	1	1	1	1
	6	0	1	0	1	1	15	0	0	0	1	1
	8	0	0	0	0	0	13	0	1	0	0	0
Negative												
	2	0	1	1	0	1	14	0	0	1	0	1
	4	1	1	0	1	0	16	1	0	0	1	0
	5	1	0	0	0	1	11	1	1	0	0	1
	7	0	0	1	1	0	12	0	1	1	1	0

New items (memory phase)											
Item number	Tail type	Back pattern	Head shape	Body type	Color	Item number	Tail type	Back pattern	Head shape	Body type	Color
17	1	1	1	0	1	21	1	0	1	0	1
18	1	0	1	1	0	22	1	1	1	1	0
19	0	1	0	0	1	23	0	0	0	0	1
20	0	0	0	1	0	24	0	1	0	1	0
25	0	1	1	0	0	29	0	0	1	1	1
26	1	1	0	1	1	30	1	0	0	0	0
27	1	0	0	1	1	31	1	1	0	0	0
28	0	0	1	0	0	32	0	1	1	1	1

The first three attributes mentioned in the rule were diagnostic and specified category membership. The last two attributes, body type and color, were non-diagnostic because both of their values appeared equally often in each category. Their purpose was to maximize similarity between old and match items. In opposition to Brooks and his colleagues, attributes that span the entire stimulus were chosen to be non-diagnostic. This was done to ensure that the participants had every chance of noticing these attributes no matter where on the stimulus their attention was focused when applying the rule.

The basic idea of the rule paradigm is to give participants a clear, perfectly predictive rule at the onset of the experiment. The rule was disjunctive, requiring that at least two out of three attributes values be present in a stimulus for it to belong to a category. With three binary attributes, there are eight possible rules. All were used in Experiment 1 (see Table 9). For instance, rule 1 stated that if a creature had two or three of the following attributes: a cane-shaped tail, stripes, and a parabolic-shaped head, then it belonged in the "Tremblay" category. Otherwise, it belonged in the "Beaulieu" category. These different rules change the category membership of the stimuli and whether they are positive or negative items, but they do not determine items serving as training vs. transfer items. To balance this factor, the exemplar sets 1 to 8 and 9 to 16 served in turn as old items and match items.

The purpose of using all these rules was to balance the assignment of physical attributes to the abstract values given in the categorical structure, so that all items serve in all categories, and in all conditions of the experiment. This prevents any effect found from being due to the unforeseen saliency of a particular attribute value or of a particular combination of attribute values.

Table 9.

The rules given to participants in Experiment 1.

	Rules
1-	A cane-shaped tail, stripes, and a parabolic-shaped head;
2-	A cane-shaped tail, stripes, and a oval-shaped head;
3-	A cane-shaped tail, spots, and a parabolic-shaped head;
4-	A stair-shaped tail, stripes, and a parabolic-shaped head;
5-	A cane-shaped tail, spots, and a oval-shaped head;
6-	A stair-shaped tail, stripes, and a oval-shaped head;
7-	A stair-shaped tail, spots, and a parabolic-shaped head;
8-	A stair-shaped tail, spots, and an oval-shaped head.

The rule was given to participants in one of two orders: tail, back pattern and head; or head, back pattern and tail. Examination of the categorical structure shows that these two orders avoid a possible bias in the application of the rule. Allen and Brooks (1991) and Regehr & Brooks (1993) partially balanced the attribute values used to satisfy the categorization rule, but the verbal specification of the rule always presented the diagnostic attributes in the same order: leg length, body type and spots. As can be seen in Table 3, if the participants tested the attributes in the order specified in the rule, then they were forced to consider all three attributes before reaching a conclusion about negative items, whereas they only needed two tests to verify positive items. This possible bias could have contributed to the poorer performance obtained with negative old and negative match items. By contrast, the two rule presentation orders used in the present experiment were balanced with respect to the expected number of attribute tests for positive and negative items.

For participants in the naming condition, the stimuli were also given "first names". All were chosen to be short, common names. They were: Luc, Carl, Jean, Louis, Guy, Marc, Alex, David. The names were partially counter-balanced with regard to their association with particular stimuli. For half of the participants, these first four names were associated with items 1 to 4 respectively, and the last four with items 5 to 8. For the others, these first four names were associated with items 5 to 8 respectively, and the last four with items 1 to 4. Although complete counter-balancing would have been ideal, the number of participants necessary to implement such a control was unrealistically large.

For the training and transfer phases, we chose the 16 items with the same logical structure as Allen & Brooks (1991). Each transfer item was matched with a training item on every attribute except back pattern. Table 8 shows match items next to their corresponding training items. The sixteen remaining items that may be constructed from combinations of five binary attributes were used in the recognition test. These items also matched with a training item on the diagnostic attributes, but they were different because of the new pairings of non-diagnostic attributes used. In all, five types of stimuli were produced:

1. *Positive olds:* Item first seen in the training phase and for which the corresponding transfer item is in the same category.

2. *Negative olds*: Item first seen in the training phase and for which the corresponding transfer item is in the opposite category.

3. *Positive match*: Item first seen in the transfer phase that is in the same category as the training item to which it is most similar.

4. *Negative match*: Item first seen in the transfer phase that is in the opposite category to the training item to which it is most similar.

5. *New items:* Items seen only in the recognition test which are neither old nor match items.

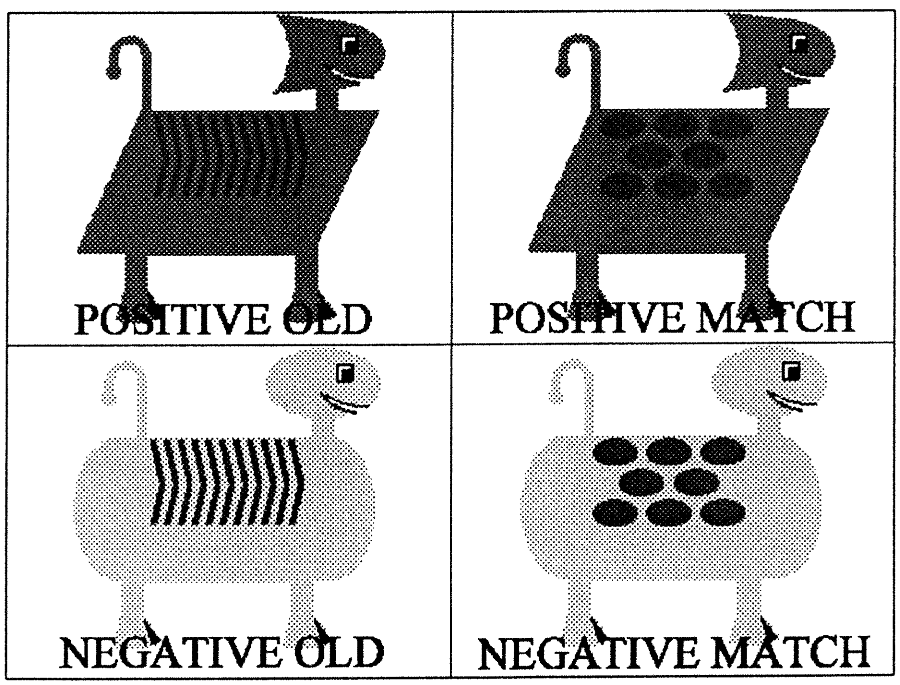


Figure 8. Black and white examples of the stimuli used in Experiment 1. The labels apply for participants given rule 1 (see Table 9).

Examples of the first four types of exemplar are given in Figure 8. All items appeared on a black background.

Procedure.

The participants were tested individually. All instructions and stimuli were presented on 14" VGA monitors connected to 386dx or 486dx IBM compatible computers. The program MEL Professional v.2.01 (Schneider, 1989) was used to give the experimental instructions, present the material and record the participants' answers and response times.

Participants were assigned to one of two conditions: the *standard condition* or the *naming condition*. Both conditions included five phases. In the first phase, the participants were given the categorization rule and instructed to classify the animals accordingly. They were instructed to answer as quickly as possible while being as accurate as possible (as was the case in Allen and Brooks, 1991; and Regehr and Brooks, 1993). Participants were given 40 trials divided into five blocks. The same procedure was used in Regehr and Brooks (1993) and an equal number of classification decisions were requested in Allen and Brooks (1991). Each block involved the presentation of the eight old items in random order. All trials proceeded as follows. First, a fixation point appeared in the center of the screen for 1500 ms. An old item was then presented and participants had to classify the stimulus by pressing the appropriate key ("A" for

Beaulieu or “L” for Tremblay) on the keyboard. The stimulus remained on display for another 2000 ms and feedback pertaining to accuracy was added. For correct answers, this feedback was the category name. For wrong answers, a short buzzing tone accompanied the correct category name. Once the 2000 ms had elapsed, the screen was cleared. The inter-stimulus interval was 1000 ms.

Participants in the *standard condition* received only the training phase instructions previously described. However, participants in the *naming condition* received further instructions. Like participants in the standard condition, they were told that their primary task was to classify the creatures according to the rule. In addition, they were told that the feedback for any given stimulus would include both the category name (the family name) and the first name. They were told to take the two second period to try and learn the first name of each individual creature. The goal of these instructions was to favor the individuation of the creature and thus, the memorization of the exemplar as whole entities. Otherwise, the naming condition was identical to the standard condition.

In the transfer phase, the four positive old, the four negative old, the four positive match and the four negative match items were presented once each in random order. This was the procedure used in Regehr and Brooks (1993). As Allen and Brooks (1991) noted, this procedure might increase the chance that participants realize the presence of conflicting transfer items. We chose Regehr and Brooks’ method, preferring to avoid potential order effects. This phase of the experiment proceeded like the training phase except that stimuli disappeared as soon as an answer was given and that no feedback was given. The procedure used to this point duplicates the basic rule paradigm devised by Brooks and his colleagues.

In the third phase of the experiment, the participants were given an additional 15 blocks (120 trials) of practice. Those in the standard condition continued to classify the creatures according to the rule and those in naming condition still tried to memorize the creatures' first names. This supplementary practice was given in order to give the participants more time to memorize the exemplars and thus, increase the possibility that negative match effects be found. Note that, over both training phases, participants saw and categorized the creatures on a total of 160 trials. This is more than in Allen and Brooks' (1991) Experiment 1, in which participants saw the creatures on 120 trials, but classified them on only 40 occasions.

The fourth phase of the experiment was another test phase in which positive old, negative old, positive match and negative match items were presented once in random order as was done in the first test phase. This provided an opportunity to test for negative match effect after 20 practice blocks.

The last phase was an explicit recognition memory test. Participants were shown the eight old items, the eight match items and the sixteen new items in a random order.

Their task was to determine whether each particular item had been seen in the training phases of the experiment (identified as "*those parts of the experiment in which you were getting feedback*") or whether the item was not shown in these phases. Hence, this direct memory test forced participants to explicitly discriminate old items from match and new items. Responses were given by pressing the "1" key for old items and the space bar for the other types of items. No feedback was given concerning response accuracy.

Then, the participants in the naming condition were given a paper and pencil test to evaluate how well they had learned the creatures' first names. Each of the eight training stimuli was printed in color in one of eight numbered rectangles on a sheet of white paper (see Appendix A). A second sheet, identical to the first except that the rectangles were empty, was also given to the participants (see Appendix B). They were told to write the first and last name of the creatures in the rectangles, which corresponded to the stimuli printed on the first sheet. They could use each first name only once.

Finally, it must be added that between the second match phase and the memory test, stimuli built using only the three rule attributes were presented once. Remember that without non-diagnostic attributes, it is impossible to distinguish old from match items. Hence, there remained only four positive and four negative items as defined by the rule attributes. The goal of showing these stimuli was to obtain a baseline measure of the application of the rule to positive and negative items. However, as these data were not very informative, they will not be presented.

Results

Classification task

Separate analyses of variance (ANOVAs) were performed on the error rates and on the response times to compare performance on old and match items presented in the test phases. These analyses involved a (2) X 2 X 2 design with two within-subjects factors: phase (old vs. match) and value (positive vs. negative), and one between-subjects factor: training condition (standard vs. naming). These analyses are similar to those of Allen and Brooks (1991) in that they assess the variance of positive old and negative old items separately. However, this design compares the data for both old and match items, and positive and negative items in the same analysis. We believe this is the type of analysis most fit to evaluate Brooks and his colleagues' hypothesis, because finding a negative match effect involves finding a difference between negative match and negative old items that is greater than the difference between positive match and positive old items. Classification data obtained after five blocks and twenty blocks of practice were analyzed separately. Error trials were eliminated for response time analyses.

Five blocks.

The error data after five training blocks pertaining to phase and value are presented in Figure 9. As can be seen, error rates were all below 10%. The interaction

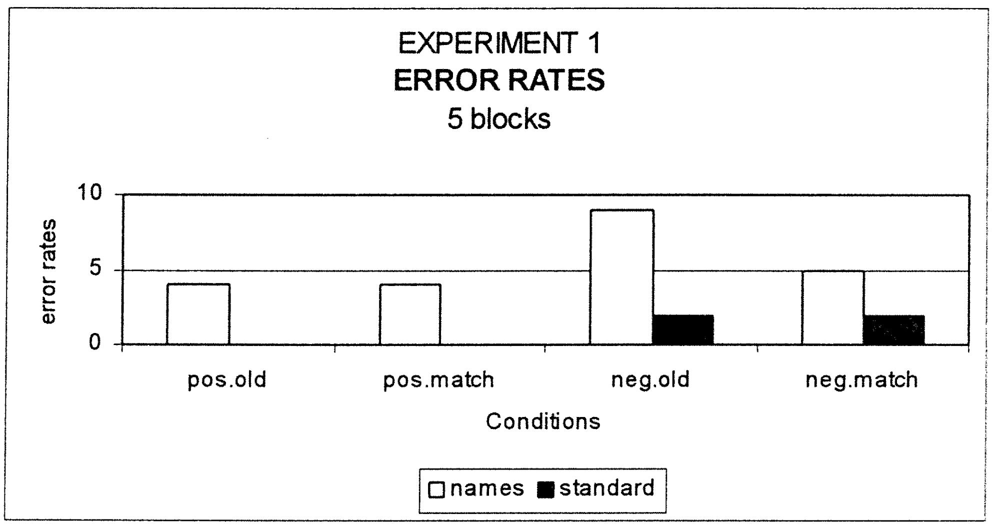


Figure 9. Error rates after five blocks of training for both the names and standard conditions.

between training condition, experimental phase and item value was not significant $\underline{F}$ (1, 62) = 0. 99, $\underline{MS_e}$ = .009, $\underline{p} < 0.323$, nor was the interaction between experimental phase and item value, $\underline{F}$ (1, 62) = 0. 44, $\underline{MS_e}$ = .004, $\underline{p} < 0.509$. These results are comparable to those found in Regehr and Brooks' Experiment 2A in which small error rates were also found (5% on average for positive and negative old items vs. 4% in this experiment, and 8% for negative match items vs. 4% in this experiment). However, the results differ drastically from those reported in Allen and Brooks' Experiment 1 in which the error rates were of approximately 20% for negative old items and 45% for negative match items. Finally, a main effect was found for training condition, $\underline{F}$ (1, 62) = 6.27, $\underline{MS_e}$ =

0.14, $p < 0.015$. Participants in the naming condition made more mistakes than the participants in the standard condition (6% vs. 1%).

The response time data obtained after five training blocks are presented in Figure 10. First, as was the case for error rates, a main effect was found for training condition, $F(1, 62) = 4.97$, $MS_e = 3818360$, $p < 0.03$. Participants in the naming condition showed greater response time latencies than the participants in the standard condition (1395 ms vs. 1151 ms). Considering that the participant in the naming condition also made more classification errors, these response times were clearly not due to a speed-accuracy trade-off. Learning the first names while also categorizing the

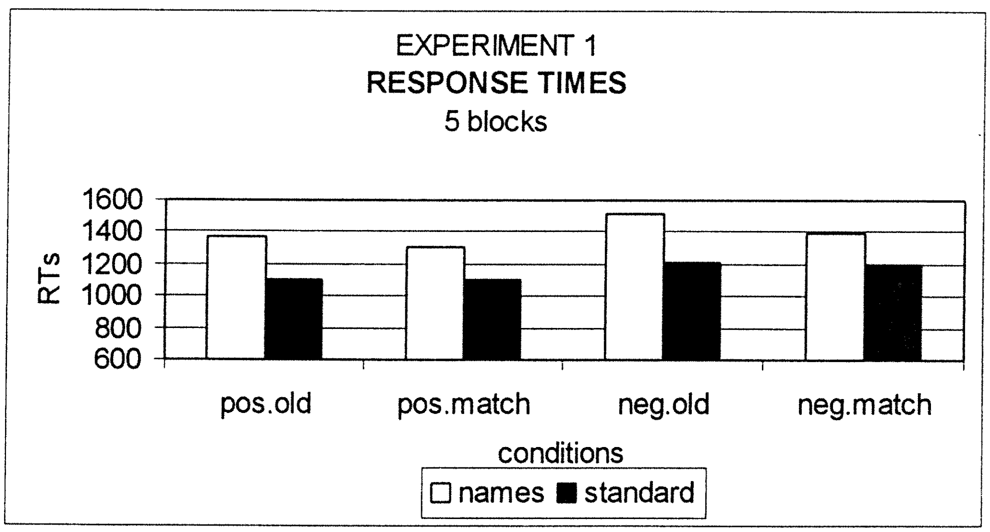


Figure 10. Response times after five blocks of training for both the names and standard conditions.

stimuli apparently made the task more difficult.

The interaction between training condition, experimental phase and item value was not significant $\underline{F}$ (1, 62) = 0. 139, $\underline{MS_e}$ = 10534, $\underline{p} < 0.71$, nor was the interaction between experimental phase and item value, $\underline{F}$ (1, 62) = 0.39, $\underline{MS_e}$ = 29473, $\underline{p} < 0.54$. Response times for negative old items were 1362 ms vs. 1294 ms for negative match items (an improvement of 68 ms). Hence, the results did not replicate Allen and Brooks' large difference between negative old and negative match items (approximately 1250 ms vs. 1600 ms).

Twenty blocks.

The error data after twenty training blocks are presented in Figure 11. Once

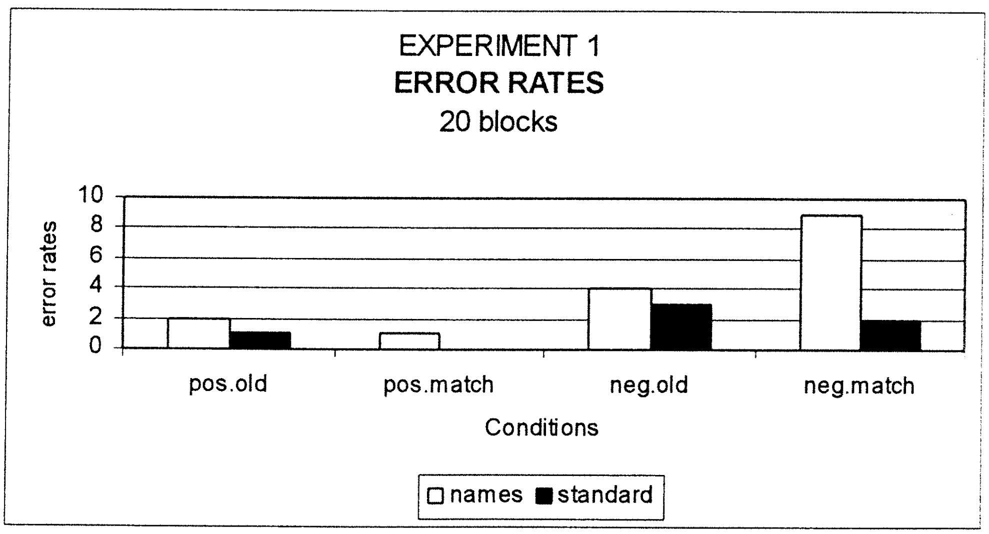


Figure 11. Error rates after 20 blocks of training for both the names and standard conditions.

more, error rates were below 10%. The interaction between training condition, experimental phase and item value was significant, $\underline{F}$ (1, 62) = 4.62, $\underline{MS_e}$ = 0.02, $\underline{p}$ <

0.035. A decomposition of the interaction revealed that the phase by value interaction was significant in the naming condition, $\underline{F}$ (2, 124) = 3.96, $\underline{MS_e}$ = 0.02, $\underline{p} < 0.02$, but that it was not in the standard condition, $\underline{F}$ (2, 124) = 0.49, $\underline{MS_e}$ = 0.00, $\underline{p} < 0.611$. Further

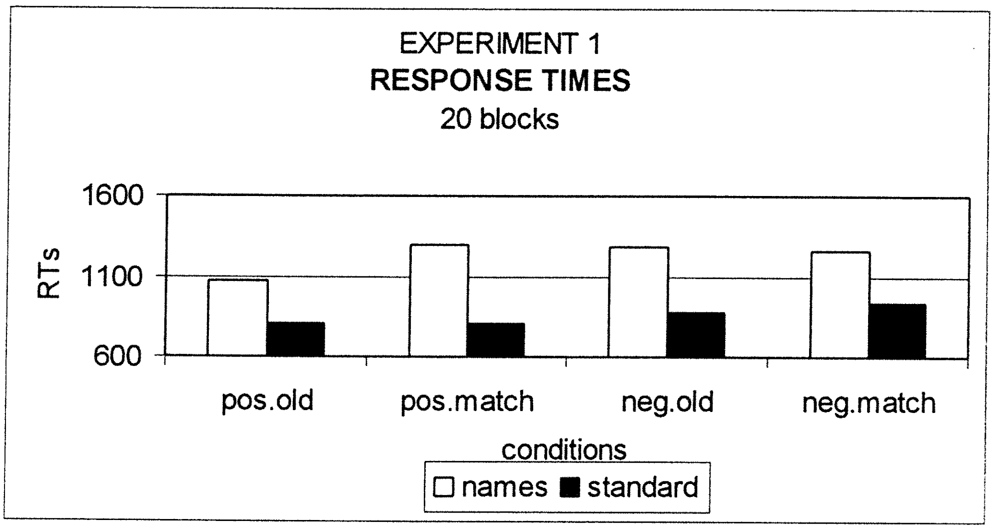


Figure 12. Response times after 20 blocks of training for both the names and standard conditions.

decomposition of the significant effect for the naming condition showed a significant difference between negative old (4% error rate) and negative match items (9% error rate), $\underline{F}$ (1, 62) = 4,73, $\underline{MS_e}$ = 0.04, $\underline{p} < 0.03$. However, there was no difference between positive match (1%) and positive old items (2%). Therefore, with prolonged training and instructions to individualize the stimuli, there was a small, but significant increase in error rates for negative match items. Finally, a main effect was found for training condition, $\underline{F}$ (1, 62) = 4.00, $\underline{MS_e}$ = 0.04, $\underline{p} < 0.05$ with participants in the naming condition still showing less accuracy than the participants in the standard condition (4% vs. 1%).

The response time data obtained after twenty training blocks are presented in Figure 12. The interaction between training condition, experimental phase and item value was significant $\underline{F}$ (1, 62) = 7.415, $\underline{MS_e}$ = 3961158, $\underline{p}$ < 0.01. The decomposition of the interaction showed a significant phase by value effect for the naming condition, $\underline{F}$ (2, 124) = 8.35, $\underline{MS_e}$ = 2436157, $\underline{p}$ < 0.001. However, further decomposition showed that the difference between negative match (1290 ms) and negative old items (1303 ms) was not significant, $\underline{F}$ (1, 62) = 0.3, $\underline{MS_e}$ = 15763, $\underline{p}$ < 0.58. Also, the phase by value interaction was not significant in the standard condition, $\underline{F}$ (2, 124) = 0.5, $\underline{MS_e}$ = 226140, $\underline{p}$ < 0.61. Response times were 940 ms for negative match items and 883 ms for negative old items. Finally, participants' average response times in the naming conditions were much slower (1231 ms) than those of participants in the standard condition (859 ms), $\underline{F}$ (1, 62) = 13.7, $\underline{MS_e}$ = 8875216, $\underline{p}$ < 0.001. Nonetheless, there was no evidence to suggest a negative match effect in response times.

Interestingly, a main effect was found for phase, $\underline{F}$ (1, 62) = 5.38, $\underline{MS_e}$ = 274428, $\underline{p}$ < .024. Response times for old items (1011ms, SD = 434ms) were significantly faster than those for match items (1078ms, SD = 236ms). This effect was not present after five blocks, $\underline{F}$ (1, 62) = 0.83, $\underline{MS_e}$ = 137239, $\underline{p}$ < .37. This suggests that after twenty blocks of training, participants had some memory for training exemplars, which included information about the non-diagnostic attributes. Otherwise, training and transfer items are undistinguishable.

Memory test

Table10 presents the percentage of "old" vs. "other" responses given to old, match and new items. One participant's data in the naming condition was not included in the analysis because the data file was lost. The global performance was close to random (44% correct) and participants in the naming condition (48% correct) were slightly superior to their counterparts in the standard condition (41% correct).

Table 10

Experiment 1 memory test results after 20 blocks of training for both the standard and naming conditions.

		Responses	
		Old	Other
Conditions			
Standard			
	Item type		
	Old	72%	28%
	Match	63%	37%
	New	74%	26%
Naming			
	Item type		
	Old	79%	21%
	Match	55%	45%
	New	66%	34%

Signal detection analyses (Coombs, Dawes, and Tversky, 1970) were conducted to evaluate the participants' ability to discriminate "old" exemplars from "others". The results showed that participants could not explicitly distinguish the stimuli in either the standard condition (d'= 0.25, B = 0.96) or the naming condition (d'= 0.01, B = 1.06).

Naming test

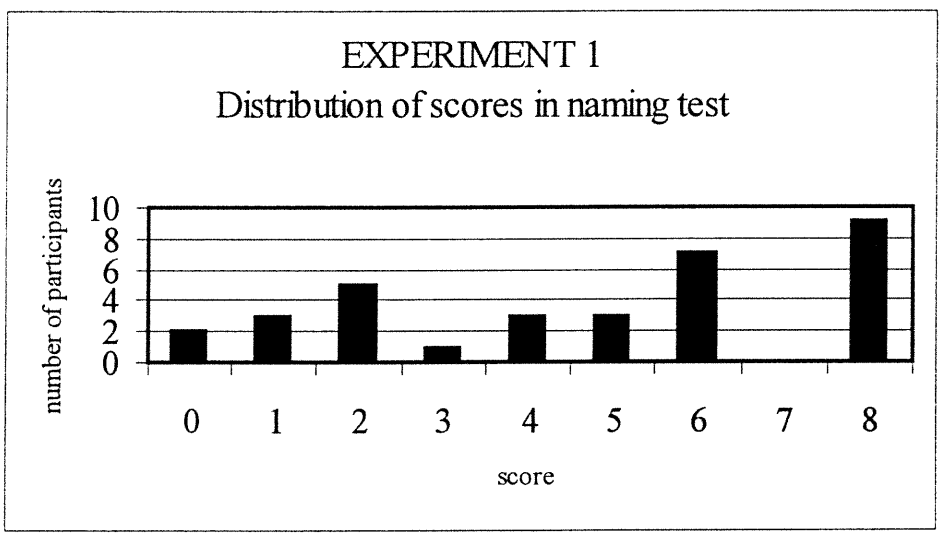


Figure 13. Distribution of the success scores in the naming test.

Participants in the naming condition received a paper and pencil test in which they were asked to identify the eight training phase items by their first names. Success on this task was rated on a score of 0 (no good answers) to 8 (perfect score).

The average was 4.75 (59%) with a standard deviation of 2.76. The distribution of these scores is given in Figure 13. As can be seen, nine participants (28%) perfectly identified the eight stimuli, seven (22%) identified six out of eight stimuli correctly, and 16 participants (50%) failed to correctly identify more than five stimuli. Thus, these scores indicate great variability in the performances between subjects. Therefore, it is possible that the participants who achieved high success rates in the naming task actually produced the negative match effects observed for error rates. The response times of the same participants may also exhibit a negative match effect when the results are not averaged with those of the participants who did poorly in the naming task.

To verify these possibilities, a (2) x 2 x 2 ANOVA was conducted separately on the error rates and the response times in the classification task after twenty blocks of practice under the naming condition. There were two within-subject factors: phase (old vs. match) and value (positive vs. negative), and one between-subjects factor: success in the naming test as determined by a median split. Participants who obtained a score of five or less were grouped together (and named *the low success group*) and compared with participants who obtained a score of six or more (*the high success group*).

For error rates, the interaction between phase, value, and success in the naming test was not significant, $\underline{F}\ (1, 30) = 0.38$, $\underline{MS_e} = 0.002$, $\underline{p} < .542$. The differences between negative match and negative old items were 4% in the high success group and 6% in the low success group. The same interaction also failed to be significant for response times, $\underline{F}$ (1, 30) = 0.24, $\underline{MS_e}$ = 21207, $\underline{p}$ < .63. The differences between negative match and negative old items were 18ms in the high success group and –80ms in the low success group. Although the high success group's response times were, on average, 163ms slower than those shown by the low success group, this difference was not significant, $\underline{F}$ (1, 30) = 0.73, $\underline{MS_e}$ = 844499, $\underline{p}$ < .40. Hence, the data shows that even when high success participants are teased apart from low success ones, there is no evidence that the capacity to successfully identify the stimuli prompts an increased influence of memorized exemplars on the application of rule.

In the previous section, it was noted that old items were categorized faster than match items after twenty blocks of training. Thus, it is possible that high success in naming the stimuli allows for even faster classification times for old items presumably because these items are recognized more easily than the other items. However, there was no evidence to support this claim, $\underline{F}$ (1, 30) = 0.2, $\underline{MS_e}$ = 17898, $\underline{p}$ < .66.

There was also the possibility that participants in the high success group were superior to those in the low success group with regards to the memory test. To evaluate this, the memory test data was analyzed separating low and high success participants. A 1 x 3 ANOVA was conducted on the mean number of correct responses to assess the performances in the explicit memory test. The between-subject factor was condition type. For correct responses, a significant difference was found, $\underline{F}$ (1, 29) = 9.8, $\underline{MS_e}$ = 0.01, $\underline{p}$ < .004. The high success group produced more good answers (54%, SD = 13%)

than those in the low success group (41%, SD = 11%). Participants in the high success group correctly identified 9% more old items, 22% more match items and 12% more new items. It may also be observed that data for the participants in the low success group is very similar to that observed for participants in the standard condition (see Table 10). Hence, the high success group's accuracy was 25% greater than that of the low success group. Although general performance is mediocre, these data show that recognition was superior in the naming condition.

Secondary analyses of rule types and items sets.

As discussed in the method, all possible rules were used and the item sets used as old and match items were counter-balanced to avoid obtaining effects due to the

Table 11

Experiment 1memory test results after 20 blocks of training for participants in the low and high success group of the naming condition.

		Responses	
		Old	Other
Naming condition			
Low success			
	Item type		
	Old	74%	26%
	Match	67%	33%
	New	72%	28%
High success			
	Item type		
	Old	83%	17%
	Match	45%	55%
	New	60%	40%

uncontrolled saliency of a particular attribute or attribute combinations. Nonetheless, 8 x 2 x (2 X 2) ANOVAs were conducted at five blocks and twenty blocks on error rates and response times to determine if these variables influenced the categorization process. There were two between-subject factors: the rules (1 to 8), and the item sets (items 1 to 8 vs. items 9 to 16) which were alternately used during training and transfer; and two within-subject factors: phase (old vs. match), and value (positive vs. negative).

For error rates, rules and item sets did not interact singly or jointly with phase and value at five blocks (all $\underline{F}$s < 1.6). At twenty blocks, there was a tendency for item sets to interact with phase and value, $\underline{F}\ (1, 48) = 3.1$, $\underline{MS}_e = 0.004$, $\underline{p} < .08$. However, the interaction did not stem from differences between negative match and negative old items. When items 1 to 8 served as old items, error rates were 5% for negative match and 3% for negative old items and they were 5% for negative match and 4% for negative old items when items 9 to 16 served as old items. The other interactions involving these factors were not significant (both $\underline{F}$s < 1.8).

For response times after five blocks, rules significantly interacted with phase and value, $\underline{F}\ (1, 48) = 3.5$, $\underline{MS}_e = 151970$, $\underline{p} < .004$. The interaction was decomposed to verify if there were significant effects of phase and value within the individual rules. Rule 6 and 7 (see Table 9) yielded such effects ($\underline{F}$s > 4.7). However, in both cases, negative old item response times were slower than negative match response times (1817 ms for negative old vs.1314 ms for negative match items with rule 6 and 1276 ms for negative old vs.1186 ms for negative match items with rule 7). The other rules did not yield this effect (all $\underline{F}$s < 0.45). Other interactions involving rules and item sets were not significant (all $\underline{F}$s < 2.6). At twenty blocks, item set interacted with phase and value, $\underline{F}$ (1, 48) = 3.93, $\underline{MS}_e$ = 146617, $\underline{p}$ < .0053. The interaction was decomposed by item set conditions. There was a significant interaction between phase and value when item 1 to 8 served as old items, $\underline{F}$ (2, 96) = 3.3, $\underline{MS}_e$ = 157452, $\underline{p}$ < .04, but not when items 9 to 16 served as old items, $\underline{F}$ (2, 96) = 2.4, $\underline{MS}_e$ = 114066, $\underline{p}$ < .10. However, the difference between negative old (1046ms) and negative match (1026ms) was not responsible for the significant interaction when items 1 to 8 served as old items. Finally, the other interaction involving rules and item sets were not significant (both $\underline{F}$s < 1.5).

Hence, although there is evidence that some variation in the results might have stemmed from the different rules and item sets, it was not the case that these factors systematically varied the possibility of obtaining negative match effects.

Discussion

The goal of Experiment 1 was to re-evaluate the role of exemplar-based learning on rule-based categorization using Allen & Brooks' (1991) rule paradigm. It presents a strong test of Brooks and his colleagues' hypothesis that incidentally learned non-diagnostic attributes influence the application of a categorization rule. Possible confounds resulting form the order of presentation of the rule attributes were eliminated. No context was presented, leaving only the attributes comprising the exemplars to learn.

The fact that the non-diagnostic attributes spanned the entire stimulus ensured that they were within the participants' gaze. Furthermore, participants were given longer training to increase exemplar memory. Some were even asked to learn the exemplars' "first names" to individuate the stimuli. Yet, evidence was minimal to support the idea that exemplar memory was influencing conscious category decisions.

For error rates, the results obtained after five blocks of training replicate those of Regehr and Brooks' Experiment 2A (1993). Error rates were below 10% and there was no significant difference between conditions. However, they were very different from those reported in Allen and Brooks' Experiment 1 (1991) in which a 45% error rate for negative match items was found. Our results held under several conditions and tests. First, participants with twenty blocks of practice in the standard condition were not more likely to respond incorrectly when they were shown negative match items than participants with five blocks of practice. In principle, this increased opportunity to memorize the exemplars should have made the effect more present.

Participants in the naming condition did show a negative match effect, but the 9% error rate obtained for negative match items is far from the 45% obtained in Allen and Brooks' (1991) Experiment 1. The modest effect obtained in this condition may have been generated by a tendency for the participants in the naming condition to pay more attention to non-diagnostic attributes than those in the standard condition. Contrary to the classification task in which participants were told to use certain attributes, learning the creatures' first names posed no such constraints. Participants could focus on the attributes of their choosing to identify the creatures. There are many solutions to successfully complete this task. For example, a participant could select the three diagnostic attributes; or tail type and back pattern and body type; or tail type and head shape and color; etc. Hence, in trying to learn the creatures' names, some participants may have paid attention to the non-diagnostic attributes, thus creating an exemplar memory of the old items that created the small negative match effect observed.

The difference found between negative match and negative old items in the naming condition and not in the standard condition lends support to this explanation. As was discussed in the Introduction, Brooks and his colleagues' explanation of the negative match effect in the standard condition is not plausible because there is no reason for the participants to pay attention to the non-diagnostic attributes. But in the naming condition, the experimental instruction gave the participants a motive to do so. The large negative match effects that will be found when attention is systematically given to non-diagnostic attributes (later in Experiment 3) will also validate this account. Finally, the fact that the high success group in the naming condition did not generate more negative match effects than the low success group simply reflects the first group's use of better name learning strategies without there being a difference in the attention given to non-diagnostic attributes. In all, the error rate data show that, contrary to Allen and Brooks' claim, exemplar memory may influence analytical classification only if some attention is given to the non-diagnostic attributes.

The response times obtained after five training blocks also failed to yield a negative match effect, by contrast to those obtained in Allen and Brooks (1991) Experiment 1 and in Regehr and Brooks (1993) Experiment 3. There is one possible explanation for the discrepancy with these two other studies. First, the presentation order of the rule attributes used in Allen and Brooks and Regehr and Brooks may have created a bias against negative items (see the method section). If the participants tested the attributes in the order given by that rule, then they were forced to consider all three attributes before reaching a conclusion about negative items, whereas they only needed two tests to verify positive items. Two other factors, which we will examine in Experiment 3, may also have contributed to Allen and Brooks early findings. One is the presence of the contexts and stories that were introduced to encourage memorization of the exemplars. The other is the very specific testing order used by Allen and Brooks.

After twenty blocks of training, there was still no evidence of a negative match effect on the response times obtained in the standard condition and in the naming condition. Hence, the increased practice did not succeed in making negative match items create the hesitation in responding anticipated by Brooks and his colleagues.

The most unexpected result of Experiment 1 is that classification times were systematically shorter for the old items seen in the test phases compared to the transfer items. This difference cannot be attributed to the logical structure of the rule attributes, which was identical for old and match items. It must be attributed to some memory of the exemplars. The fact that the difference was of similar magnitude for the standard and the naming conditions suggests that the effect is independent of deliberate learning strategies. Rather, it appears to result from the repeated exposure to the training stimuli. This, in turn, suggests that the resulting perceptual representation of the attribute combinations making each training exemplar facilitates their processing in the test phases. However, these representations do not seem to interfere with the use of the categorization rule.

Hence, Brooks and his colleagues' hypothesis that exemplar memory may influence conscious category decision after limited training and with little attention was not supported. Negative match effects were not found with the exception of error rates in the naming condition after twenty blocks of practice. However, this small effect may be explained by the increased attention given to the exemplars because of the naming instructions.

Experiment 2

Our primary goal in presenting Experiments 2A and 2B is to compare performance in the standard rule paradigm and in the induction or concept formation paradigm. These experiments were in fact carried out before Experiment 1. Less care was taken to counterbalance the values of the rule attributes: one rule was used in

Experiment 2A and two rules were used in Experiment 2B. However, the analyses performed on the results of Experiment 1 suggest that the nature of the attributes values specified in the rules is not a determining factor. Another difference with Experiment 1 is that the amount of training varied between subjects so that the subsequent recognition test allows measuring explicit exemplar memory at different stages of learning in both paradigms. Finally, the number of participants involved in Experiments 2A and 2B is much larger than in Experiment 1, thereby providing more power to the statistical tests. The increased number of participants in conjunction with the smaller number of rules used also yields less variable results for each individual item. The second major goal of Experiment 2 is precisely to go beyond averaged results and to consider performances on each individual item. By comparing each old item to its corresponding match item, it should be possible to detect negative match effects for some exemplars that are not necessarily revealed in analyses based on averaged data. However, we entertain the competing hypothesis that non-diagnostic attributes are not stored in a way to generate a conflict between exemplar memory and rule application. In this case, item analyses that compare training with transfer items solely on the basis of the rule attributes and that ignore the non-diagnostic attributes should show a clear relationship between response times and rule application.

Experiment 2A

Method

Participants.

Eighty-two students at the Université de Montréal participated in the study. Participants in the 10 block learning condition received 3$ as compensation for their time and those in the 20 block condition received 5$. Two participants were dropped. One participant was excluded because of past participation in an experiment involving similar material and another participant was excluded because the instructions relative to the classification task were misunderstood.

Materials.

The stimuli were very similar though not identical to those in Experiment 1. The five binary attributes were: body type (parallelogram or curved), spots (small and circular, or large and oval), tail type (circular or cane shaped), texture (dotted or striped), and color (green or yellow). The logical description of the categorical structure is identical to that given in Experiment 1 (see Table 8).

The standard rule paradigm was used with only one rule type. If an animal had two or three of the following attributes: a body in the shape of a parallelogram, small circular spots and a circular tail, then it was classified as a Maurice. Otherwise, it was classified as a Henri. The rule was given to participants in the exact order mentioned above to avoid the bias for negative items (see Experiment 1 method section).

Item 1 to 8 in categorical structure (see Table 8) served as old items and items 9 to 16 served as match items. As was the case in Experiment 1, each transfer item was matched with a training item on every attribute except spots. Only eight additional items were reserved for the recognition test (items 17, 19, 21, 23, 26, 28, 30 and 32 in Table 8). These manipulations produced the five types of stimuli found in Experiment 1: positive old items, negative old items, positive match items, negative match items, and new items (see Appendix C for example of the stimuli).

Procedure.

Participants were randomly assigned to one of two training conditions: 10 block or 20 blocks. The participants were tested individually. All instructions and stimuli were presented with 386dx or 486dx IBM compatible computers using MEL Professional v.2.01 (Schneider, 1989).

The experiment was conducted in three phases. In the training phase, the participants were first given the categorization rule and instructed to classify the animals accordingly. Then, participant in the 10 block condition were given 80 trials of training, whereas participants in the 20 block condition were given 160 trails. Ten blocks of training were chosen instead of 5 (as in Experiment 1) to provide a number of learning trials that falls between the 40 categorical decisions required in Allen and Brooks' (1991) Experiment 1 and the 120 presentations of the training stimuli. Each block involved the presentation of the eight old items in random order. All trials proceeded as follows. First, a fixation point appeared in the center of the screen for 1500 ms. An old item was then presented and participants had to classify the stimulus by selecting the appropriate letter ("M" or "H") on the keyboard. The stimulus disappeared from the screen and feedback pertaining to accuracy was given. Participants were instructed to answer as quickly as possible while being as precise as possible. The inter-stimulus interval was 1000 ms.

In the transfer phase, the four positive match and four negative match items were presented once each in a random order. The old items were not presented in the test phase. Hence, in the following analyses, the last block of training items was compared to transfer items to assess differences between training and transfer items, and positive and negative items. Given this order, the match items cannot have any influence on the performance obtained with the old items. This phase of the experiment proceeded like the training phase except that no feedback was given.

The last phase was an explicit recognition memory test. Participants were shown the eight old items, the eight match items and the eight new items in a random order. Their task was to determine whether each particular item had been seen in the first part of the experiment (the training phase), the second part of the experiment (the transfer phase) or whether the item was new. Responses were given by selecting the appropriate number on the keyboard (1, 2 or 3). No feedback was given concerning response accuracy.

Results

Classification task

Error rates and response times obtained on the last block of the training phase and in the transfer phase were submitted to 2 X (2 X 2 X 2) ANOVAs involving one between-subjects factor: condition (10 learning blocks vs. 20 learning blocks), and three within-subjects factors: category (Maurice vs. Henri), phase (old vs. match), and value (positive vs. negative). This design is identical to that used in Experiment 1 except for the category factor, which was not included in earlier analyses. Two participants were dropped from all analyses because their average response times were three standard deviations above the sample average. For each participant, trials (correct and incorrect) for which response times were three standard deviations above the person's average were also excluded (1.2% of the sample). Finally, error trials were eliminated for response time analyses. This created empty cells in the response time analyses for three additional participants in the 20 block condition. They were also dropped from the ANOVA.

Errors.

The error data pertaining to condition, phase and value are presented in Figure 14. As can be seen, error rates were all below or equal to 6%. There was no interaction between condition, phase and value $\underline{F}$ (1, 76) = 0.748, $\underline{MS_e}$ = 0.01, nor was there any

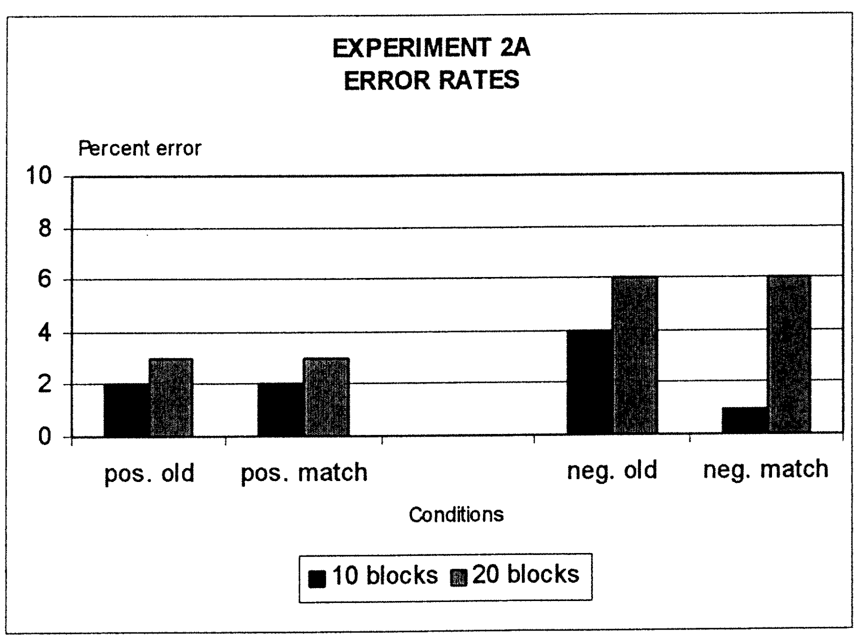


Figure 14. Error rates after ten and twenty blocks of training in Experiment 2A.

interaction between phase and value, $\underline{F}$ (1, 76) = 0.269, $\underline{MS_e}$ = 0.01. These results are similar to those found in Experiment 1.

Response times.

The response time data pertaining to group, phase and value are presented in

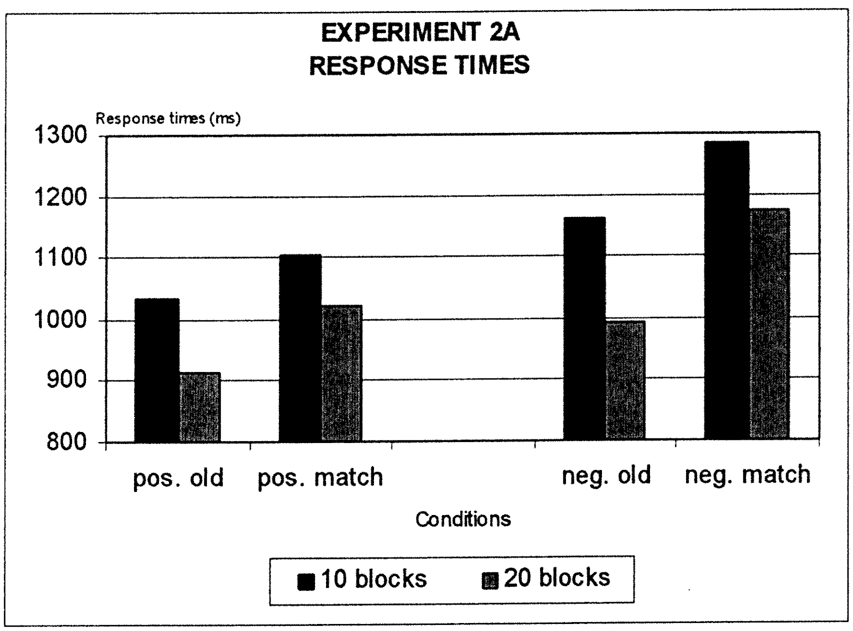


Figure 15. Response times after ten and twenty blocks of training in Experiment 2A.

Figure 15. The interaction involving phase and value approached significance, $\underline{F}$ (1, 73) = 3.108, $\underline{MS_e}$ = 46596, $\underline{p}$ < 0.09. The average response times for negative old items were 1081ms vs. 1231 ms for negative match items. These increased response times for negative match items are in the direction predicted by Brooks and his colleagues. The difference between negative match and negative old items is also much larger than the

56 ms difference found in the standard condition of Experiment 1 after 20 blocks. Furthermore, the difference between training conditions failed to be significant, $\underline{F}$ (1, 73) = 0.142, $\underline{MS}_e$ = 46596, $\underline{p}$ < .71 and the training condition did not interact with the other factors, either singly or jointly. The statistical power afforded by the data of 75 subjects and the reduced variability of having only one rule type yielded a negative match effect that was close to being significant.

Other significant results included an interaction between item value and category, $\underline{F}$ (1, 73) = 5.385, $\underline{MS}_e$ = 66052. Items that were in the Maurice category and in the positive conditions averaged 973 ms, negative Maurice items averaged 1060 ms, positive Henri items averaged 1067 ms, and negative Henri items 1251 ms. This interaction suggests that certain items took longer to classify following the given rule than other items. There was also a significant main effect for experimental phase, $\underline{F}$ (1, 73) = 14.118, $\underline{MS}_e$ = 151162 (old items were 119 ms faster than match items). This result is similar to that found in Experiment 1, though it is more difficult to interpret due to the presence of feedback for old items that was not given for match items.

Item analyses

The analyses conducted at this point have yielded results quite similar to those found in Experiment 1. Error rates and response times did not show a negative match effect. Nevertheless, the difference between negative match and negative old items for response times was 150 ms, showing a clear tendency in favor of Brooks and his colleagues' hypothesis. Yet, the effects obtained for the factors category (Maurice vs. Henri) and value (positive vs. negative) suggest that the logical structure of the rule attributes, which is not balanced over these factors, also contributed to the results. The time needed to test the rule on different combinations of attributes also seems to bring a minor contribution to performance. Hence, there is evidence to support both an exemplar as well as a rule-based interpretation of the results. To decide between the two, we looked at the response times obtained for each individual training and transfer item instead of solely focusing on the means.

The top portion of Figure 16 contrasts the response times obtained with the old items and their corresponding match items after 10 training blocks. The bottom portion contrasts the same response times after 20 blocks. The labels in the Figure identify the items by the value of their three rule attributes. The left-hand part of the Figure shows positive items whereas the right-hand part shows negative items. The two full lines show response times for old items in ascending order and the two dotted lines show response times for their corresponding match items.

As can be seen, the results are similar in both training conditions. Negative match items include two exemplars (100 and 001) that have longer response times than

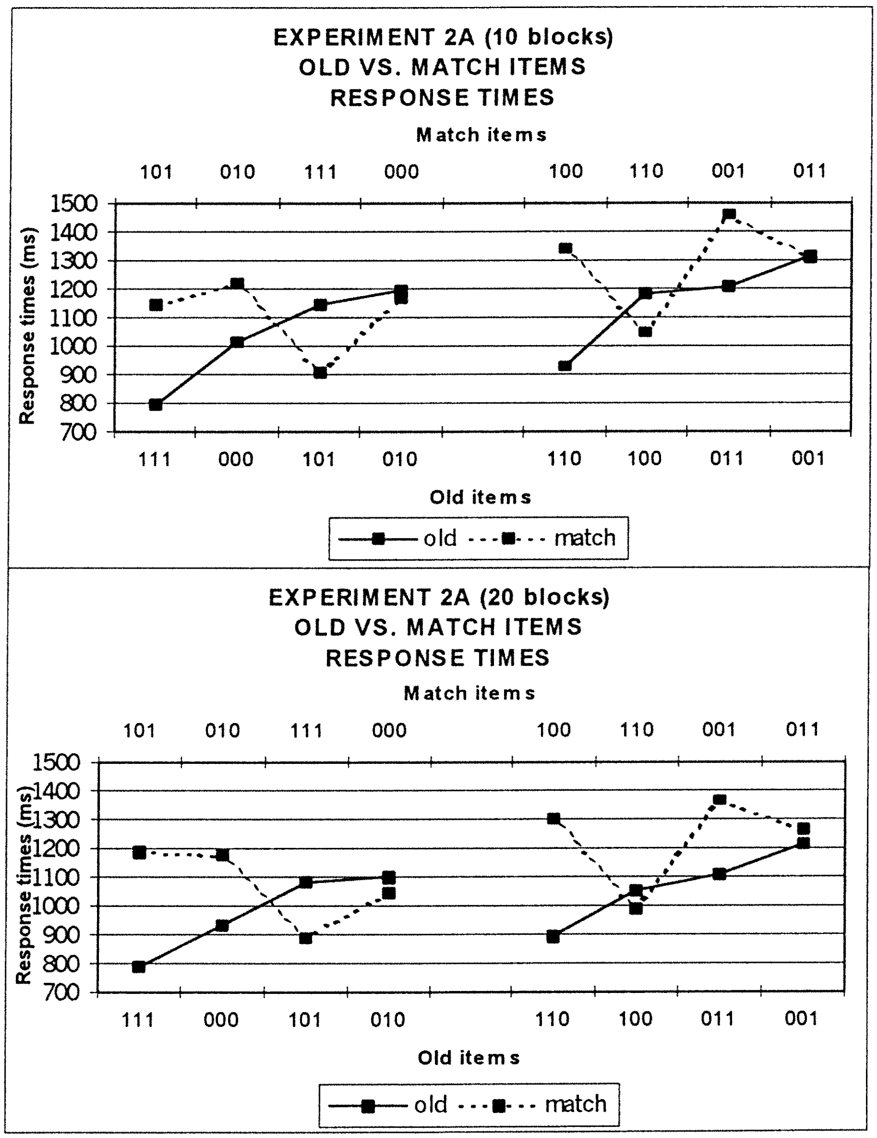


Figure 16. Item analyses for response times comparing old and match items in Experiment 2A.

their corresponding old items (110 and 011 respectively). Surprisingly, positive match items also include two exemplars (101 and 010) that have longer response times than their corresponding old items (111 and 000 respectively). Both

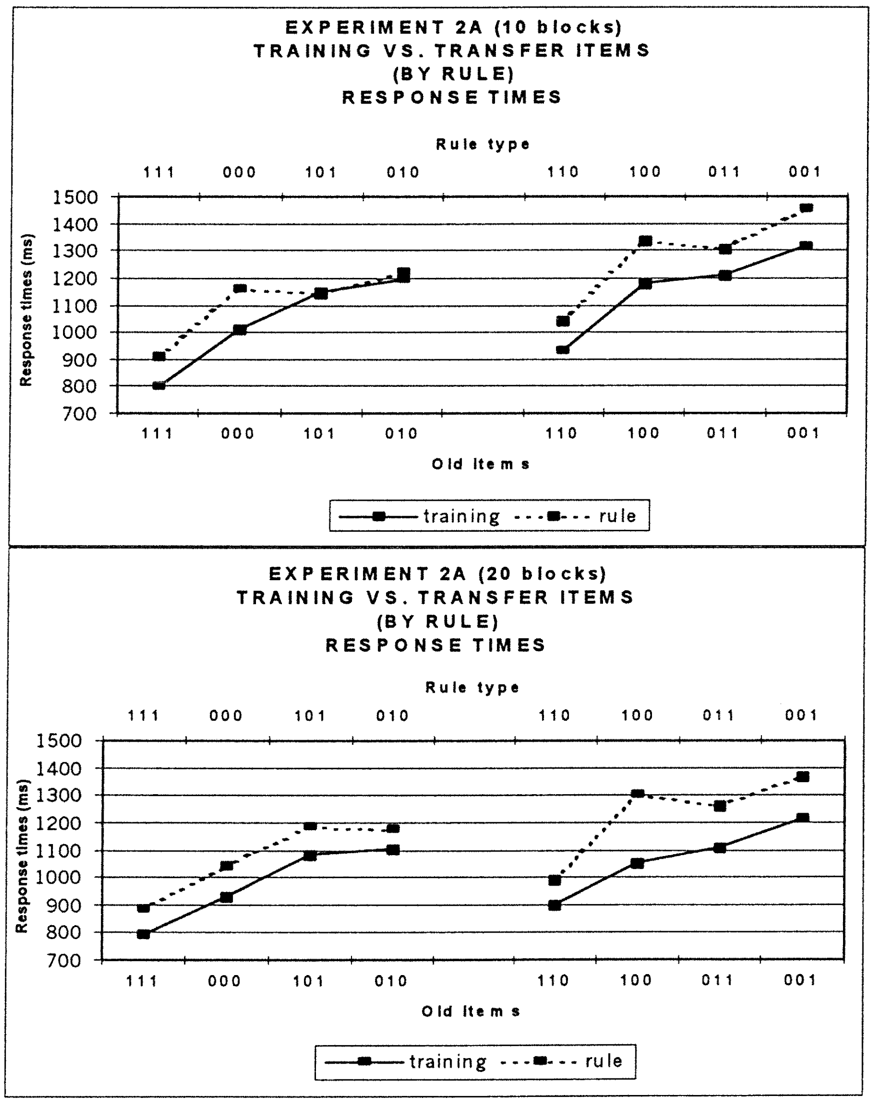


Figure 17. Item analyses for response times comparing training and transfer exemplars by rule attributes only in Experiment 2A. Non-diagnostic attributes are ignored.

positive and negative match items include one exemplar that has shorter response times than their corresponding old items, although the difference is somewhat more pronounced in the positive condition (see the difference between items 111 and 101 vs. the difference between 110 and 100). Finally, both conditions also include one match exemplar with approximately the same response times as the corresponding old exemplar. The correlation between old and match items is r = 0.06 after 10 blocks and r = -0.11 after 20 blocks. By contrast with what Allen & Brooks' theory would have led to expect, this item analysis shows no obvious relationship between the response times obtained with the old and their corresponding match items.

The training and transfer items were also compared on the basis of the three rule attributes only. For example, this analysis compares item 1 (1 1 1 0 0) with item 9 (1 1 1 1 1) whereas the previous analysis compared item 1 with item 10 (1 0 1 0 0). This comparison is based on the assumption that only the rule attributes are important in explaining the differences in response times between positive and negative items. Note that matching the items in this way does not change the mean results reported in the classification task section because this analysis involves the same items for each type of stimuli: positive old (1, 3, 6 and 8), negative old (2, 4, 5 and 7), positive match (9, 10, 13 and 15), and negative match (11, 12, 14 and 16). The results of this analysis are presented in Figure 17. The results obtained after 10 blocks of training are shown in the top part and the results obtained after 20 blocks are shown in the bottom part. Once more, the left-hand part of the Figure shows positive items whereas the right-hand part shows negative items. The full lines show response times for training items in the same ascending order as before. However, the dotted lines now show response times for transfer items that have the same combination of diagnostic attributes (the same rule type). All transfer items, both positive and negative, have longer response times than their corresponding old items except for two positive exemplars in the 10 block condition (101 and 010) that have approximately equal response times to their corresponding training items. The correlation between training and transfer items is $r = 0.94$ after 10 blocks and $r = 0.95$ after twenty blocks. Hence, this analysis shows a strong relationship for response times between training and transfer items.

Memory test

Table 12

Experiment 2A memory test results after 10 blocks and 20 blocks of training.

	Responses		
Conditions	Old	Match	New
10 blocks			
Old	60%	24%	16%
Match	63%	18%	19%
New	60%	24%	16%
20 blocks			
Old	60%	24%	16%
Match	66%	17%	17%
New	66%	18%	16%

Mean percentage of old, match and new responses given to the various types of stimuli are presented in Table 12. The data from the 10 block training condition are in the top portion and those of the 20 block condition are in the bottom portion. The results are almost identical in both training conditions. The mean percentage of correct responses is approximately 60% for training items, 18% for match items and 16% for new items. Though these results could initially be taken as evidence that the participants have some memory of training phase items, the results in fact show only a response bias much like that found for participants in the standard condition of Experiment 1. Indeed, among all the item types, only 33% of the exemplars selected as being training items had in fact been seen in the first phase of the experiment. Clearly, this is the level predicted by chance. The rest of the "old" responses were false alarms equally distributed over items belonging to the transfer phase and the memory phase. Hence, participants favored responding that items came from the training phase without showing any capacity to explicitly discriminate the exemplars.

Discussion

As in Experiment 1, there was little evidence to support the idea that exemplar memory was influencing the application of the rule. Once more, error rates were very small and the difference between negative old and match items for response times was not statistically significant. However, the 150 ms difference found after 20 blocks of training could be taken as weak evidence that the conflict between exemplar memory and rule application proposed by Brooks and his colleagues is taking place. However, the item analyses showed very little relationship between training and match items.

Remember that, according to Brooks and his colleagues, negative match items generate slower response times because they are similar to items seen in the training phase but belong to the opposite category. This was found to be the case for two negative match items, but it was also true for two positive match items. Also, the correlation between old and match items was close to zero. Hence, there was no systematic disadvantage for negative match items. Furthermore, some of the data found with the positive items are highly problematic for Brooks and his colleagues' thesis. For instance, take item number 1 (see the top portion of Table 8). Its logical description is 1 1 1 0 0. This is the prototype of the Maurice category, having all three diagnostic attributes. Its matched item is item number 10 and its logical description is 1 0 1 1 1. We observe that the average response time for the matched item is 400 ms slower than for the old item. This seems difficult to reconcile with Allen & Brooks' thesis because it would seem that the memory of the prototypical Maurice from the training phase is adversely influencing the time to classify its matched item in the same category. Such results are hardly compatible with similarity-based accounts of categorization that include non-diagnostic attributes in the computation of similarity. It seems that there is even less support for Allen & Brooks' explanation of the negative match effect in the item analyses that there is in the averaged data.

A more reasonable explanation is that the application of the rule to the three diagnostic attributes determines response times. Indeed, when training and transfer items are compared by rule attributes only, a clear relationship between the two emerges. It may be seen as an instance of applying the verbal rule to the features composing the stimuli. This type of explanation has been successful in accounting for response time data in previous research. For example, Martin and Camarazza (1980), in evaluating the role of necessary and characteristic attributes in categorization, showed that participants' response times resulted from the sequential application of a classification rule following a decision tree. Each branch in the tree represented the time associated with testing the presence of a given rule attribute and the total time to apply the rule was simply the sum of all the tests.

The results are also compatible with a similarity-based account, such as the GCM, provided that the non-diagnostic attributes receive little or no weight. The challenge facing rule-based and similarity-based accounts is to explain the differences in response times among old (and among new) items. Unfortunately, our results do not unambiguously support one view over the other. Proponents of a similarity-based account would be comforted by the fact that the prototypes of both categories yielded shorter response times than most of the other items. Proponents of a rule-based account could argue that the Maurice category yielded shorter response times, on average, than the Henri category because the items in the Maurice category contain more of the attribute values specified in the rule than the items in the Henri category. This is the case because the Henri category was defined by alternate values not specified in the verbal rule given to the participants. Rule-based and similarity-based mechanisms might even have been cooperating in producing the results obtained, provided once again that the non-diagnostic attributes are ignored by a similarity-based process. What the results of Experiment 1 and of the present experiment clearly rule out is that non-diagnostic attributes generate a conflict between rule-based and similarity-based processes when no special instruction are given to direct attention on these attributes.

Experiment 2B

Up to this point, several variations of Allen and Brooks' (1991) rule paradigm have been used. The number of training trials has been varied; slightly different stimuli have been used; and a naming condition was added. However, no attempt has been made to reproduce Allen and Brooks' (1991) induction results. As discussed in the Introduction, it is well known that in induction tasks, previously learned exemplars influence the classification of transfer material. To show that the results obtained with the rule paradigm could not be explained solely by similarity-based processing, Allen and Brooks (1991, Experiment 1) ran an induction task, which used the same categorical structure and stimuli as the rule task. Participants were shown the *builders* and the *diggers*, and they were told to infer the category membership of the creatures with the help of the feedback provided. Otherwise, the experiment was identical to the rule paradigm experiment. The results showed a very strong negative match effect for error rates. Participants made 86% errors on the negative match items compared with only approximately 12% for negative old items. There was no difference in response times between negative match and negative old items, however. Allen and Brooks argued that this was to be expected because the effect was entirely similarity-based. Hence, participants did not hesitate to *wrongly* categorize the negative match items on the basis that they shared four out of five attributes with a corresponding training phase exemplar. Regehr and Brooks (1993, Experiment 2A) replicated the effect with similar stimuli, but the difference in error rates between old and negative match items was less extreme (16% vs. 50%). Brooks and his colleagues' did not analyze their results as a function of the logical structure of the stimuli. Hence, the goal of Experiment 2B is to verify that a negative match effect for error rates may be obtained in the induction task with the present stimuli and categorical structure both in averaged data and items analyses. We will show that participants' behavior is very different in the induction paradigm and in the rule paradigm.

Method

Participants.

Eighty students at the Université de Montréal participated in the study. Participants in the 10 block condition received 3$ as compensation for their time and those in the 20 block condition received 5$. However, one participant was dropped from all analyses because the average response time data were three standard deviations above the group mean.

Materials.

The material was identical to that used in Experiment 2A except that two classification rules were used: *rule A* (parallelogram-shaped body, small circular spots, and circular tail) and *rule B* (curved body, small circular spots, and circular tail). Notice that rule A was the one used in Experiment 2A. Rule B was used to counterbalance items values so that all the items that are positive with rule A become negative and vice versa. Although the rules defined category membership, participants were not aware of them.

Procedure.

Participants were randomly assigned to one of two training conditions: 10 block or 20 blocks, and to one of the two rule conditions. The participants were tested individually on 386dx or 486dx IBM compatible computers using MEL Professional v.2.01 (Schneider, 1989).

The experiment was conducted in three phases. In the training phase, the participants were told that they would see creatures from two families (the *Maurice* and the *Henri*) and that their task was to determine to which of these families each creature belonged. However, they were informed that, initially, they would have no basis on which to base their decisions and that they would have to guess. They were told that they would receive accuracy feedback after each response and that this would eventually help them to classify the creatures in the correct family. Then, participants in the 10 block condition were given 80 trials of training, and participants in the 20 block condition were given 160 trails. Each block involved the presentation of the eight old items in random order. Otherwise, the training phase, the transfer phase and the memory phase proceeded as in Experiment 2A.

Results

Classification task

Error rates and response times obtained on the last block of the training phase (old items) and in the transfer phase (match items) were submitted to 2 X (2 X 2) ANOVAs involving one between-subjects factor: condition (10 learning blocks vs. 20 learning blocks), and two within-subjects factors: phase (old vs. match), and value (positive vs. negative). Error trials were eliminated for response time analyses. This created empty cells in the response time analyses for four participants in the 10 block condition and 13 participants in the 20 block condition. Hence, they were dropped from the ANOVA.

Errors.

The error data pertaining to condition, phase and value are presented in Figure 18. It can be observed that the error rates are much higher than in the rule task. The average error rate after 10 blocks of practice was 38% (SD = 28%) and it was 34% (SD = 31%) after 20. Hence, the task was difficult.

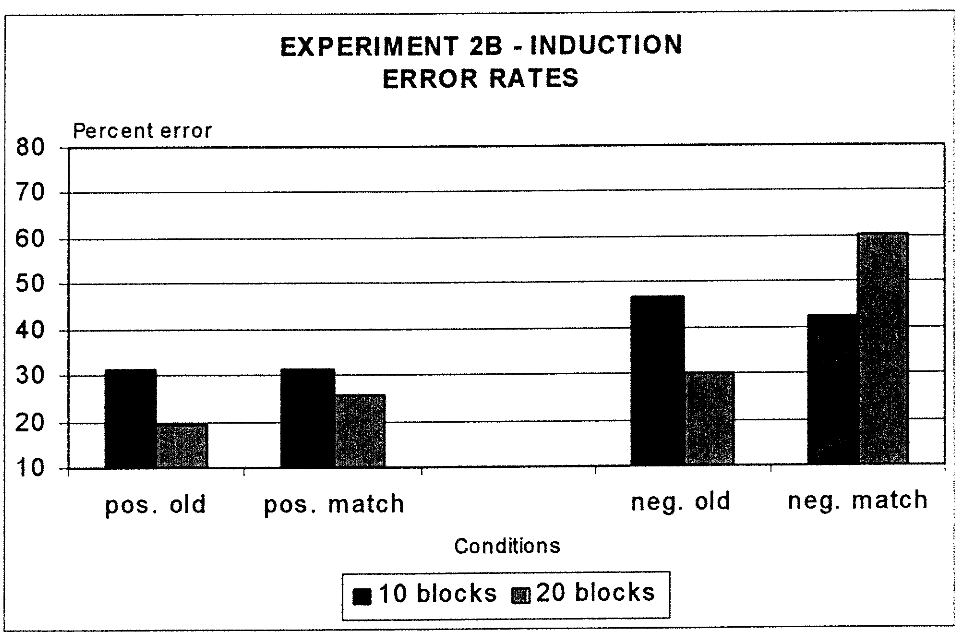


Figure 18. Error rates in the induction task at 10 blocks and 20 blocks in Experiment 2B.

The interaction between condition, phase and value, $\underline{F}$ (1, 77) = 5.491, $\underline{MS_e}$ = 0.40, $\underline{p}$ < .02 was significant, but the interaction between phase and value was not, $\underline{F}$ (1, 77) = 2.509, $\underline{MS_e}$ = 0.18, $\underline{p}$ < .12. A decomposition of the ANOVA showed that the phase by value interaction was not significant after 10 blocks, $\underline{F}$ (1, 77) = 2.2, $\underline{MS_e}$ = 0.25, $\underline{p}$ < .15, but that it was highly significant after 20 blocks, $\underline{F}$ (1, 77) = 19.9, $\underline{MS_e}$ = 2.34, $\underline{p}$ < 0.0001. For negative items, the difference between match and old was 30%, whereas the difference was only 7% for positive items. Hence, contrary to Allen and Brooks (1991, Experiment 1) and Regehr and Brooks (1993, Experiment 2A), a negative match effect was not found after a short period of training, even though this experiment included ten blocks of training instead of the five used by Brooks and his colleagues. This discrepancy may be explained by Allen and Brooks' use of a sixth non-diagnostic attributes, the backgrounds. Their role in generating negative match effects will be discussed in Experiment 3. Nonetheless, after 20 blocks of training, the expected interaction was clearly present.

Response times.

The response time data related to condition, phase and value are presented in Figure 19. As was the case for Allen and Brooks (1991, Experiment 1) and Regehr and Brooks (1993, Experiment 2A), there were no significant phase by value interaction, $\underline{F}$ (1, 60) = 0.246, $\underline{MS_e}$ = 177495, $\underline{p}$ < .62. This was the case because the difference between old and match items was equally large for positive and negative items. The interaction involving condition, phase and value was not significant either, $\underline{F}$ (1, 60) = 0.182, $\underline{MS_e}$ = 131724, $\underline{p}$ < .671. Contrary to what was found in Experiments 1 and 2A,

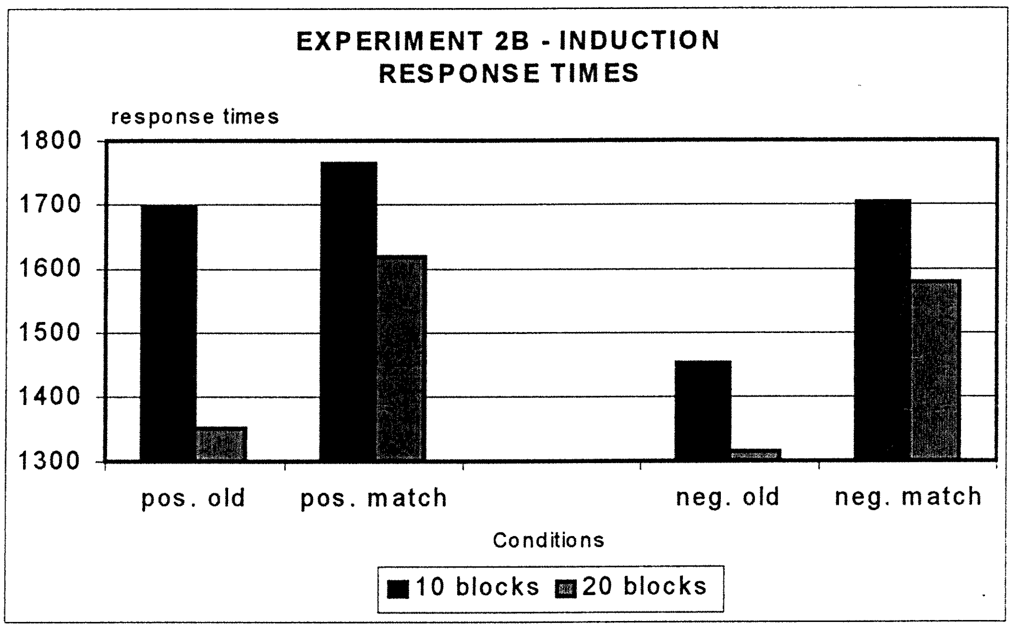


Figure 19. Response time data in the induction task after 10 and 20 blocks of training in Experiment 2B.

there was no main effect for phase, $\underline{F}$ (1, 60) = 0.7, $\underline{MS_e}$ = 2913092, $\underline{p}$ < .237.

Item analyses

The results in the induction paradigm are quite different from those obtained with the rule paradigm. After twenty training blocks, the error rates for negative items were twice as large in the last transfer phase as they were in the last training block. By contrast, for positive items, the error rates were only about 30% larger on the transfer phase items than on the last training block. This shows the strong influence of previously learned exemplars on categorization behavior. This influence may also be seen in an analysis, which compares error rates by item.

Figure 20 shows the error rates obtained with rule A. The results for rule B are in the Appendix D. The results obtained after ten blocks with the old items and their corresponding match items are in the top portion. The bottom portion shows the same contrast after 20 blocks. The labels in the Figure identify the items by the value of their three diagnostics attributes. The left-hand part of the Figure shows positive items whereas the right-hand part shows negative items. The two full lines show error rates for old items in ascending order and the two dotted lines show error rates for their corresponding match items.

After ten blocks, the results are very similar to those found in the rule paradigm. There is no systematic disadvantage for negative match items. The correlation computed between old and match items was $r = -0.7$ for rule A and $r = -0.14$ for rule B. After 20 blocks, the pattern of results that corresponds to what would be expected from the hypothesis that exemplar memory is influencing the classification of transfer material appears. Error rates for all negative match items are larger than the corresponding old item, whereas error rates for positive match items

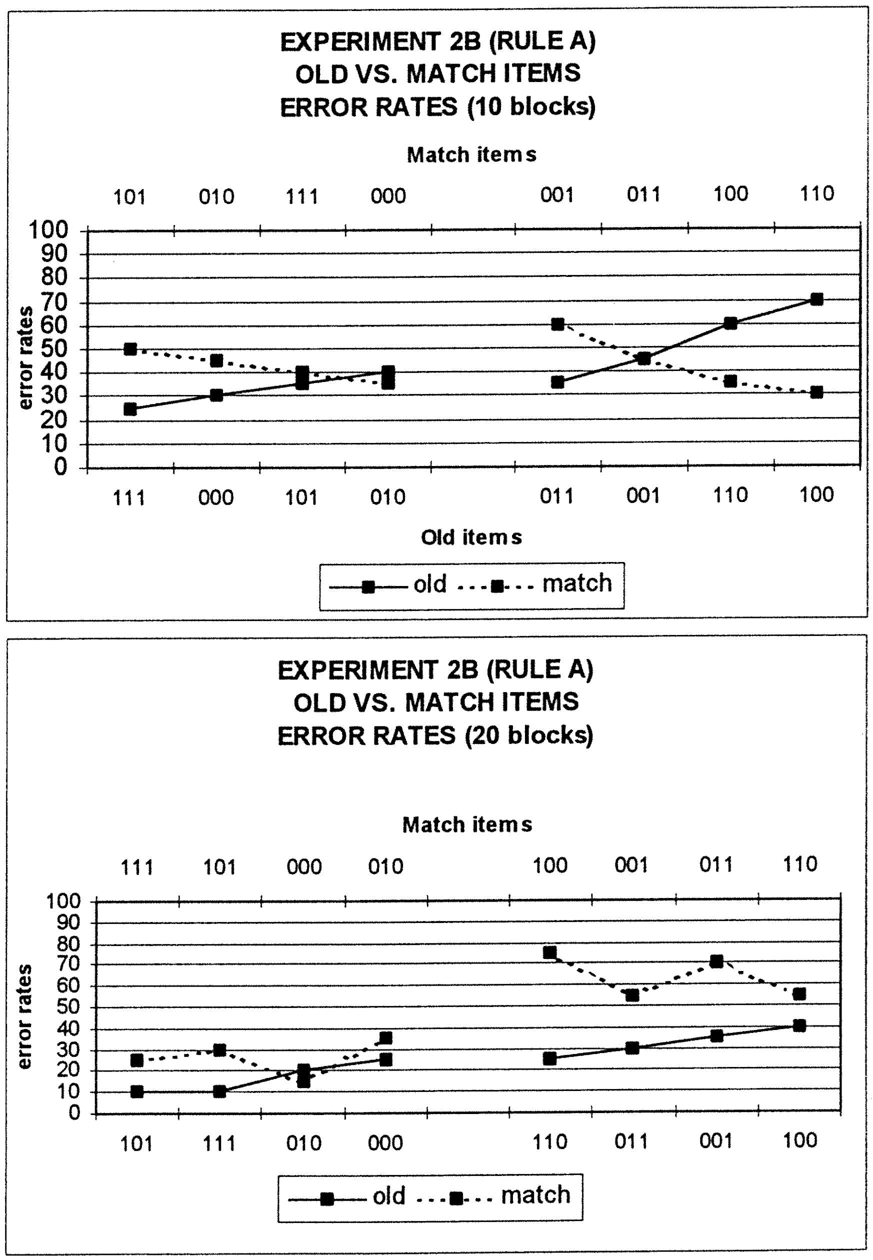


Figure 20. Item analyses for error rates comparing old and match items in Experiment 2B.

are comparable to the corresponding old items. The correlation between old and match items was r = 0.68 for rule A and r = 0.08 for rule B.

Figure 21 shows the analysis of rule A data when comparing training and corresponding transfer items of the basis of the rule attributes only (the graphs for rule B are in the Appendix E). The 10 block training condition is in the top portion and the 20 block one is in bottom portion. The left-hand part of the Figure shows positive items whereas the right-hand part shows negative items. The two full lines show error rates for training items in ascending order and the two dotted lines show error rates for their corresponding transfer items. In opposition to Experiment 2A, this analysis does not appear to offer a better description of the relationship between training and transfer material (with perhaps the exception of the rule B comparison after 10 blocks, see Appendix E). After ten blocks, as in the old vs. match comparison, there is no obvious way to relate training error rates to transfer error rates. The correlation between training and transfer items is r = -0.31 for rule A and r = 0.79 for rule B. After 20 blocks, the systematic disadvantage for negative match items is as clear as it was with the old vs. match item analyses. The correlations were r = 0.75 for rule A and r = 0.25 for rule B. Although the data corresponds to expectations concerning the difference between negative old and negative match items, it may be observed that the correlation data are much less impressive than in those found in Experiment 2A.

Thus, with twenty blocks of practice, both the averaged data and the item

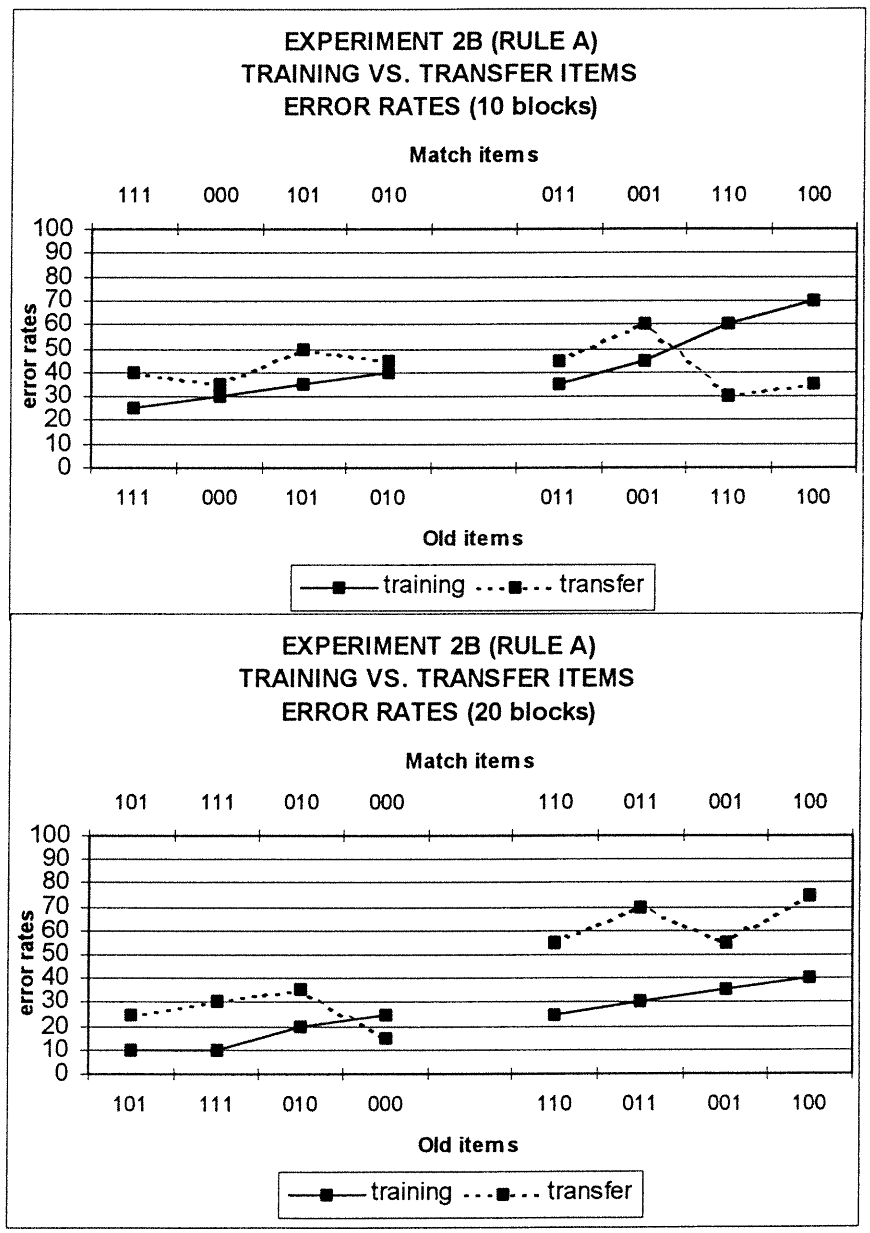


Figure 21. Item analyses for error rates comparing training and transfer exemplars by rule attributes only in Experiment 2B. Non-diagnostic attributes are ignored.

analyses support a similarity-based explanation of the negative match effect in the induction task. Yet, it is tempting to consider the possibility that, as in Experiment 2A, rule application behavior is sufficient to explain the data.

In concept formation tasks, it is well known that participants start out with simple, most often single attribute rules (Bruner et al., 1956, Ahn and Medin, 1992, Nosofsky et al. 1994). Then, with practice and feedback, the number of attributes used increases if the sought after rule requires it. In this experiment, participants needed to focus on three attributes to successfully classify all items in the training phase. However, they did not need to be the three rule attributes. In Table 13, all the three attribute and two attribute rules that could theoretically be held by a participant during learning are listed in the leftmost column (assuming each attribute value is associated with the correct category). These rules are identified by the number of the attribute involved. The attributes numbered 1 to 5 correspond to body type, spots, tail type, texture, and color respectively. For instance, the rule "1-2-3" includes the three diagnostic attributes. In the four right-hand columns are given the error rates that would be generated for each item type by participants following each rule consistently. It is of great interest to see that eight out of the ten three-attribute rules yield perfect performance with training stimuli. Hence, if a given participant chooses one of these rules, we may say that he has induced an optimal rule for the given stimuli set. However, these same rules generate very different error patterns once we turn to match items. As can be noticed, none of the rules, except for rule "1-2-3" leads to perfect transfer phase performance. The rules “1-2-4” and “2-3-5” yield 50% error rates for both positive and negative match items; the rule “2-4-5” yields a 100% error rate for positive match and a 0% error rate for negative match items; and rules “1-3-4”, “1-3-5”, “1-4-5” and “3-4-5” yield a 0% error rate for positive match and a 100% error rate for

Table 13.

List of three and two attributes rules that can be applied to the Experiment 2B stimuli during the learning phase of the induction task.

	Items types			
	Positive old	Positive match	Negative old	Negative match
Three attribute rules				
1-2-3	0%	0%	0%	0%
1-2-4	0%	50%	0%	50%
1-2-5	25%	50%	25%	50%
1-3-4	0%	0%	0%	100%
1-3-5	0%	0%	0%	100%
1-4-5	0%	0%	0%	100%
2-3-4	25%	50%	25%	50%
2-3-5	0%	50%	0%	50%
2-4-5	0%	100%	0%	0%
3-4-5	0%	0%	0%	100%
Two attribute rules				
1-2	25%	25%	25%	25%
1-3	0%	0%	50%	50%
1-4	25%	25%	25%	75%
1-5	25%	25%	25%	75%
2-3	25%	25%	25%	25%
2-4	25%	75%	25%	25%
2-5	25%	75%	25%	25%
3-4	25%	25%	25%	75%
3-5	25%	25%	25%	75%
4-5	50%	50%	50%	50%

negative match items. Thus, if participants were following such rules, it is possible that the average error rates both combined and concealed this rule-governed behavior.

An analysis of error rates on match items as a function of error rates on old items

To evaluate this hypothesis, the individual error rates for the 39 participants in the 20 block condition (which included rules A and B) were examined. Table 14 gives the distribution. As shown in the leftmost column, the rates on old items ranged between 0% and 62.5%. The numbers in the parentheses show the number of participants that fall in each error rate category. The two right-hand columns show the averaged error rates obtained with each of the subgroups of participants for positive and

Table 14

A distribution of participants' error rates on match items as a function of error rates on old items for Experiment 2B.

negative match items.

Distribution of error rates on old items	Averaged match phase error rates for old phase error rate categories	
	Positive	Negative
0% (9)	25%	72%
12,5% (6)	12,5%	75%
25% (12)	27%	52%
37,5% (2)	12,5%	50%
50% (9)	39%	53%
62,5% (1)	0%	50%

The results of interest are those of participants who made no errors during the training phase. Notice that their average negative match item error rates reach almost

75% compared to only approximately 25% for positive match items. This data appears to be strong evidence in favor of Brooks and his colleagues' hypothesis. However, of these nine participants, five showed a pattern of error that was perfectly consistent with the application of the rules "1-3-4", "1-3-5", "1-4-5" or "3-4-5" which yield a 0% error rate for positive match and a 100% error rate for negative match items. One more participant was nearly consistent with the last four rules enumerated. Notice that these participants generated huge negative match effects. Yet, it does not appear necessary to postulate a multiple-system model of categorization to explain their results. Two more participants were nearly consistent with the rules "1-2-4" and "2-3-5" which yield 50% error rates for both positive and negative match items. Finally, one participant was perfectly consistent with the rule "2-4-5" which yields a 100% error rate for positive match and a 0% error rate for negative match items. This participant's results are in the opposite direction with regards to Brooks and his colleagues' hypothesis. The entire set of individual error rates as a function of item type are given in Appendix F. The data for participants who did not obtain a 0% error rate for old items are more difficult to analyze unambiguously, because the participants are still in the process of inferring an adequate rule. However, it may be reasonably assumed that these participants are behaving following the same rule-based pattern.

Hence, these participants make a good case for the hypothesis that error rates with match items were directly related to the rules being used at the end of the training phase. Alternatively, a similarity-based explanation, involving a memory of the exemplars including only those attributes which were focused upon though rule induction, could also possibly be used. In this case, the exemplars memorized in the learning phase could be leading the participants to classify similar transfer exemplars in the "wrong" categories (that is, following the experimenter's design). However, as we have already mentioned, a dual mechanism explanation is not required.

Finally, this data also explains why the correlations between training and transfer items were very high in the item analysis of Experiment 2A ($r = 0.94$ after 10 blocks and $r = 0.95$ after 20 blocks), while being variable in Experiment 2B. In Experiment 2A, participants were forced to adopt the same classification rule because of the rule paradigm experimental instruction. However, in Experiment 2B, we have shown that participants were using many different rules. This greater variability in strategies caused the correlations to be smaller. Yet, because some participants adopted a training phase rule that disadvantaged negative match items, the item analyses (and the classification task data) nonetheless showed worst averaged performances for negative match items compared to negative old items.

Memory test

Mean percentages of old, match and new responses to the various stimulus types is presented in Table 15. The data from the 10 block training conditions are in the top portion and those of the 20 block condition are in the bottom portion. It is clear that participants did not recognize the old stimuli. Performances were near random after ten blocks. Once more, only a response bias favoring old items was found after twenty blocks. This supports the idea that the participants were not memorizing all five attributes comprising the exemplars, which are necessary to successfully discriminate the old, match and new exemplars. Rather, participants seemed to have been focusing on a limited subset of attributes as suggested previously.

Table 15

Experiment 2B memory test results after 10 and 20 blocks of training in the induction task.

	Responses		
Conditions	Old	Match	New
10 blocks			
Old	58%	21%	21%
Match	53%	26%	21%
New	53%	24%	23%
20 blocks			
Old	61%	15%	24%
Match	62%	18%	20%
New	65%	15%	20%

Discussion

Experiment 2B replicated the induction task results obtained by Allen and Brooks (1991) and Regehr and Brooks (1993) although more practice trials were needed. It also established that it was possible to show the influence of learning training exemplars on the categorization of transfer material in both averaged data and item analyses. However, the explanation required only a single mechanism explanation. As in Experiment 2A, the rule-based view was compelling. Participants who succeeded in perfectly classifying old stimuli showed behavior that was consistent with rule application. If we relate this to the main topic of understanding the relationship between similarity-based and rule-based mechanisms of categorization in the rule paradigm, an important doubt as to the validity of Brooks and his colleagues' hypothesis appears.

The evidence is the following. In their original experiment (Experiment 1), Allen and Brooks (1991) obtained strong negative match effects for error rates and response times. Since then, a complete replication has not been possible. In the same article, the authors were unable the reproduce the effect for response times. Then, in Regehr and Brooks (1993, Experiment 2A), they were unsuccessful in reproducing the effect for error rates and they obtained the effect for response times using a less conservative analysis. In Experiment 1 of the present study, the effect does not appear for either of the dependent variables in the standard condition. Also, the item analyses of Experiment 2A showed a strong relationship between response times applied to old items and those applied to match items which necessitated no similarity-based explanation, nor supposed conflicting mechanisms. Finally, the induction task of Experiment 2B showed that it was possible to obtain negative match effects in both averaged data and item analyses, but that an alternate rule-based explanation was available.

Hence, negative match effects in the rule paradigm have been far from robust. Yet, one crucial question remains. Why did Allen and Brooks obtain such convincing results in their original experiment? The answer resides in two of their methodological choices: the use of backgrounds to make the task ecological and the presentation order. These choices will be the focus of Experiment 3.

Experiment 3

Experiments 1 and 2A in this study and all the experiments presented in Regehr and Brooks (1993) differ methodologically from those presented in Allen and Brooks (1991) on two points: the use of backgrounds and the presentation order for transfer material.

In their original experiments, Allen and Brooks wanted their task to be realistic and ecological. They wanted to create a categorization experiment that was closer to real world tasks than the average laboratory task. Hence, they presented all stimuli on four types of salient background showing the creatures' living environments (see Figure 3). By definition, these backgrounds were non-diagnostic (see Table 3), that is, they were unrelated to category membership. However, they may be distinguished from the two non-diagnostic attributes that compose the creatures in two important respects. First, in the training phase of the experiment, after the participants had classified the stimuli on viewing the first slide, the participants were told to "remember how the animal built or dug (p. 6)" while viewing the next two slides. Then, in the transfer phase of the experiment, "...the first slide showed only the background on which the upcoming test item would be displayed. The subjects were simply to look at this background and indicate when they were ready for the second slide. The second slide showed the same background with a pair of animals on it. [...] They were also told that they might be able to use the first slide to anticipate which items were most likely to appear on the background...(p.6)'. Therefore, backgrounds had a special status compared with other non-diagnostic attributes in that during the training and transfer phases, the experimental instructions directed the participants' attention directly upon them. Now, it is one thing to claim that non-diagnostic attributes, to which no attention is given, influence categorization. It is another to say that an attribute, to which much attention is given, influences categorization even though it is non-diagnostic. Indeed, exemplar-based models would predict that giving attention to non-diagnostic attributes would increase their weight in the calculation of similarity between exemplars and thus, this would lead to larger error rates

Nonetheless, there is no doubt that Brooks and his colleagues' thesis was that non-diagnostic attributes to which no attention is directed still influence categorization. This idea goes back to Brooks' (1978) initial conception of non-analytical processes: " ...the category membership of an item is inferred from its overall similarity to a known individual or low-level cluster of individuals, where similarity is judged on the basis of aspects or configurations of the stimulus that are *not* [Brooks' italics] weighted for their criteriality for the particular concept being considered (p. 180)'. Allen continues to make the claim very explicitly in a subsequent unpublished manuscript: "the perceptual system normally encodes both the diagnostic and non-diagnostic features of a stimulus array and the record that is formed contains both types of information (Wagner and Allen, p.4)". Thus, it is possible that authors' desire to make the task ecological, inadvertently created a situation in which negative match effects were due to an attribute that was not part of the stimulus and to which participants' attention was directed. The fact that Regehr and Brooks (Experiment 2A) and the previous experiments did not include backgrounds and that they mostly failed to replicate the original results supports this hypothesis.

The other methodological difference is the presentation order of the material during the transfer phase. Remember that Allen and Brooks presented old items between the positive match and negative match items during the transfer phase (see Figure 4). They justified this procedure in the following way: "This separation into a positive and negative phase was intended to allow us to evaluate a possible generalized caution effect when the subject discovered the presence of negative matches (p. 6)". Such justification clearly rests on the a priori assumptions that participants memorize all attributes of the stimuli, both diagnostic and non-diagnostic. Otherwise, participants cannot tell the difference between positive and negative match items. The memory test results reviewed thus far show that the assumption is unwarranted in the standard version of both the rule and induction paradigms. Without any explicit memory of the exemplars, the participants cannot become cautious. So, the contrived presentation order of the material during the phase was probably not only unnecessary, but this procedure may also have biased the results. The test phase in Allen and Brooks contained 40 trials and negative old items were presented on trial 9 through 25. For instance, suppose that the general speed of responding still increases during the transfer phase. Then, there could be a greater difference in response times between the negative match and negative old items if the latter are presented late in the transfer phase and the later are presented early. This is not the case (see Figure 4). Naturally, this problem is eliminated when items were presented in a random order.

Experiment 3 duplicates the standard rule paradigm used in Experiment 1. However, the two contentious methodological points discussed above are evaluated. In one condition, backgrounds were introduced and instructions were given to attract attention to them. In this *background* condition, the presentation order for transfer material was random. A *presentation order + background* condition was also used in which the presentation order of the stimuli during transfer was as in Allen and Brooks' original study. It will be shown that these two methodological choices generated the negative match effects for error rates and response times and that, therefore, Brooks and his colleagues conclusions about the rule paradigm are unwarranted.

Method

Participants.

Sixty-four students at the Université de Montréal participated in the study. Each received 5$ as compensation for their time.

Materials.

The stimuli were drawings of fictional animals similar to those used in Experiment 1. They were built from five binary attributes: tail type (cane-shaped or stair-shaped), back pattern (stripes or spots), head shape (parabolic or oval), body type (oval or parallelogram), number of legs (two or four); and colored backgrounds which had four possible values (blue, white, yellow, and green). Colors were chosen as backgrounds instead of the "living environments" of Allen and Brooks (see Figure 3) to better control possible differences in saliency. Obviously, it is impossible to assert that the colors used were equally salient. However, the relative saliency of backgrounds composed of igloos and large coniferous trees, such as those used in Allen and Brooks (1991), is even more difficult to control. Notice that in this experiment, all the creatures were gray to avoid confusions between the color of the creature and the color of the background. The logical description of the categorical structure is identical to that given in Experiment 1 except for the addition of backgrounds (the categorical structure for this experiment is given in Appendix G).

The first three attributes were diagnostic and they were identical to those used in Experiment 1. The last three attributes were non-diagnostic. The body types were identical to Experiment 1, color in the creatures was changed to number of legs and the colored backgrounds were added. Like the other non-diagnostic attributes, backgrounds were identical for old items and their corresponding match items.

The classifications rules given to participants were the same as those used in Experiment 1 (see Table 9). Once again, creatures were classified into the “Tremblay” and “Beaulieu” categories. Only the standard rule paradigm was used. Hence, none of the participants needed to learn first names for the creatures.

For the training and transfer phase, we chose the 16 items with the same logical description as Allen & Brooks (1991). Each transfer item was matched with a training item on every attribute except back pattern (see Appendix G). As in Experiment 1, the rule was given to participants in one of two orders: tail, back pattern and head; or head, back pattern and tail to avoid biasing response times in favor of positive items. The items 1 to 8 and items 9 to 16 served in turn as old items and match items. The sixteen remaining items were used in the memory test. These items were also matched with a training item on the diagnostic attributes. However, they had combinations of non-diagnostic attributes (including backgrounds) not found in old or match items (see Appendix G). This produces the five familiar types of stimuli: positive old, negative old, positive match, negative match, and new. Examples of the exemplars are given in Appendix H.

Procedure.

The participants were tested individually. All instructions and stimuli were presented with 386dx or 486dx IBM compatible computers using the program MEL Professional v.2.01 (Schneider, 1989).

Participants were randomly assigned to one of two conditions: the *background condition* and the *presentation order + background condition*. Participants in the *background condition* received the training phase instructions given in the *standard* condition of Experiment 1. However, they were told to take the period of time, during which the stimuli remained on display after the categorical decision had been made, to notice the colored backgrounds and relate them to the creatures with which they appeared. These instructions parallel those given by Allen and Brooks' in their Experiment 1 (1991, p. 6). The condition included the five phases described in the *standard* rule condition of Experiment 1. In order, they were: 5 blocks of practice with feedback, a match phase without feedback, 15 more blocks of practice, another match phase, and explicit recognition memory test. The only differences between the procedure used for this experiment and that used in Experiment 1 were in the match phases. In addition to the instruction given in the *standard* paradigm, participants were also told that the colored backgrounds would appear before the creatures and these could be used to anticipate the upcoming creature. These instructions also parallel those given in Allen and Brooks' Experiment 1 (1991, p. 6). Furthermore, the number of match phase trials used in Experiment 1 and in this experiment differed. In Experiment 1, old and match items were presented once in random order for a total of 16 match phase trials. In this experiment, the old items were presented four times and the match items only once for a total of 40 match phase trials. This number of trials is the same as used in Allen and Brooks (1991). The stimuli in this experiment were presented in random order.

The procedure used in the *presentation order + background* condition was identical to that of the *background* condition except for the presentation order used for match phase material. Instead of presenting the old and match items randomly, they were presented in the exact order suggested by Allen and Brooks (see Figure 4): positive old items (once), positive match items (once), negative old items (four times), positive old items (three times), and negative match items (once). The presentation order of the stimuli was randomized within the five parts of the match phases.

Results

Classification task

Separate analyses of variance (ANOVAs) were performed on the error rates and on the response times to compare performance on old and match items presented in the test phases. These analyses involved a (2) X 2 X 2 design with two within-subjects factors: phase (old vs. match) and value (positive vs. negative), and one between-subjects factor: training condition (background vs. presentation order + background). Classification data obtained after five blocks and twenty blocks of practice were analyzed separately. Three participants were dropped because of incomplete and damaged data files due to software problems and the data obtained after 20 blocks of training with a fourth participant was also dropped for the same reason. Error trials were eliminated for response time analyses. This created empty cells for 11 participants after 5 training blocks and 9 participants after 20 blocks. Therefore, these participants were dropped for the response time analyses.

Five blocks.

The five block error data pertaining to condition, phase and value are presented in the top portion of Figure 22. Error rates for negative match items approximated those found in Allen and Brooks' Experiment 1. The interaction

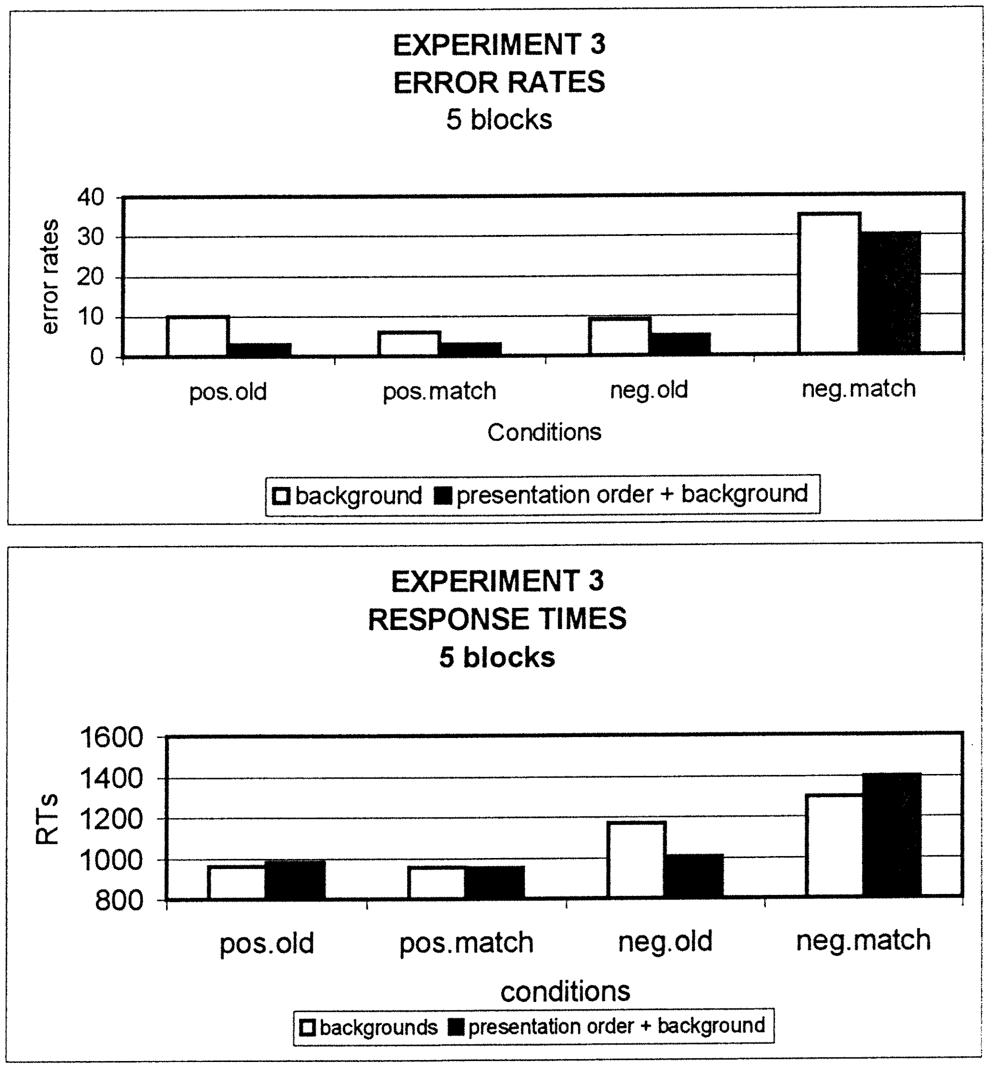


Figure 22. Error rates and response times at five blocks in Experiment 3.

between phase and value was highly significant, $\underline{F}$ (1, 59) = 23, $\underline{MS_e}$ = 1.11, $\underline{p} < 0.001$, and this result was the same in both conditions, the interaction between experimental condition, experimental phase and item value being far from significant, $\underline{F}$ (1, 59) = 0.

142, $\underline{MS}_e = .007$, $\underline{p} < 0.707$. Error rates for negative match items were 32% compared with 5% for negative old items, while being fairly constant for positive match and positive old items (5% vs. 7% respectively). As predicted, reintroducing the backgrounds into the experimental design caused the negative match effect to surge.

The response time data after five training blocks for condition, phase and value are presented in the bottom portion of Figure 22. The interaction between phase and value was not significant, $\underline{F}\ (1, 47) = 2.011$, $\underline{MS}_e = 492701$, $\underline{p} < .163$, nor was the interaction between conditions, $\underline{F}\ (1, 47) = 0.497$, $\underline{MS}_e = 492701$, $\underline{p} < .484$. Although the analysis did not yield a significant difference between negative old and negative match items, there is reason to believe that these results reproduced those obtained by Allen and Brooks (1991, Experiment 1). The difference between negative match and negative old items was 392ms in the *presentation order + background* condition. Notice that the mean difference obtained in this condition is quite similar to Allen and Brooks' result (approximately 1250ms for negative old vs. 1610ms for negative match items for a difference of 360ms, see Figure 5). However, the difference between negative match and negative old items is only 130ms in the *background* condition. Although statistically non-significant, this difference in means between the *background + presentation order* and *background* conditions is 262 ms. Hence, there is some cause to believe that presenting the stimuli in a specific order in the match phase contributes to enhance the difference between negative match and negative old items.

To explain this difference, response times were plotted by trials for both conditions in Figure 23. The trials are grouped by blocks of four. For items in the *presentation order + background* condition (the full line), the blocks show the response times for specific item types: 1 to 4 and 25 to 36 are positive old items, 9 to 24 are negative old items, 5 to 8 are positive match items and 37 to 40 are negative match items. All types of items are mixed in the *background* condition (the dotted lines) due to the random presentation order.

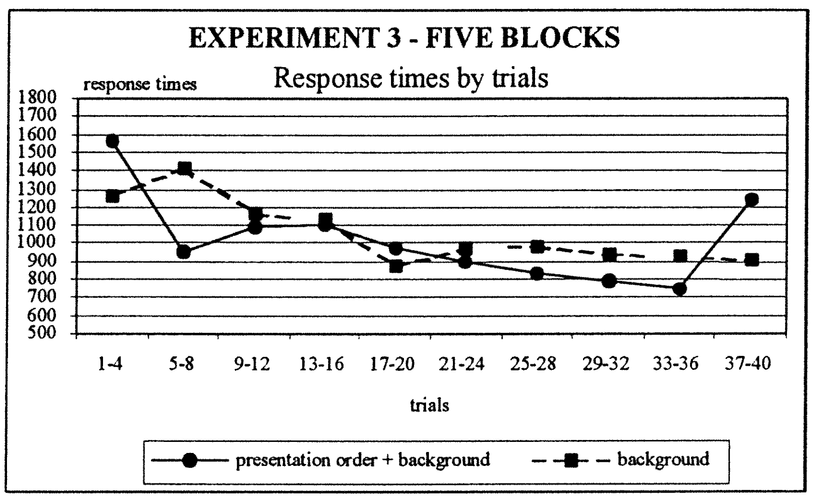


Figure 23. Response times by trials after five blocks of training in Experiment 3.

Clearly, performances across trials were not stable even though there was no feedback in the match phases of the experiment. Hence, by not randomizing the presentation order, Allen and Brooks' procedure creates a confound by which the variability related to the experimental item types is confused with the variability due to sampling in distinct portions of the match phase. Negative old items are only sampled between trials 9 to 24 in the *background + presentation order* condition, whereas sampling is constant throughout trials in the background condition. Hence, sampling of negative old response times in *background + presentation order* excludes the generally longer response times found in the first trials of the test phase. Turning back to the top portion of Figure 22, it can be seen that response times for negative old items in the *background* condition is 1168 ms vs. 1005 ms in *the background + presentation order* condition for a difference of 163 ms. Our present analysis by trials suggests that response times for negative old items are lower in the *background + presentation order* precisely because of this unequal sampling. When the sampling is random, as was done in the *background* condition, this confound is eliminated and the difference between negative match and negative old items is reduced. However, if this hypothesis is correct, response times for negative match items in the *background + presentation order* should be among the fastest because they are collected last in the test phase. However, this was not the case as response times for these items were much slower than those sampled in the trial immediately before. Hence, unequal sampling provides only part of the explanation. The recognition test, presented later, will complete the picture.

Another possible explanation that could account for the presence of a statistically significant negative match effect in Allen and Brooks' (1991) Experiment 1 and the absence of this effect in the present experiment. It relates to the specific way the data was analyzed. Contrary to our experiment in which only data obtained with the rule paradigm were analyzed, Allen and Brooks included data from both the *rule and the* *induction paradigm* in their analyses. Specifically, they conducted 2 x (2) ANOVAs, which were ran separately for positive and negative items, with one within-subjects factor: phase (old vs. match), and one between-subjects factor: paradigm (induction vs. rule). Allen and Brooks reported an interaction for the paradigm factor (see Figure 5) with negative items. It showed that the difference in response times between negative match and negative old items obtained with the rule paradigm was larger than the difference obtained in the induction paradigm. However, they did not report if the difference between negative items within the rule paradigm was statistically significant and this is the comparison that was made in the present experiment. Because we did not conduct an induction task parallel to the rule task in the experiment, it is not possible to analyze the data as Allen and Brooks did.

Twenty blocks.

The error rate data obtained after 20 blocks of training is presented in the top portion of Figure 24. The interaction between phase and value was once again significant, $\underline{F}$ (1, 58) = 33.145, $\underline{MS_e}$ = 1.51, $\underline{p} < 0.001$, and there was still no difference between conditions, $\underline{F}$ (1, 58) = 0.341, $\underline{MS_e}$ = .016, $\underline{p} < 0.561$. Error rates for negative match items were 34% compared with 3% for negative old items and they were close to nil for positive match and positive old items (1% vs. 2% respectively).

The negative match effect is unequivocally present in response times when 20 blocks of practice are given. These data are shown in the bottom portion of Figure 24.

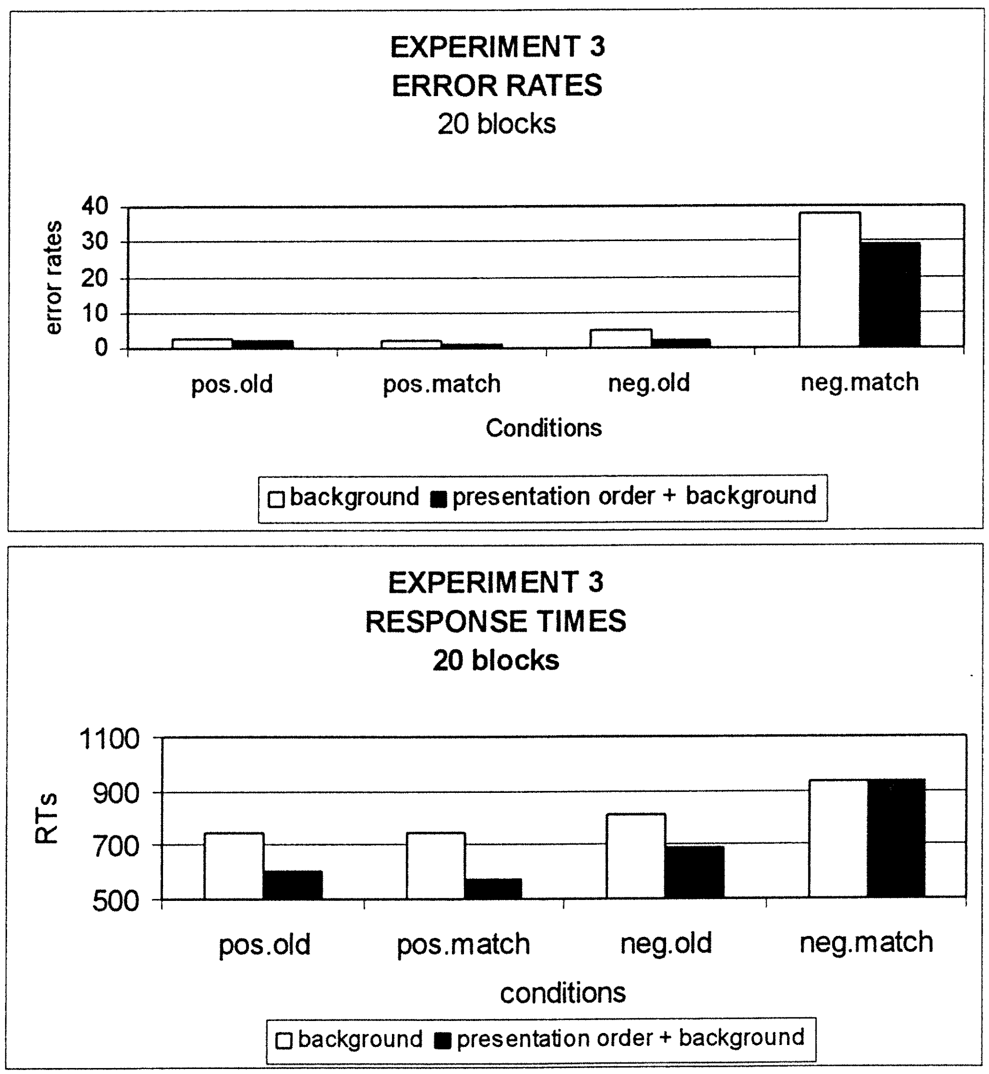


Figure 24. Error rates and response times at twenty blocks in Experiment 3.

The interaction involving, condition, phase and value was not significant, $\underline{F}$ (1, 45) = 1.7, $\underline{MS_e}$ = 39639, $\underline{p}$ < .2, but the phase and value was, $\underline{F}$ (1, 45) = 12.896, $\underline{MS_e}$ = 511169, $\underline{p}$ < .001. A decomposition of this last interaction revealed that the difference between negative match (934 ms) and negative old items (743 ms) was significant, $\underline{F}$ (2, 45) = 5.9, $\underline{MS_e}$ = 80587, $\underline{p}$ < .005. Hence, the diminished variance in response times given by an increased number of practice trials emphasized the tendencies observed after five blocks of practice. Once more, the negative match effect appeared stronger when backgrounds and the presentation order were used. The difference between negative match and negative old items in the *background + presentation order* was 249 ms vs. only 126 ms in the *background* condition. Hence, in response times as in error rates, using backgrounds and drawing the participants' attention on them, plays a large role in generating negative match effects. The non-randomized presentation order also appears to have contributed to the effect.

Finally, as in Experiment 1 and 2A, a main effect for phase was observed, $\underline{F}$ (1,45) = 7.4, $\underline{MS_e}$ = 48398, $\underline{p}$ < .009. Response times for old items were 706 ms vs. 793 for match items.

Memory test

The percentages of correct responses for the various item types are given in Table 16. Once more, global performance was poor. Participants in the *background* condition had a success rate of 51% vs. 53% for participants in the *background + presentation order* condition. Participants were somewhat superior in discriminating old items (75%) in contrast with match items (40%) and new items (47%). This result is different from the one found in Experiments 1 and 2A. Because the results were similar in both conditions, the explanation for the facilitated recognition of old items must be related to the inclusion of colored backgrounds in the training phase. Indeed, each of the four backgrounds was associated with only two different training stimuli. This creates better cues for recognizing old items because certain color and diagnostic attribute combinations are unique. For example, if we take item number one in Appendix G, it

Table 16

Experiment 3 memory test results after 20 blocks of training for both the background + presentation order and the background conditions.

Conditions		Responses: Old	Responses: Other
Background + presentation order			
	Item type		
	Old	76%	24%
	Match	60%	40%
	New	52%	48%
Background			
	Item type		
	Old	74%	26%
	Match	61%	39%
	New	54%	46%

can be seen that it is the only creature with a cane-shaped tail, stripes and a green background. Contrary to the other diagnostic and non-diagnostic attributes combinations, these associations made the old stimuli more distinctive and allowed the participants better recognition. The distinctiveness of old items created by the combination of background and certain diagnostic attribute may also help to understand unexplained data in the classification task.

To account for a larger difference between negative match and negative old items in the *presentation order + backgrounds* condition than in the *background* condition, a response time by trials analysis was conducted. It was argued that not randomizing the stimuli in the *presentation order + backgrounds* condition led to quicker response times for negative old items which consequently increased the difference between negative match and negative old items. This explanation depended on the argument that performances were improving across trials even though no feedback was provided. However, the important increase in response times for negative match items, found at the very end of the test phase was not explained. The memory test results lead us to believe that the stimulus distinctiveness created by the backgrounds combined with the repetition of old items in the match phase created this response time increase for negative match items.

Because the four negative old items, and then the positive old items, were presented repeatedly, this may have consolidated the association between the background and the rule attributes, which made the stimuli distinctive. Hence, in the test phase for negative match items, when the back patterns were changed, participants might have consciously realized that a given background and rule attribute combination should be associated with one category. Yet, the three rule attributes taken alone leads to categorize the item in the opposite category. Thus, this conscious dilemma could be responsible for the response times for negative match items in the *background + presentation order* condition. Although this argumentation is speculative, it does provide a plausible explanation of the phenomenon.

Discussion

As suspected, the backgrounds used in Allen and Brooks' original study were mostly responsible for the increased error rates and response times that led the authors to believe that exemplar memory was influencing the application of the rule. It was shown that the addition of the backgrounds along with instruction to pay attention to them in both the training and transfer phases of the experiment boosted error rates from between 5 and 10% as found in Regehr and Brooks' (1993) Experiment 2A, and in our Experiments 1 and 2A, to approximately 33% after 5 and 20 blocks of training. This error rate for negative match items falls in between the 45% reported in Allen and Brooks' (1991) Experiment 1) and the 18 to 28% reported in their Experiment 3. Hence, it was not the influence of a memory trace of learning phase exemplars that included the two non-diagnostic attributes that led participants to make a large number of classification errors for negative match items. Rather, it was the attention that they were giving to the non-diagnostic background that led them astray.

For response times, the addition of backgrounds was once again the main cause of the negative match effect observed after 20 blocks. It seems that the repetition of old items in the test phase might have also contributed in increasing the effect for participants in the *presentation order* + *background* condition.

The remaining question is the following. What do these results mean in the larger context of understanding the relationship between similarity-based and rule-based mechanisms of categorization? First, it seems clear that Brooks and his colleagues' version of a multiple-system model of categorization is unfounded. In agreement with other single and multiple models of categorization, and with the implicit learning literature, there was little learning of non-diagnostic attributes when the participants' attention was not brought to bear directly upon them. There was some evidence that a perceptually based memory trace for these attributes developed. Indeed, in the experiments presented, there was a systematic response time advantage for old items. However, this advantage was unrelated with the presence or absence of a negative match effect. This is true because the advantage for old items was found in the present experiment in which there were increased error rates and response time latencies for the negative match items. But, the advantage for old items was not found in the *standard* condition of Experiment 1, nor in Experiment 2A; and it was either absent or weak in the *names* condition of Experiment 1. Hence, it appears that perceptual representations of the stimuli that include both diagnostic and non-diagnostic information are stored during learning. However, contrary to the belief of Brooks and his colleagues, these representations do not interfere with the application of a well-known, practiced categorization rule. They merely cause slightly faster response times for old items in the match phases.

Nonetheless, the present experiments' elegant design still does bring evidence to support a multiple-system model of categorization. Indeed, if participants were behaving in a strictly analytical way, then even with the backgrounds, there would have been no increased error rates for negative match items. Hence, even if the task is not very realistic (or ecological, this is somewhat ironic considering Allen and Brooks initial goals), the experiment shows that forcing participants to attend to non-diagnostic information will adversely influence their categorical decisions even if they master the rule.

However, there remains one problem with our interpretation that denies the role of unattended non-diagnostic attributes in the categorization process. Three experiments presented in Regehr and Brooks (1993: Experiments 1D, 3A and 3B, see the Introduction) provide the arguments. Regehr and Brooks' Experiment 1D was identical in design to the standard condition of our Experiment 1. There were no backgrounds, nor was there a special presentation order. However, the attributes that composed the creatures were idiosyncratic. That is, if the logical value of the attribute for body shape stated "curved body", then each particular training exemplar with a curved body had its own unique curved body. In the experiments presented until now, there was always only one way to physically implement each logical value of the attributes. For instance, all creatures having a curved body had the same curved body. With idiosyncratic attributes, Regehr and Brooks succeeded in obtaining negative match effects for error rates and response times (see Table 5). The results for error rates (9% for old items vs. 33% for negative match items) are particularly strong and warrant explanation. Furthermore, Regehr and Brooks' Experiments 3A and 3B replicated these results and showed that modifying the non-diagnostic attributes in the match phase of the experiment eliminated the negative match effect for error rates and response times. The authors argued that this was strong evidence that the non-diagnostic attributes, and not only the attributes comprising the rule, were responsible for the negative match effect. This clearly contradicts the conclusion drawn from the present series of experiments.

Experiment 4 confirms the validity of our conclusions by showing that Regehr and Brooks' Experiment 1D, 3A, and 3B results were due to the idiosyncratic nature of the rule attributes.

Experiment 4

Regehr and Brooks (1993) realized that there were two possible interpretations for their Experiment 1D. First, consistent with the hypothesis formulated in Allen and Brooks (1991), the negative match effect could be due to a conflict between exemplar memory that included non-diagnostic information and the application of the categorization rule. However, it is also possible that the effect be due to the idiosyncratic nature of the attributes singled out by the rule. Indeed, an inspection of the stimuli used (see panel D in Figure 6) shows that match items are not only maximally similar to their corresponding old items because they share four out of five logical attributes, but also because they have the same idiosyncratic attributes. Hence, only one old item and its match item will share exactly the same body type, the same legs, the same neck, etc. Therefore, it is possible to obtain a conflict when an old item and its corresponding match item share some idiosyncratic rule attributes (e.g. number of legs and body type) but that the rule says they belong in different categories. For instance, seeing a creature with a particular body or leg, one might say "*it's in that category*" while the rule might be saying "*No, it's in the other category*". The important difference between the stimuli used in Allen and Brooks (1991, Experiment 1) and in Regehr and Brooks (1993, Experiment 2A) is that the diagnostic attributes, which receive attention every time the rule is applied, can cause a conflict due to their idiosyncratic nature. This was not the case in the other experiments. Nevertheless, if this is the correct explanation for the effect, it does not require postulating a powerful similarity-based mechanism that learns non-diagnostic information without attention and that uses this information to override conscious rule application.

To support the hypothesis that non-diagnostic attributes play an important role in the generation of the negative match effect, Regehr and Brooks designed two follow-up experiments. Experiment 3A was a replication of Experiment 1D. Experiment 3B was also similar except that the non-diagnostic attributes were modified in the transfer phase. Regehr and Brooks found a negative match effect in the first experiment only (see Table 6). Clearly, this is consistent with their hypothesis.

However, there were problems with their methodology. Most importantly, Regehr and Brooks used different training sets in Experiments 3A and 3B and did not control for the possible differences in learning between the two. Hence, the difference between Experiments 3A and 3B may have been unrelated to Brooks and his colleagues' hypothesis, stemming instead from dissimilar training experience. This shortcoming is not trivial, especially from Regehr and Brooks' perspective. Changing non-diagnostic attributes produces creatures with vastly different global looks. Regehr and Brooks supposed that the individuality of the creatures played an important role in generating the conflict between old and match items, but creature salience, attractiveness, and ease of memorization were not controlled when using different sets of training exemplars. This makes the results very difficult to evaluate adequately.

For example, consider the top stimuli for training sets A and B in Figure 7. A salient non-diagnostic feature such as the "looped neck" of the top creature in training set B might cause participants to adopt a conscious exemplar-specific strategy of categorization such as: "if it has a looped neck, then it is in category X" more than the very small neck of the creature in training set A. Because the saliency of the non-diagnostic attributes is not balanced, there is no control for the possible differences in strategy use for participants in the two training sets.

The existence of this potential methodological problem is supported by the fact that the differences in median response times obtained in Regehr and Brooks' (1993) Experiments 3A and 3B came not from increased response times for negative match items (1300 ms for 3A vs. 1271 ms for 3B), but rather from a substantial increase in response times for old items (811 ms vs. 1194 ms). Hence, in Experiment 3B, it appears that the effect was lost not because the negative match items showed greater latencies, but because old items did. This result is not the one expected from Brooks and his colleagues' multiple-system view of categorization.

This methodological problem could have been avoided by adding a second transfer set with the same non-diagnostic attributes as training set B (see Figure 25) and by including both training and transfer sets in a single analysis of variance. This procedure would have controlled for possible training phase differences and would have yielded an unequivocal answer to Regehr and Brooks' experimental question. Finally, notice that the comprehension of the results are further obscured by Regehr and Brooks' use of an analysis which merged all old items into a single cell. This analysis compared the results obtained with negative match items to those obtained with all old items, both positive and negative. As discussed in the Introduction, finding a difference between old items and negative match items might be due to the fact that positive old items yielded faster average response times (or fewer errors) which, in turn, diminished the average response times (or error rates) for old items, thus yielding the effect. However, this cannot be used to provide clear support for Regehr and Brooks' hypothesis, which expects a difference between negative match and negative old items not present among positive items. That is why the design, which includes positive and negative, and old and match items was used in all our previous experiments.

Experiment 4 replicated Regehr and Brooks' (1993) Experiments 3A and 3B. However, a second set of transfer items made of the diagnostic attributes of transfer set A and the non-diagnostic attributes of training set B was added. In short, there were two learning sets and two transfer sets, which were crossed over different groups of participants. This allows to separate effects due to learning from those due to the transfer conditions. Also, the results obtained with positive old and negative old item were merged in some analyses, following Regehr and Brooks' procedure, and they were kept separate in other more conservative analyses in line with those used in Experiments 1, 2A and 3 of this dissertation. Finally, match effects were tested after 5 blocks and 20 blocks of training. The results show that after five blocks, there is no evidence of negative match effects when the legitimate 2 (positive vs. negative) X 2 (old vs. match) test is used. At 20 blocks, even the more conservative analyses revealed a significant negative match effect, even in conditions in which the non-diagnostic attributes had been changed. This leads to the conclusion that the negative match effect was simply due to the idiosyncratic nature of the diagnostic attributes.

Method

Participants.

Sixty-four students at the Université de Montréal were randomly assigned to one of four groups. Each participant received 3$ as compensation for their time.

Materials.

The stimuli in the training sets A and B, and the transfer set that we will call A) were identical to those presented in Regehr and Brooks' (1993) Experiments 3A and 3B. The stimuli in the second transfer set (called B) were created by assembling the diagnostic attributes of the transfer set A with the non-diagnostic attributes of training set B. Thus, this created four experimental conditions with two training sets, A and B, and two transfer sets, also A and B (see Figure 25). The addition of transfer set B provides a test of Regehr and Brooks' hypothesis that a creature's individuality is an important factor in obtaining a conflict between exemplar memory and rule application while controlling for the potential differences in the training sets. Indeed, each training

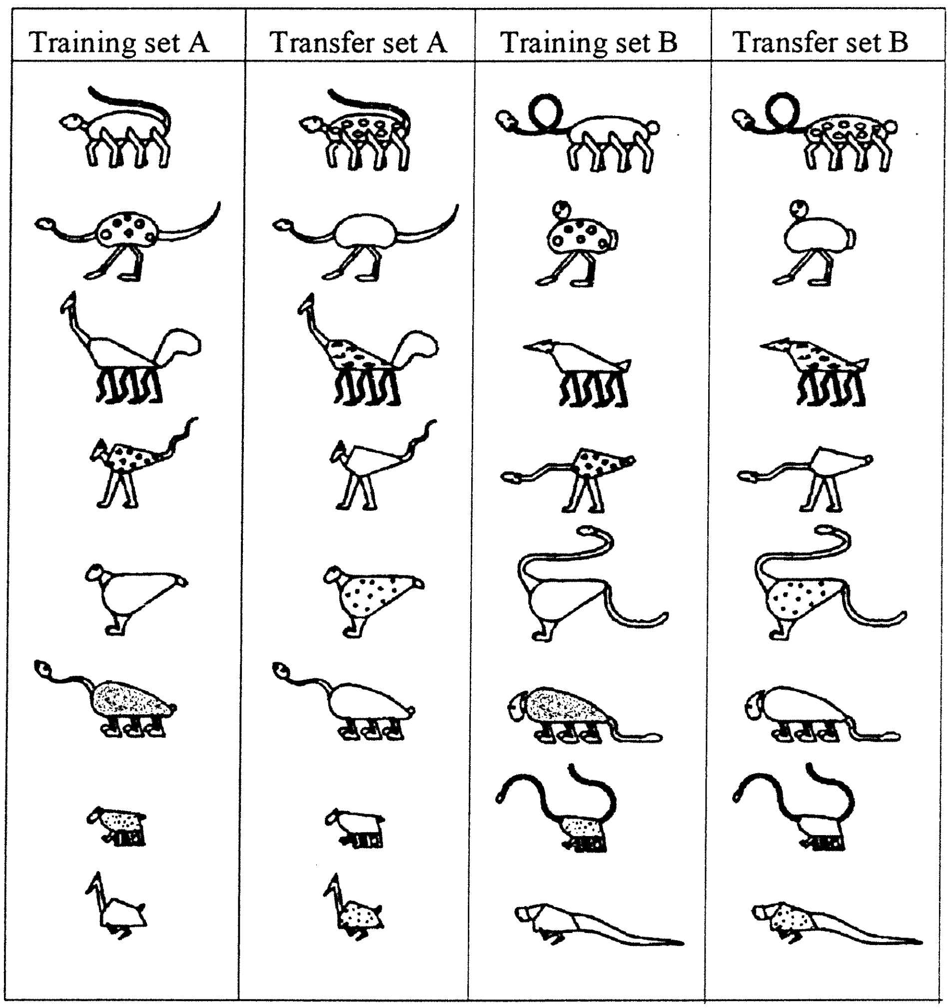


Figure 25. Training and transfer stimuli for Experiment 4. Note- The stimuli from the training sets A and B, and from the transfer set A are from G. Regehr, & L. R. Brooks, 1993, *Journal of experimental psychology: General, 122*, p. 94. Copyright by the American Psychological Association.

set is compared with a transfer set in which the non-diagnostic attributes remain constant (training set A → transfer set A, and training set B → transfer set B) and with a transfer set in which the non-diagnostic attributes are changed (training set A → transfer set B, and training set B → transfer set A).

The stimuli were drawings of fictional animals built from five binary attributes: number of legs (six or two), body type (angular or curved), spots (present or absent), neck length (long or short), and tail length (long or short). The logical description of the categorical structure is the same as the one used in Experiment 1 (see Table 8). Notice that all attributes are idiosyncratic. That is, each attribute value is implemented in a unique way for each training exemplar. For instance, in training set A, the attribute value "long neck" is shared by four creatures. Yet, each long neck is different. This is true for all other attributes values.

The first three attributes specified category membership. Once more, if an animal had two or three of these diagnostic attributes, then it was classified as a "Maurice". Otherwise, it was classified as a "Henri". As in Regehr & Brooks (1993, p. 107), four different rules were used to counter-balance the attribute assignment of the logical description. These rules were:

(1) Six legs, angular body, and spots present.
(2) Two legs, angular body, and spots present.
(3) Six legs, angular body, and spots absent.
(4) Two legs, angular body, and spots absent.

The last two attributes, neck length and tail length, were non-diagnostic because they both appeared equally often in each category. Their purpose was to maximize similarity between old and match items. As for all experiments in Regehr and Brooks (1993), no backgrounds were used.

Old items were selected from the training set A or B and match items were selected from transfer set A or B. In conditions in which the non-diagnostic attributes remained unchanged, the old and match items were perfectly matched on all attributes except spots, as was the case in all previous experiments using the rule paradigm. This created the four familiar item types: positive old, negative old, positive match, and negative match. However, in conditions in which the non-diagnostic attributes were modified, only the attributes "body type" and "number of legs" were identical for both old and corresponding match items.

For the explicit recognition memory test, items in Figure 25 unseen during the training and transfer phases were used as new items. For example, participants who received the training set A and transfer set A, were presented creatures from the training and transfer sets B as new items in the memory test. Hence, although the specific creatures used changed over conditions, there were always 16 new items.

Procedure.

The participants were tested individually. All instructions and stimuli were presented on 14" VGA monitors connected to 386dx or 486dx IBM compatible computers. The program MEL Professional v.2.01 (Schneider, 1989) was used to give the experimental instruction, present the material and record the participants' answers and response times.

Participants were assigned one of two training sets: A or B; and one of the two transfer sets: A or B. Participants in the four possible training and transfer set combinations were led through five identical experimental phases. These were identical to those used in Experiment 1 (excluding the phase in which exemplars without the non-diagnostic attributes were presented). To begin, the participants were given the categorization rule and instructed to classify the animals accordingly. They were given 40 trials divided in five blocks. However, unlike Experiment 1, the stimuli disappeared as soon as the participants' answers were recorded and so, the written feedback was displayed alone on a dark background.

In the transfer phase, the eight old items, the four positive match and four negative match items were presented once each in random order. This phase of the experiment proceeded like the training phase except that no feedback was given. At this point, Regehr and Brooks' (1993) Experiment 3A and 3B had been duplicated.

In the third phase of the experiment, the participants were given an additional 15 blocks (120 trials) of practice and then received another test phase identical to the first one. This provided an opportunity to test for a negative match effect after 20 practice blocks with feedback.

The last phase was an explicit recognition memory test. Participants were shown the eight old items, the eight match items and the sixteen new items in a random order. Their task was to determine whether each particular item had been seen in one of the first four phases of the experiment or whether the item was new. Contrary to the recognition test used in the previous experiments, participants had to discriminate between new items vs. old and match items. This was done to evaluate if the participants for whom the non-diagnostic attributes remained the same in the training and transfer phases could discriminate new items from other items with better accuracy than the participants for whom the non-diagnostic attributes were changed. Responses were given by selecting the appropriate key on the keyboard ("1" or "space bar"). No feedback was given concerning response accuracy.

Results

Classification task

Five blocks

The error rates and response times results were first analyzed following the method suggested by Regehr and Brooks (1993, Experiments 1D, 3A, and 3B), except that all training and transfer set combinations were included in the same design. The 2 X 2 X (3) ANOVAs included one within-subjects factor: item (old [both positive and negative] vs. positive match vs. negative match) and two between-subjects factors: training set assignment (A vs. B), and transfer condition (*same non-diagnostic attributes* in the training and transfer phases vs. *different non-diagnostic attributes* in the training and transfer phases). Concretely, this last between-subjects factor pitted participants having received training set A followed by transfer set A or training set B followed transfer set B against A followed by B and B followed by A. It provides a strong test to evaluate the role of the non-diagnostic attributes in the emergence of negative match effects. Error rates and response times were analyzed separately. Also, error trials were eliminated for response time analyses. This created an empty cell for one participant whose data were dropped from the analyses.

Table 17 shows the error rates and response times in the four training and transfer set combinations used. The main effect for item type was very significant for both error rates, $\underline{F}$ (2,120) = 14.7, $\underline{MS_e}$ = 0.3148, $\underline{p}$ < 0.001, and response times, $\underline{F}$ (2,120) = 14.9, $\underline{MS_e}$ = 4702276, $\underline{p}$ < 0.001. Mean error rates were 9% for old items,

Table 17

Experiment 4 error rate and response time results obtained after five blocks of training.

	Training set			
	A		B	
Transfer set	A	B	A	B
Measure				
Error rates				
Old	8%	9%	10%	9%
Positive match	11%	9%	17%	9%
Negative match	22%	20%	31%	16%
Negative match - old	14%	11%	20%	7%
Response times (ms)				
Old	2091	1539	2279	1882
Positive match	1908	1824	2412	2050
Negative match	2403	2091	3062	2281
Negative match – old	312	552	783	399

Note. Following the method suggested by Regehr and Brooks (1993) which involves comparing negative and positive match items to all old items combined.

12% for positive match items, and 22% for negative match items and the response times, in the same order were 1947ms, 2048ms, and 2459ms. Yet, the item type factor did not interact singly or jointly with the factors training set or transfer condition in either error rates (all $\underline{F} < 1.4$) or response times (all $\underline{F} < 1.5$). Hence, both measures showed a disadvantage for negative match items, which support the idea of a conflict between similarity-based and rule-based modules of categorization. However, there was no evidence to suggest that the effect was less pronounced when the non-diagnostic attributes were different compared to when they were the same after five blocks of training using Regehr and Brooks' analyses.

It was suggested that averaging all positive and negative old items into a single cell of the ANOVAs could potentially inflate the difference between old and negative match items. To explore this possibility, a second type of analysis was used to compare to data obtained after five blocks. The 2 X 2 X (2) X (2) ANOVAs included the same two between-subjects factors: training set (A vs. B) and transfer condition (non-diagnostic attributes the *same* vs. *different*). However, the within-subjects factors were those described in Experiment 1: phase (old vs. match) and value (positive vs. negative). As was explained previously, this test is more adequate to evaluate negative match effects because it compares the difference between negative old and negative match items to that between positive old and positive match items.

The error rate data after five training blocks did not reveal an interaction between phase and value, $\underline{F}\,(1, 60) = 0.381$, $\underline{MS_e} = 0.006$, $\underline{p} < 0.539$. The overall difference between negative match (22%) and negative old items (13%) was 9% and the difference between positive match (12%) and positive old (5%) was 7%. These factors were not involved in any interaction involving training set (both Fs < 0.4), but there was a very slight tendency for phase, value and transfer condition to interact, $F\ (1, 60) = 2.579$, $MS_e = 0.04$, $p < 0.114$. This data are presented in Figure 26. The means show that the participants for whom the non-diagnostic attributes were the same in the transfer phase were much further from showing a negative match effect (15% for negative old vs. 18% for negative match) than those for whom they were different (12% vs. 27%). This result

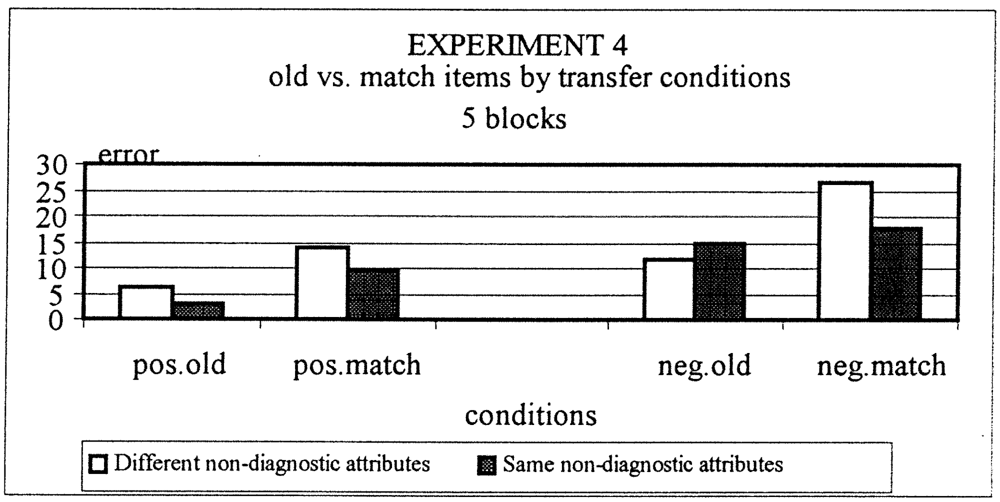


Figure 26. Error rates obtained after five blocks of training in Experiment 4 for positive and negative items, and old and match items by transfer condition.

is opposite to what would be expected from Regehr and Brooks' hypothesis concerning the role of non-diagnostic attributes.

The response time data did not support Brooks and his colleagues' hypothesis either. The phase and value factors did not interact with each other, $F\ (1, 59) = 2.159$, $MS_e = 445510$, $p < 0.147$, nor did they interact with the factors training set, transfer conditions or both (all $\underline{F}$s < 1.2). Globally, the difference between negative match (2426 ms) and negative old items (2005 ms) was 421ms and the difference between positive match (2034 ms) and positive old items (1781 ms) was 253ms. Thus, whether one uses the 2 (positive vs. negative) X 2 (old vs. match) analysis, similar to that used by Allen and Brooks (1991) or the 1 x 3 analysis (old vs. positive match vs. negative match) used by Regehr and Brooks (1993), the interpretation of the data changes greatly.

Yet, comparing the analyses suggested in Allen and Brooks (1991) and Regehr and Brooks (1993) lends strong support to the validity of the methodological remarks first raised in the Introduction. For both error rates and response times, the 2 X 2 design failed to yield a reliable negative match effect. However, the 1 x 3 analysis leads to conclude the exact opposite. Of the two types of analyses, the 2 X 2 design is clearly the more legitimate one. Indeed, the theory put forth by Brooks and his colleagues requires that there be a difference between negative match and negative old items and that this difference should be larger than the difference between positive match and positive old items. Only a 2 X 2 design can test for the presence of such an effect. By pooling together the results obtained with positive and negative old items, the 1 x 3 design provides biased comparisons with negative match items. As it turns out, the positive old items systematically produced the shortest response times. So, the fact that the negative match effect became significant in the 1 x 3 analysis shows that the effect is not due to a difference between negative match and negative old items, which was not reliable in the 2 X 2 analyses, but to the difference between negative match and positive old items.

Once again, notice that this last difference is totally irrelevant in the context of Brooks and his colleagues' theory.

Twenty blocks

The error rate and response time data obtained after 20 blocks of training is shown in Table 18. It was analyzed following the 2 X 2 X (2) X (2) ANOVAs design

Table 18

Experiment 4 error rate and response time data obtained after twenty blocks of training.

	Training set			
	A		B	
Transfer set	A	B	A	B
Measure				
Error rates				
Positive old	3%	1%	0%	0%
Positive match	8%	9%	8%	9%
Negative old	6%	5%	14%	5%
Negative match	34%	20%	30%	20%
Response times				
Positive old	1074	797	1094	986
Positive match	1458	1317	1471	1612
Negative old	1021	840	1151	944
Negative match	1905	1546	2110	2010

which included the within-subjects factors: training set and transfer condition, and the two between-subjects factors: phase and value. Eliminating error trials for the response time data created empty cells for three participants. Their data were dropped form the response times analyses.

For error rates, the interaction between phase and value was significant, $\underline{F}$ (1, 60) = 9.8, $\underline{MS_e}$ = .21, $\underline{p}$ < 0.003. A decomposition of this interaction revealed that the 20% difference between negative match (26%) and negative old items (6%) was also significant, $\underline{F}$ (4, 60) = 10.4, $\underline{MS_e}$ = .03, $\underline{p}$ < 0.001. However, phase by value did not interact with training set or transfer condition either singly or jointly, (all $\underline{F}$s < 1.4). Hence, the prolonged learning period produced a reliable negative match effect, unrelated to whether the non-diagnostic attributes were the same or different.

The same pattern of results was obtained for response times. The phase by value interaction was significant, $\underline{F}$ (1, 57) = 16, $\underline{MS_e}$ = 2769969, $\underline{p}$ < 0.0001, and the decomposition of this interaction showed that the 904 ms difference between negative match (1893 ms) and negative old items (989 ms) was also significant, $\underline{F}$ (1, 57) = 16, $\underline{MS_e}$ = 2769969, $\underline{p}$ < 0.0001. Once more, there were no significant interactions of this phase by value interaction with training set, transfer condition or both (all $\underline{F}$s < 1).

Therefore, the prolonged training period did provide evidence, using the truly appropriate analyses, that there could be a conflict between exemplar memory and rule application. Because these effects were identical whether the non-diagnostic attributes were different or the same, therefore the idiosyncratic rule attributes must necessarily have caused the conflict.

Memory test

Table 19

Experiment 4 memory test results after 20 blocks of training for all four training set and transfer condition groups.

	Responses	
Conditions	Old	Other
Same non-diagnostic attributes		
Training set A		
Item type		
Old	95%	5%
Match	83%	17%
New	6%	94%
Training set B		
Item type	92%	8%
Old	90%	10%
Match	13%	87%
New		
Different non-diagnostic attributes		
Training set A		
Item type		
Old	91%	9%
Match	57%	43%
New	43%	57%
Training set B		
Item type		
Old	89%	11%
Match	58%	42%
New	51%	49%

Table 19 presents the percentage of "new" vs. "other" responses given to old, match and new items. The data is shown for each of the training sets and transfer conditions. Remember that the task was to distinguish old and match items from new items. The global performance was 77%. Hence, the task was easier than in Experiments 1, 2A and 3. The success rate of participants for whom the non-diagnostic attributes were the same in the training and transfer phase was 91%. Although their success rates were slightly lower for match items compared to old items, it is clear that the idiosyncratic nature of the non-diagnostic attributes provided strong cues to distinguish new items from the other ones. Participants whose non-diagnostic attributes were different, had less success (63%) in distinguishing new items from the old and match items. As the participants in the other condition, they identified the majority of old items correctly (90%). However, their responding for match and new items was close to random. Hence, the participants could not say when the match items had been seen during the experiment. As for the new items, if their non-diagnostic attributes were the same as those for the old items, they were judged old. Otherwise, they were judged new.

Discussion

Experiment 4 showed that Regehr and Brooks' (1993) Experiments 1D, 3A and 3B were not problematic for the conclusions reached in Experiments 1, 2A and 3. It was suggested that, in the rule paradigm, participants may a form perceptual representation of the training stimuli that includes non-diagnostic information, but this memory does not override the conscious will of participants to apply a practiced rule. At face value, Regehr and Brooks' experiments that used idiosyncratic attributes and individualized stimuli challenged this view. They showed negative match effects without salient backgrounds or a specific presentation order and they also seemed to show that the non-diagnostic attributes were playing a crucial role in generating the effect. However, the present reproduction did not support these authors' interpretation. After five blocks of training, it was shown that there was no effect when the phase by value analyses were used. After twenty blocks, the effect was present, but it was generated whether or not the non-diagnostic attributes had been maintained. Hence, the conclusion reached for Experiments 1, 2 and 3 stands: non-diagnostic attributes, to which little attention is given during learning, does not influence rule application. Nonetheless, this last experiment does provide clear evidence that categorization is guided by a multiple-system process. Indeed, as in all the other rule paradigm experiments, a perfectly predictive rule was given to participants from the onset. If their behavior had been strictly rule governed, then there would have been no increased error rates or response time latencies for negative match items. Yet, after 20 blocks of training, negative match effects for both these variables were found whether or not the non-diagnostic attributes were the same. Because, altering the back pattern creates match items, only two attributes remain constant in all conditions: the body and the legs. The evidence suggests that these attributes were memorized and that it was this information that conflicted with rule application. In opposition with the prior experiments, this happen because: (1) attention was focused on these attributes during learning; (2) these attributes were idiosyncratic and systematically associated with the same category.

These data lead to another interesting question. In all the previous rule paradigm experiments, the attributes that were not mentioned in the rule did not carry any information with regards to category membership. So, there is no reason to give these attributes attention or to use them in categorizing the stimuli. Yet, what would happen with attributes that have cue validity with respect to category membership even though they are not mentioned in the rule. We explore this question in the following experiment.

Experiment 5

The capacity of characteristic attributes to influence concept acquisition and categorization has been shown in different ways that include typicality effects (Rosch and Mervis, 1975; Rips, Shoben, and Smith, 1973) and induction tasks (Posner and Keele, 1968). Another paradigm that has been used for this purpose is the simultaneous sorting task developed by Ahn and Medin (1992). It involves showing participants a set of exemplars and asking them to sort them into groups that seem natural. To test whether characteristic attributes are important in this type of task, Ahn and Medin presented participants with a critical set of exemplars that had a family resemblance structure (given in Table 20). Similarity-based approaches predict that participants will put exemplars 1 to 5 in one category and exemplars 6 to 10 in the other, because these grouping maximize inter-stimulus similarity. However, participants did not follow this pattern of behavior. Rather, they used a unique dimension to classify the items. For instance, on the basis of attribute A, they placed the exemplars 1 to 4 in one category. Then, they placed the exemplars 5 to 9 in the other category.

Table 20

Stimuli used in Ahn and Medin' (1992) simultaneous sorting experiment (Experiment 1, Set B).

	Category A					Category B			
Attributes	A	B	C	D		A	B	C	D
Exemplar					Exemplars				
1	0	0	0	0	6	2	2	2	2
2	0	0	0	1	7	2	2	2	0
3	0	0	2	0	8	2	2	1	2
4	0	1	0	0	9	2	0	2	2
5	2	0	0	0	10	1	2	2	2

Note. Categories A and B have family resemblance structures, because each exemplar has many attributes in common with the other members of its category. Yet, none of the attributes singly determine category membership.

Finally, exemplar 10 was put in the second category because it shared more attributes in common with members of that category. Ahn and Medin concluded that "...people do not assimilate probabilistic structures but rather organize them in terms of discrete structure plus noise (p. 81)". Hence, given the freedom to build their own categories, people made little use of family resemblances.

Ahn and Medin (1992) used these data to support their two-stage model of category construction. The idea is that people tend to adopt a one-dimension sorting or "1-D sorting" strategy. That is, people choose a salient dimension and divide the exemplars into two categories accordingly. Afterwards, if there are some exemplars that cannot be classified with the selected dichotomy (e.g. there is a medium object and the person is sorting objects as big or small), then these items are put in the category to which they show greatest similarity

At face value, this result is troublesome for similarity-based views. Indeed, if participants only use one salient dimension to classify objects and largely ignore other dimensions, then models resting upon the supposition that categorization depends on making judgment on a number of attributes become suspect. However, it might be argued that Ahn & Medin's (1992) simultaneous sorting task does not truly reflect the way people form concepts and categorize objects. First, we do not have simultaneous access to all the exemplars of two competing categories when we acquire concepts. For example, imagine having to learn the concept "prime number" by placing all the primes on one side of a table and all non-primes on the other, this would surely prove to be a very long process... This is not only true for mathematical concepts, but also for natural categories and artefacts. Rather, it is more plausible to believe that concepts are acquired by experiencing objects one at a time. For instance, suppose the first dog a child sees is a Golden Retriever. If one follows a standard similarity-based account of categorization, the child will commit different aspects of the dog to memory: color, size, shape, behavior, etc. Then, when the child meets other kinds of dogs, the "dog equals Golden Retriever" tentative concept is challenged and the concept is adapted to take in consideration these new experiences. Therefore, it is possible that multiple feature comparisons only occur when exemplars are shown one at a time. Another related problem with the simultaneous sorting task is that seeing all the exemplars at once might promote 1-D sorting because there is too much information to learn, analyze, compare and contrast at the same time. Hence, participants could be using a laboratory specific strategy which is: "in order to look consistent and to follow the experimenter's instruction, choose one dimension and classify all the items accordingly". Hence, it can be argued that the simultaneous sorting task does not provide an ecologically valid way by which to study categorization.

Regehr & Brooks (1995) pursued this idea and varied the sorting tasks used to evaluate the generality of Ahn & Medin's two-stage model. First, they showed that 1-D sorting in simultaneous sorting task is a robust phenomenon. They replicated Ahn & Medin' results with many different sets of stimuli. 1-D sorting was obtained even when the stimuli were "holistic blobs" thought to have no obvious features. However, a second task called "match-to-standard" did succeed in producing a larger amount of categorizations that used multiple features. Participants were shown the prototypes of two categories. Then, they were given a stack of cards containing other exemplars. They were told to categorize these exemplars using the two prototypes or "standards" in a way that seemed sensible and natural. Participants were not allowed to look through the stack in advance, nor were they allowed to see previously classified exemplars. Regehr & Brooks argued that this task was more realistic because it was comparable to categorizing a given object by comparing it to available instances, which are the “memorized standards”. The results obtained using five types of stimuli were the following: 53% of participants used family resemblances to classify the exemplars and only 21% followed the two-stage model. The remaining 25% used other strategies. Hence, the match-to-standards task does promote categorizations that involve multiple features. Yet, it can be observed that only half of the participants actually used family resemblances. This can hardly be said to be a general principal. Furthermore, when participants were asked to simultaneously sort exemplars that had just been classified in a match-to-standards task, most reverted to 1-D sorting. Thus, participants do not confer family resemblance attributes a special status, even when they have been used in previous categorizations.

Another criticism can be made however of most sorting tasks whether they involve simultaneous sorting or match-to-standards. In real life, categorization has a purpose. We form categories in order to understand the world so that our behavior may be better adapted to it. Hence, we are motivated to carefully observe creatures and objects, to seek similarities and differences, and most importantly, to create explanations that tell us why things are the way they are (Murphy & Medin, 1985; Carey, 1985). In sorting experiments, there is no such purpose and participants are allowed to sort the stimuli as they please. As Lassaline & Murphy (1996) note: "Presumably, family resemblance categories are initially more difficult to construct, because they require attending to and integrating information about values on multiple dimensions (p.96)". Hence, it is possible that people actually build categories that integrate multiple attributes in the real world, whereas in the laboratory, they simply pick a salient dimension and sort accordingly.

To verify this hypothesis, Lassaline & Murphy (1996) conducted a simultaneous sorting task with three groups. The first was given the standard task of sorting all the stimuli at the same time in a way that seemed natural. The second group answered induction questions about the stimuli before beginning the sorting task. For example, participants were asked questions such as "If a vehicle has bench seats, what kind of top does it have? (p.97)". Lassaline & Murphy suggested that this is akin to the real world categorization process because participants must focus their attention on all the attributes and seek causal explanations that underlie the relationship between the attributes. A third group answered frequency questions about the stimuli before the sorting task such as "How many vehicles have bench seats?" The frequency group served as a control for the induction group because it forced participants to look at each individual stimulus without making the inferences necessary to real world categorization. The results supported Lassaline & Murphy's hypothesis. 54% of participants in the induction group used family resemblances to sort the stimuli versus 17% for the frequency group and 19% for the control group. Hence, participants can use multiple features to categorize items even in a simultaneous sorting task. However, this paradigm also has a shortcoming. By going through the list of questions that were put to the participants, one has the impression that the experimenters were asking "leading question" to witnesses in a court case. Participants were lead to notice the relationships between attributes that make up the stimuli. In the end, one is left wondering if the participants sorted the items using family resemblances by thinking "it must be what the experimenter wants". Also, one might question if people are systematically that analytical in everyday categorization and if they would have noted all these relationships without the prompting.

Our review of the previous experiments leads to the following conclusions. First, the use of a unique dimension to classify exemplars in categorization tasks in which participants can make their own categories is a robust finding. It was the dominant strategy used in all experiments, which included a variety of stimulus types. Secondly, it is difficult to make participants deviate from this strategy. If the experimental conditions allowed participants to focus on individual items either by showing one stimulus at a time or by asking questions that forced careful analysis of individual stimuli, then more sorting that involved multiple features was observed. Otherwise, participants overwhelmingly used one dimension to classify the items.

However, one factor that remained constant in all of these experiments may explain the propensity for 1-D sorting. In all tasks, all the participants that were building

1-D categories were never given the opportunity to show that they had learned additional information concerning the other attributes of each item. This is the case because there was no transfer phase in these experiments where the participants were asked to categorize items on the basis of attributes other than the chosen salient dimension. Hence, it is possible that information about characteristic attributes is encoded and processed during categorization, but that these experiments do not provide a context in which this knowledge may be demonstrated. In this study, we propose an experiment to address this question.

A one-dimension rule sorting task

We designed a categorization task in which the stimuli belonging to one of two categories were presented one at a time. In the training phase, participants were shown creature-like stimuli similar to those in Experiments 1, 2A and 3. Participants were told to classify the creatures using a one-dimension rule, which perfectly determined category membership. In essence, they were instructed to do 1-D sorting. Each creature also included another attribute that was perfectly correlated (PC attribute) with the rule attribute and five "family resemblance" attributes (FR attribute) that were correlated 80% of the time with the rule attribute. However, these attributes were not mentioned in the experimental instructions. Thus, participants were not aware that the PC and FR attributes could potentially be used to classify the creatures.

In the transfer phase, the participants were shown test items that were highly similar to training items. However, the rule attribute was removed. The participants were instructed to classify the creatures in the same two categories as before, as well as they could. This procedure created a situation in which they were made to show any additional knowledge they had gained about the creatures, because only the attributes not mentioned in the training instructions remained.

Manipulating the PC and FR attributes produced two transfer conditions. First, in the *correlated attribute test phase (COR test phase)*, participants were shown test items without the rule attribute. This condition is a minimal test for the hypothesis that characteristic attributes can be learned in categorization tasks and subsequently be used because only the relationship between the rule attribute and the PC attribute needs to be learned in order to classify successfully all items in this transfer phase. Then, in the *family resemblances only test phase (FR only test phase)*, participants were shown test items that excluded both the rule attribute and the PC attribute. This is a stronger test of the role of characteristic attributes in categorization, because the correlation between each FR attributes and the rule attribute in the test phases is 66%. Therefore, using only one attribute correctly to classify the test items yields a score of 66% correct, using two attributes yields 93% correct and a minimum of three attributes is needed to correctly classify all test items. Thus, if it is supposed that all FR attributes are integrated in the representations of the categories during the training phase, then the success rates should be a 100% even in the FR only test phase. Finally, in both transfer phases, if characteristic attributes are not learned, then 1-D sorting based on the most salient dimensions should occur as predicted by Ahn & Medin's (1992) two-stage model.

Method

Participants.

Ninety-six students at the Université de Montréal participated in the study. Participants were randomly assigned to one of eight conditions. They received 3$ as compensation for their time if they had 40 training trials and 5$ if they had 160.

Materials.

The 10 training phase stimuli were drawings of fictional creatures built from seven binary attributes divided in two categories. The logical structure of these categories is given in Table 21. Each animal was composed of one "rule attribute" which determined category membership, one attribute that was perfectly correlated with the rule attribute and five family resemblance attributes that co-varied 80% of the time with the rule attribute. In the training phase, the rule attribute was given to the

Table 21.

Logical description of the stimuli used in Experiment 5.

	Category A							Category B						
Attributes														
	RULE	PC	FR1	FR2	FR3	FR4	FR5	RULE	PC	FR1	FR2	FR3	FR4	FR5
Training stimuli														
	0	0	0	0	0	0	1	1	1	1	1	1	1	0
	0	0	0	0	0	1	0	1	1	1	1	1	0	1
	0	0	0	0	1	0	0	1	1	1	1	0	1	1
	0	0	0	1	0	0	0	1	1	1	0	1	1	1
	0	0	1	0	0	0	0	1	1	0	1	1	1	1
COR and FR only test phase stimuli														
	~	0 or ~	0	0	0	0	1	~	1 or ~	1	1	1	1	0
	~	0 or ~	0	0	0	1	0	~	1 or ~	1	1	1	0	1
	~	0 or ~	0	0	1	0	0	~	1 or ~	1	1	0	1	1
	~	0 or ~	0	1	0	0	0	~	1 or ~	1	0	1	1	1
	~	0 or ~	1	0	0	0	0	~	1 or ~	0	1	1	1	1
	~	0 or ~	0	0	0	1	1	~	1 or ~	1	1	1	0	0
	~	0 or ~	0	0	1	1	0	~	1 or ~	1	1	0	0	1
	~	0 or ~	0	1	1	0	0	~	1 or ~	1	0	0	1	1
	~	0 or ~	1	1	0	0	0	~	1 or ~	0	0	1	1	1
	~	0 or ~	1	0	0	0	1	~	1 or ~	0	1	1	1	0
	~	0 or ~	0	0	1	0	1	~	1 or ~	1	1	0	1	0
	~	0 or ~	0	1	0	1	0	~	1 or ~	1	0	1	0	1
	~	0 or ~	1	0	1	0	0	~	1 or ~	0	1	0	1	1
	~	0 or ~	1	0	0	1	0	~	1 or ~	0	1	1	0	1
	~	0 or ~	0	1	0	0	1	~	1 or ~	1	0	1	1	0

Note. For training stimuli, the "RULE" attribute was given to participants as the classification rule. It perfectly determined category membership. The "PC" was perfectly correlated with the "RULE" attribute, but it was not mentioned in the rule. The "FR" or family resemblance attribute co-varied 80% of the time with the rule attribute. The "COR" test phase stimuli were created by removing the "RULE" attribute and the "FR only" test phase stimuli were created by removing both the "RULE" and "PC" attributes.

participants as the basis for classifying the creatures. The PC and FR attributes were not mentioned. In each category, all exemplars had six attributes in common with the prototypes ("1" or "0" for all attributes). However, the prototypes were not presented. Two examples of the stimuli are given in Figure 27.

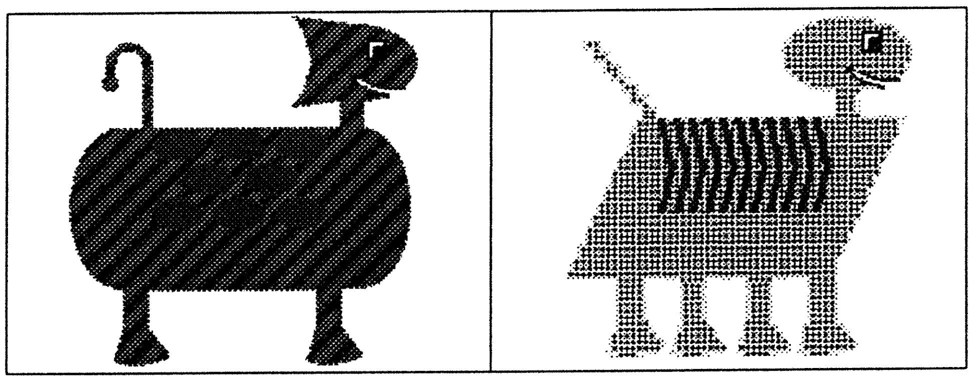


Figure 27. Black and white examples of the stimuli used in the training phase of Experiment 5. Note that the left-hand stimulus was green and the right-hand one was yellow.

Though all the stimuli were built using the same abstract categorical structure, two sets of exemplars were created to vary the type of rule attribute and PC attribute in the training phase. In the "global" condition, the rule attribute and the PC attributes were realized using dimensions that spanned the entire stimulus. They were color and texture. The remaining family resemblance attributes were body type, head, spots, legs and tail. In the "local" condition, the rule attribute and the PC attribute were realized using dimensions that were confined to a limited part of the stimulus. They were body type and legs. The family resemblance attributes were color, head, spots, texture, and tail.

Both of these conditions were separated in two depending on whether the rule attribute was *highly salient* (color and body type) or only *moderately salient* (texture and legs). Attribute saliency was determined by analyzing the post-experimental interviews of participants who had selected a 1-D sorting strategy in a similar experiment (Lacroix & Larochelle, Unpublished manuscript). This analysis revealed that body type and color were the two most popular choices, whereas texture and legs were only moderately popular. These manipulations yielded four experimental conditions: the global and highly salient rule attribute condition, the global and moderately salient rule attribute condition, the local and highly salient rule attribute condition, the local and moderately salient rule attribute condition. The conditions are displayed with the corresponding attributes in Table 22. Notice that manipulating rule attribute saliency did not require creating two extra stimulus sets, because the rule and PC attributes are perfectly correlated during training. Hence, the global training phase stimuli set can be used for both the highly salient and moderately salient conditions. The same is true for the local training phase stimuli set.

Table 22

The "RULE" and "PC" attributes used in Experiment 5.

			Attribute type	
			Global	Local
Attribute saliency	High	Rule attribute	Color	Body type
		PC attribute	Texture	Legs
	Moderate	Rule attribute	Texture	Legs
		PC attribute	Color	Body type

For the COR test phase, all attributes except the rule attributes were used to build the stimuli. Thus, four stimuli sets corresponding to the four experimental conditions were created with the global or local, and highly salient or moderately salient rule attribute removed. In the FR only phase, both the rule and PC attributes were removed. This created two more stimuli sets: one for the participants in the global conditions and one for the participants in the local conditions.

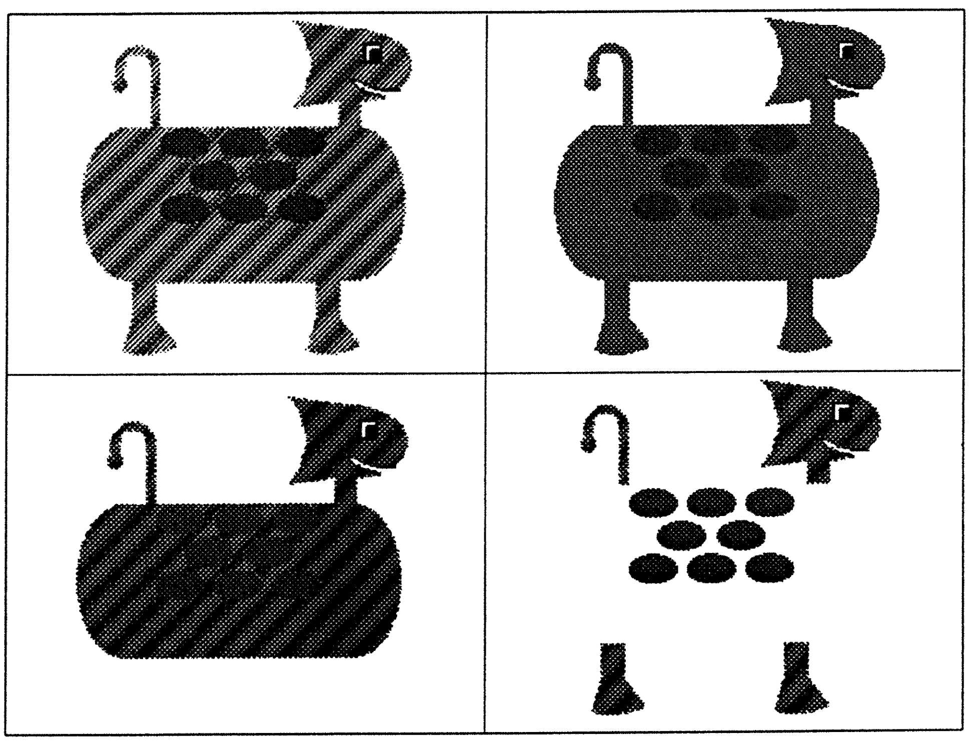


Figure 28. Black and white examples of the stimuli used in the "COR" test phase of Experiment 5. The top left-hand stimulus has color removed, the top right-hand stimulus has texture removed, the bottom left-hand stimulus has legs removed and the bottom right-hand stimulus has body removed. The "FR only" test phase stimuli were created in the same way except that either color and texture, or legs and body were both removed.

In all conditions, there were 30 transfer stimuli: 10 transfer exemplars had configurations of FR attributes that had been seen in the training phase and 20 configurations were new. All items appeared on a black background. Examples of the stimuli shown in the COR test phase are given in Figure 28.

Procedure.

The participants were tested individually. All instructions and stimuli were presented on 14" VGA monitors connected to 386dx or 486dx IBM compatible computers. The program MEL Professional v.2.01 (Schenider, 1989) was used to present the experimental instructions, material and record the participants' answers and response times.

Participants were randomly assigned to one of the four rule conditions and to one of two training length conditions: 4 blocks or 16 blocks. The experiment was conducted in three phases. In the training phase, the participants were told that their task was to familiarize themselves with the experimental setting. This directive was given in order to minimize overtly analytical behavior as was the case in Lassaline & Murphy's (1996) experiment. The goal was to verify how much information the participants would acquire about FR attributes while they were sorting the stimuli with a one-dimension rule. The participants were instructed to classify the creatures that would be presented following a simple rule. Participants in the global condition were directed to use color or texture depending on whether they had been assigned a highly or moderately salient rule. Participants in the local condition were directed to use the body or the legs. They were also asked to answer as quickly as possible without making any mistakes and to stay concentrated even if they found the task easy.

Each block was composed of the 10 training exemplars presented in random order. The number of learning trials (40 and 160) makes these training conditions comparable to those presented in Experiments 1, 2A, 3 and 4. All trials proceeded as follows. First, a fixation point appeared in the center of the screen for 1500 ms. An exemplar was then presented and participants had to classify the stimulus by selecting the appropriate letter on the keyboard. No feedback pertaining to accuracy was given. The inter-stimulus interval was 1000 ms.

In the first part of the transfer phase (the COR phase), the participants were told that they would see members of the two same families of creatures, but that the creatures would appear without the rule attributes (which was appropriately named in each of the conditions). Their task was to try to classify the creature in the same families as before. They were told to do their best, to respond spontaneously, and to guess if necessary. The 30 transfer items that were without the rule attribute were presented once each. Therefore, these items could be classified on the basis of the PC or FR attributes. The procedure was identical to that used in the training phase.

The last part of the experiment (the FR only phase) was identical to the COR phase except that the rule and the PC attributes were removed from the transfer items. Hence, their correct classification required knowledge of the FR attributes.

Finally, post-experimental interviews were conducted in which participants were asked how they had proceeded to classify the stimuli in the COR test phase and in the FR only test phase.

Results

Training phase

To evaluate potential differences between conditions due to the two sets of training stimuli that were used, separate analyses of variance were performed on the percentage of correct classifications and on response times obtained in this phase of the experiment. These analyses involved a 2 x 2 design with two between-subjects factors: feature type used as the rule (global vs. local) and training length (40 trials vs. 160 trials). All training trials were included, but incorrect responses were excluded from response time analyses.

Classification accuracy

The average percentage of correct classification was 98.1% with a standard deviation of 0.02%. The interaction between the factors was not significant, $\underline{F}$ (1,92) = 0.25, $\underline{MS_e}$ = .0001, $\underline{p} < 0.62$, nor was the main effect for feature type, $\underline{F}$ (1, 92) = 0.37, $\underline{MS_e}$ = .0002, $\underline{p} < 0.55$. However, the main effect of practice length showed a tendency towards significance, $\underline{F}$ (1,92) = 2.73, $\underline{MS_e}$ = .0015, $\underline{p} < 0.1$. Participants with 160 trials of practice (98,2%) were slightly more accurate than participants with 40 trials (97.9%).

Response times

The average response time was 488ms with a standard deviation of 298ms. There was a significant feature type by training length interaction, $\underline{F}$ (1, 92) = 3.95, $\underline{MS_e}$ = 342757, $\underline{p} < 0.05$. A decomposition of the interaction showed that global attributes showed faster response times after 160 trials of training (407ms) than after 40 trials (591ms), $\underline{F}$ (1, 94) = 4.8, $\underline{MS_e}$ = .4907958, $\underline{p} < 0.03$. However, local stimuli did not produce faster response times after 160 trials of practice (504ms) than after 40 trials of practice (450ms), $\underline{F}$ (1, 94) = 0.4, $\underline{MS_e}$ = .35812, $\underline{p} < 0.53$. Thus, both training sets yielded fairly similar results, although participants in the global conditions who had 160 trials had faster average response times.

Transfer phases

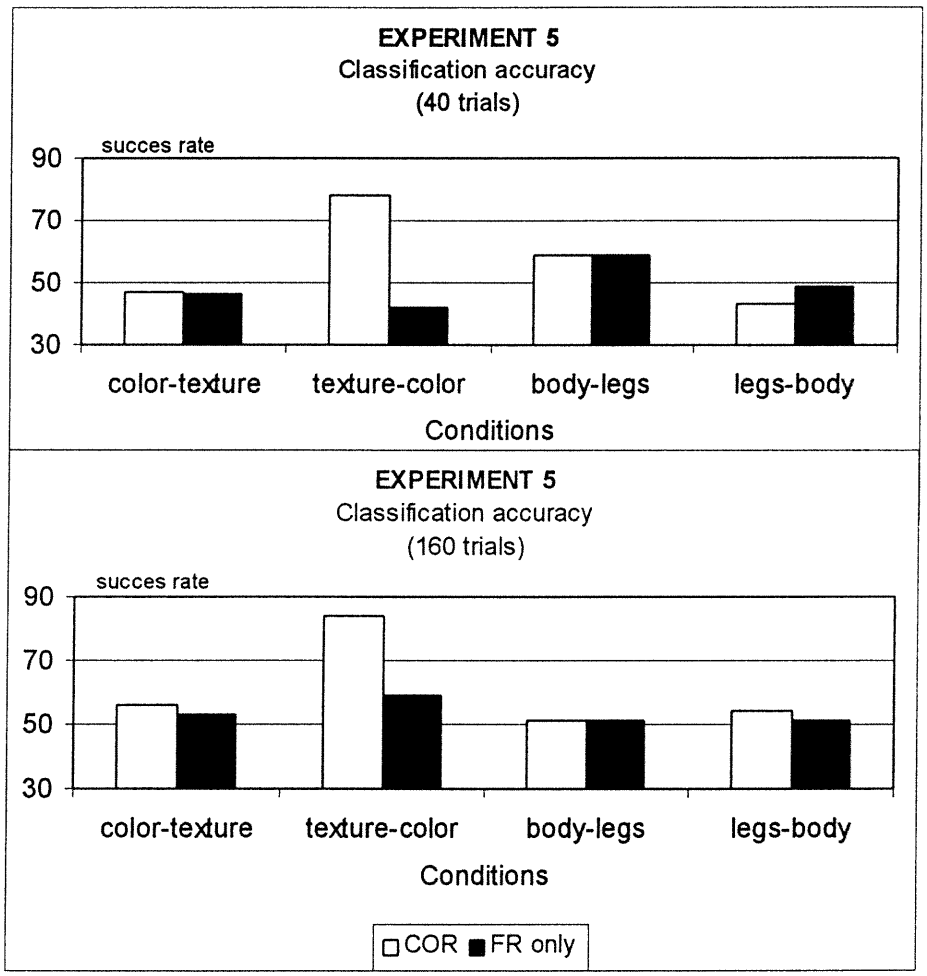


Figure 29. Classification accuracy obtained after 40 and 160 trials in the "COR" and "FR only" test phases of Experiment 5.

Separate analyses of variance were performed on the percentage of correct classifications and on the response times to assess performances in the transfer phases.

These analyses involved a 2 X 2 X 2 X (2) design with three between-subjects factors: training length (40 trials vs. 160 trials), feature type used in the rule (global vs. local), saliency of the rule attribute (high vs. moderate); and one within-subject factor: test phase (COR vs. FR only). The very large error rates made the response time data difficult to interpret. Therefore, they will not be reported.

Classification accuracy

Classification accuracy for all conditions is presented in Figure 29. The top portion shows the data obtained after 40 trials of practice and the bottom portion shows the data obtained after 160 trials. As can be observed, general performance was poor (55.1%). The interaction involving all four factors was not significant, F (1, 80) = 1.8, MS_e = 0.04, p < 0.18. However, there was a three-way interaction between feature type, saliency and test phase, F (1, 80) = 10.9, MS_e = 0.24, p < 0.01. The decomposition of this effect showed that the interaction between feature type and saliency was significant in the COR test phase, F (1, 80) = 16.6, MS_e = 0.68, p < 0.001, but that it was not in the FR only test phase, F (1, 80) = 0.96, MS_e = 0.02, p < 0.33. This was to be expected, as the features defining the type and saliency conditions are absent in the FR test phase. Further decomposition showed that when the feature type was global, there was a significant difference between the high saliency and moderate saliency rule attribute conditions, F (1, 80) = 21.1, MS_e = 0.87, p < 0.001, but that was not the case when the feature type was local, F (1, 80) = 1.2, MS_e = 0.05, p < 0.28. In other words, in the COR test phase, when the rule attribute was texture and the PC attribute was color, participants' classification accuracy was much greater (81%) than in the other conditions (52%). In the FR only test phase, participants were equally poor in classifying the stimuli (52%). The training length factor did not interact with any other factor, although a tendency was found for the interaction between training length and feature type, F (1, 80) = 3.2, MS_e = 0.13, $p < 0.08$. A decomposition of that interaction showed that participants with global feature rules showed higher classification accuracy rates (63%) than those with local feature rules (52%) after 160 trials of training, F (1, 80) = 7.3, MS_e = 0.28, $p < 0.09$, but this was not the case after only 40 trials, F (1, 80) = 0.1, MS_e = 0.01, $p < 0.71$ (53% for global feature rules vs. 52% for local feature rules).

The analysis of classification accuracy shows that participants did not generally learn much about the COR and FR attributes in the one-dimension classification rule task. The general accuracy rate was almost random (55%). The only exception was for participants who received texture as the classification rule along with color as the PC attribute. In this condition, success rates soared to 81%. This result appears to show that color was being used to successfully classify the stimuli. Yet, there is little support for the hypothesis that family resemblance attributes, or even attributes perfectly correlated with category membership are implicitly learned in a way to allow successful category decisions once the rule attribute is removed.

However, the averaged data does not allow us to evaluate if the participants were simply guessing, or if they were sorting using a 1-D rule as predicted by Ahn and Medin's (1992) two-stage model of category construction. Their model predicts that when people are free to construct their own categories, they will choose a salient dimension and sort accordingly. If this were the case, then we would expect each participant to have success rates of 0% or 100% if they are using a PC attribute, and 33% or 66% if they are using a FR attribute. Naturally, these numbers suppose that participants are being consistent without necessarily being correct with regards to category membership. This type of responding could explain our results. First, it would yield average success rates of approximately 50%, which is what was found. The exception would be for participants who had color as the PC attribute. In this condition, we would expect a high number of participants with a 100% success rate. To explore this hypothesis, we constructed a frequency distribution of classification accuracy by participant shown in Figures 30 and 31. This allowed us to determine how the participants were responding during the transfer phase and thus evaluate the two-stage model.

Classification accuracy frequency distributions

Figure 30 shows the data for participants whose rule attributes were global. In the top panel, the PC attribute is texture and in the bottom panel, the PC attribute is color. Figure 31 shows the data for the local attribute rule data. In the top panel, the PC attribute is legs and in the bottom panel, the PC attribute is body type. The white bars show the data for the COR test phase and the gray bars show the data for the FR only test phase. The data obtained after 40 trials and 160 trials were merged.

Participants were placed in one of seven slots corresponding to the following

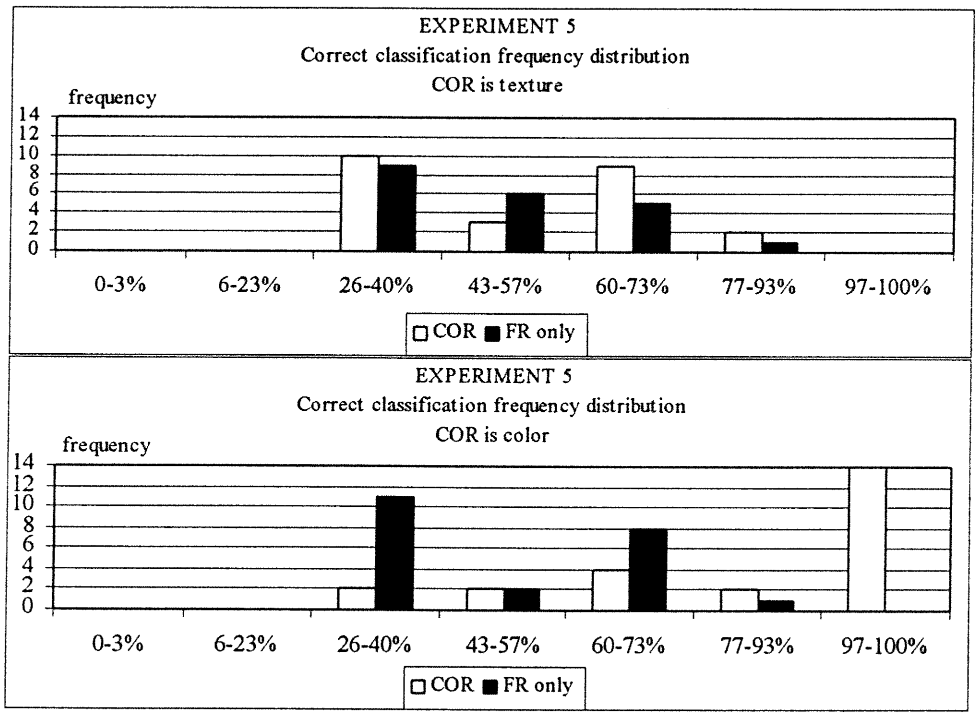


Figure 30. Correct classification frequency distributions for the "COR" and "FR only" phases of Experiment 5. Data obtained after 40 and 160 trials were merged. The top portion of the figure shows the data for the condition in which the COR attribute was texture and the bottom portion shows the data for the condition in which the COR attribute was color.

patterns of responding. First, participants who achieved scores of 43 to 57% inclusively were grouped together. We considered that these participants were responding randomly allowing one or two answers to deviate from perfect stochastic behavior. Secondly, participants who scored from 26 to 40% or 60 to 73% inclusively were grouped into two slots, which correspond to 1-D sorting behavior with one of the FR attributes (and allowing one or two inconsistent answers). The first 1-D sorting group (26 to 40%) was systematically classifying the animals in the wrong category while the second group (60 to 73%) was systematically correct. Thirdly, participants who obtained scores of 3 or 0%, and of 97 or 100% were placed in slots representing 1-D sorting using the PC attribute. The participants in the first of these groups were systematically wrong and those in the second were systematically right. Finally, participants who got scores of 6 to 23%, or 77 to 93% inclusively were put together in slots corresponding to multi-dimensional sorting.

Over all conditions, most participants answered randomly or using a one-dimension rule: 26% were random, 35% correctly used a 1-D attribute to classify the stimuli (COR and FR combined), and 32% incorrectly used a 1-D attribute. Hence, most participants were classifying the stimuli using a one-dimension rule if they were not guessing. As for the remaining participants, 5% showed scores consistent with the correct use of more than one attribute to classify the stimuli and 2% were consistent with an incorrect use of multiple attributes.

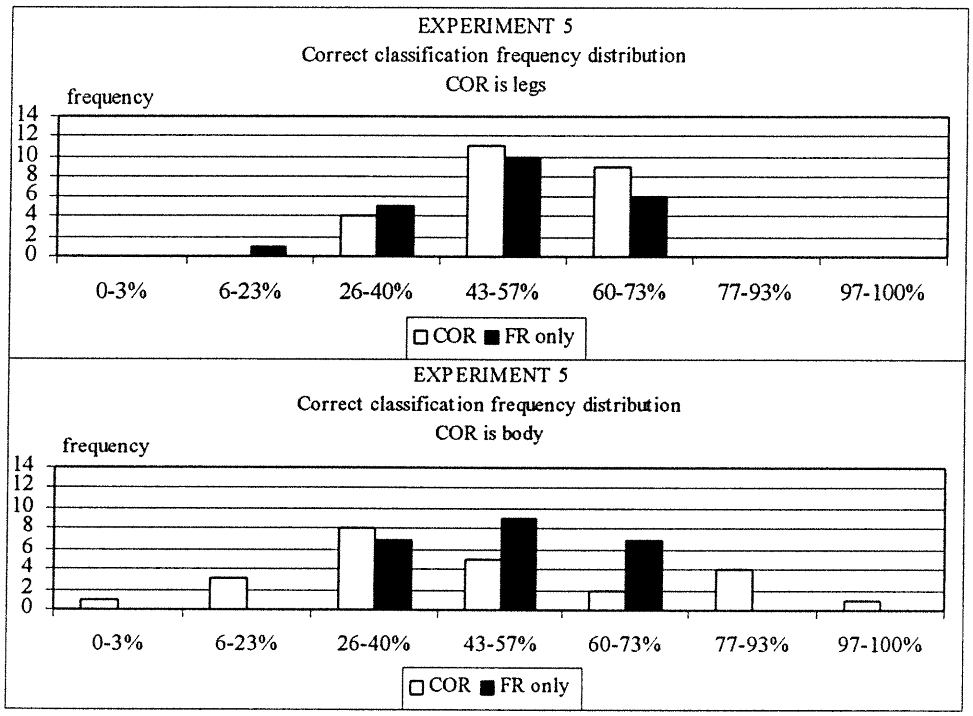


Figure 31. Correct classification frequency distributions for the "COR" and "FR only" phases of Experiment 5. Data obtained after 40 and 160 trials were merged. The top portion of the Figure shows the data for the condition in which the COR attribute was legs and the bottom portion shows the data for the condition in which the COR attribute was body.

Differences between the conditions were also observed. Most importantly, in the COR test phase, when the PC attribute was color, 14 participants achieved success rates of 97% or 100% vs. one participant for all three other conditions combined. This clearly shows that the 81% success rate in this condition was due to participants encoding the relationship between the colors and the category. The fact that color was perfectly correlated with the rule attribute is crucial to this result. In the local conditions, color was a FR attribute. Post-experimental interviews revealed that this attribute was chosen

22 times by the participants (COR and FR only test phases combined) to classify the stimuli. Yet, the accuracy rates were equal or above 60% only 45% of the time. Therefore, participants were successful in correctly associating a given color and the appropriate category only if this attribute was perfectly diagnostic.

A second difference may be found in participants' distribution between those with global rule attributes and those with local rule attributes. Merging data from the COR and FR only test phases, a chi-square test which compared the data for global and local conditions for the seven performance slots showed a significant difference, $\underline{X}^2$ (6, $\underline{N} = 184) = 28$, $\underline{p} < .0001$. In the global conditions, 29% of participants were using a FR attribute to classify test stimuli correctly, 35% were using a FR attribute incorrectly and 14% were categorizing randomly. In the local condition, 26% of participants were using a FR attribute to classify test stimuli correctly, 26% were using a FR attribute incorrectly and 38% were behaving randomly. Participants in local conditions were much more likely to classify the stimuli randomly than their counterparts in global conditions. However, the reliability of this effect might be disputed on the grounds that five cells with less than five observations were included in the chi-square. To eliminate this statistical violation, we reanalyzed both factors including only frequencies that fell in the random responding slot or in the 1-D sorting with FR attributes slots. This did not change the result, $\underline{X}^2$ (2, $\underline{N} = 154) = 10.4$, $\underline{p} < .005$. There is no obvious explanation for this result. It is not simply a matter of participants in the color as PC attribute condition falling less often in the random behavior slot because of the large number of participants that are 100% correct. Comparing the global condition in which texture was PC attribute to the both local conditions, it is observed that there are still much more participants of the local conditions that fall in the random behavior slot. A possible explanation is that participants in the local condition had a higher level of random responding because they were convinced they had not learned anything about the stimuli during training, because their attention was focused on a specific part of the animals. In comparison, participants in the global conditions got an overall view of the stimuli. Therefore, they might have had the impression of learning something about the FR attributes without it being necessarily the case.

The frequency distribution of classification success rates appears to support Ahn and Medin two-stage model of category construction. A majority of participants chose an attribute to sort the test stimuli in a consistent way. The near random success rates observed in the averaged data occurred because participants did not associate a given attribute value with the correct category in most cases.

Discussion

The goal of this experiment was to determine whether information about characteristic attributes was encoded in a one-dimension rule sorting task thereby allowing correct category decisions once the rule attribute is removed. With one exception, the answer was no. Participants were generally using a one-attribute rule, but they were as likely to classify the stimuli correctly as they were to do so incorrectly. Nonetheless, when texture was the rule attribute and color was the PC attribute, more than half of the participants successfully classified the stimuli. The design of the experiment makes this result interesting. First, it was not due to color being a global attribute because participants did not learn texture when it was the PC attribute and color was the rule attribute. Secondly, it was not the fact that color was perfectly diagnostic of category membership alone that yielded the result. All other conditions had an equally diagnostic attribute in the COR test phase which could have been used to classify the stimuli. Yet, only one participant in those conditions chose the PC attribute to successfully sort all the stimuli. It was also shown that color was not associated with category membership when it was a FR attribute in the local conditions. In these conditions, post-experiment interviews showed that participants who chose color to classify the stimuli were as likely to put them in the right category as the were to put them in the wrong one. Hence, in the one-dimension rule sorting experiment, participants did not learn characteristic attributes in a way that allowed them to classify the stimuli except when color was perfectly correlated with category membership.

These results are consistent with those of our previous experiments: unattended attribute are not learned in a way to influence category decisions. This is true whether the attributes are non-diagnostic as in Experiments 1 and 2A or whether they are characteristic as in this experiment. As in Experiment 4, it was possible for participants to learn attributes only if their attention was focused upon them. In this experiment, when texture was the rule and color was the PC attribute, more participants succeeded in correctly classifying the stimuli shown in the COR test phase. Notice, however, that this condition forced participants to process incidentally color information each time they applied the rule, much as the idiosyncratic rule attributes were processed in Experiment 4. In the local conditions, the participants did not learn the PC and FR attributes in a way to help them classify transfer phase items because the application of the rule did not make them attend to these features. The longer training periods did not change this last fact simply because unattended attributes stayed unattended throughout the training phase. By comparison, in Experiment 4, the longer training period promoted exemplar-based effects because it gave the participants more time to learn the idiosyncratic rule attribute for each stimulus. Finally, the only data more difficult to explain within this framework are those that were obtained when texture was the PC attribute. In principle, we would have expected it to be learned in the same way as color was when it was the PC attribute. This did not occur. It must be speculated that recognizing the texture types promoted a process by which the colors were associated with the categories and that this information was stored in a way to allow for the correct classification of the COR test phase stimuli, but that the opposite relationship between these two attributes did not promote such a process. Explaining this result in more detail will require further studies.

The results also bring convergent validity to Ahn and Medin's (1992) two-stage model of category construction. The majority of participants did pick a single attribute in order to classify the stimuli. However, the conclusions reached in Experiments3 and 4 lead us to disagree with Ahn and Medin's claim that family resemblances do not influence categorization. In criticizing Brooks and his colleagues' multiple-system of categorization, we noted that similarity-based effects on rule application stem from the attributes to which attention is given. In simultaneous sorting tasks, participants focus on a single attribute. If this attribute has only two possible values, as was the case in our experiment as well as in Ahn and Medin's experiments, then there is no variability on that attribute between the exemplars in each category. Thus, similarity-based effects are not expected in this case. If it had been shown that attributes not included in the rule could be learned in a way to influence classification decisions, then simply having stimuli with a family resemblance structure would have generated the effects. Because this does not occur, similarity-based effects in these sorting task could only take place if the rule attribute had many possible values. For example, if the rule attribute were body shape (curved vs. straight, as in Experiment 4), then we might expect more typical curved and straight body types to influence classification accuracy and response times after a certain period of training. We conclude that Ahn and Medin did not find an influence of family resemblances on rule application because the logical structure of the stimuli, and the process by which exemplars are learned, did not afford them an opportunity to do so.

General discussion

The goal of this dissertation was to explore further the relationship between rule-based and similarity-based mechanisms of categorization. It was hypothesized that the application of a rule could provide a strong constraint on exemplar learning and on the influence of exemplar memory on categorical decisions. Our investigation and analysis of the results obtained using multiple variants of the rule paradigm provided evidence for both these hypotheses.

The standard condition of Experiment 1 and Experiment 2A provided extensive evidence that unattended non-diagnostic attributes are not learned in a way to generate a conflict between exemplar memory and rule application. These results clearly contradicted Allen and Brooks (1991, and Regehr and Brooks, 1993) hypothesized multiple-system model of categorization which proposed that: (1) exemplars are learned quickly, incidentally, and without regard to the diagnostic value of the features; (2) the memory trace of these exemplars is stored in way to slow down decision times and mislead people to wrongly apply a perfectly predictive and practiced categorization rule.

First, our replication of the rule paradigm entirely failed to generate negative match effects whether 40 trials or 160 trials of practice were used. Second, the item analyses, which were shown to be sensitive to exemplar-based effects in Experiment 2B, also failed to reveal negative match effects. Third, the memory test showed that participants could not explicitly discriminate old from match items. Because it is necessary to have memorized the non-diagnostic attributes to successfully complete this task, this test contributed further evidence that Brooks and his colleagues' proposed multiple-system view was unwarranted. Instead, when the non-diagnostic attributes were excluded, the item analyses of Experiment 2A showed a strong relationship between the response times obtained for training phase exemplar and the corresponding transfer phase items. Thus, the response times to apply the rule were directly related to the time taken to test the individual rule attributes and determine category membership.

The naming condition of Experiment 1, Experiment 2B and Experiment 3 showed that non-diagnostic attributes influence category decisions when participants focus some attention on them. First, Experiment 3 showed where the original Allen and Brooks experiment had gone wrong. The use of salient backgrounds as a sixth non-diagnostic attribute and the instructions to pay attention to these backgrounds in both the training and transfer phases of the experiment combined to created the important negative match effects for error rates and response times. These two factors were essential in eliciting the effects as Experiment 1 and 2A showed that otherwise, they did not occur. Second, the naming condition of Experiment 1 and Experiment 2B showed that directives that lead some of the participants to focus on the non-diagnostic attributes also create negative match effects. For participants in the naming condition of Experiment 1, the results showed a small, but reliable negative match effect after a prolonged period of training. This occurred because a number of participants used name learning strategies that involved non-diagnostic attributes, as there were no constraints as to the attribute that could be used to accomplish this task. Hence, the attention given to the non-diagnostic attributes in relation with the naming directives was sufficient to generate the negative match effect observed for error rates. A similar explanation may be used to account for the induction task results of Experiment 2B. Contrary to the rule paradigm, the induction task allows the participants to infer the rule using the attributes of their choice. It was shown that many rules involving non-diagnostic attributes could lead to perfect categorization of the training stimuli, while leading to disastrous performances of the transfer stimuli. Because a sufficient number of participants induced rules that included non-diagnostic attributes, the averaged data yielded clear negative match effects for error rates and response times. Yet, the rule-based explanation eliminated the need to postulate a multiple-system account as was done by Brooks and his colleagues.

Experiment 4 showed a clear interaction between rule application and exemplar memory. In this paradigm however, the effect was not due to the non-diagnostic features. Indeed, as in Experiment 1, these attributes were not targeted by the rule. Hence, this information about the exemplars was not stored in a way to create a conflict. Rather, it was the idiosyncratic nature of the rule attributes, which caused the effect. The systematic attention given to these attributes through rule application succeeded in creating representations, which could influence category decision for transfer items. Notice that in principle, the same effect could have taken place in Experiments 1 and

2A. However, in those experiments, the physical implementation of the stimuli perfectly reflected the underlying categorical structure. This made exemplar-based similarity effects stemming from the rule attributes difficult to obtain. Indeed, analysis of the exemplars, non-withstanding the non-diagnostic attributes, show that all members of both categories were equally distant from the prototypes except for the prototypes. Hence, these non-prototypical exemplars would not be expected to produce important similarity-based effects. As for the prototypes, one could have predicted lower error rates and response time latencies (see Posner and Keele, 1968). However, the rule paradigm confounded this particular effect with rule application, as only two tests of attributes were necessary to determine category membership for these items. Thus, attribute testing also predicted lower error rates and response time latencies. This is why the results of Experiments 1 and 2A did not help in specifying the relationship between rule-based and similarity-based mechanisms of categorization.

Using the one-dimension rule paradigm, Experiment 5 showed that characteristic attributes, including those perfectly correlated with category membership, were generally not more likely to be learned through rule application than non-diagnostic attributes. In the local conditions, characteristic attributes did not lead to the correct classification of the transfer attributes because they had received little attention during training. In the global conditions, only participants who classified stimuli for which color was perfectly correlated with category membership succeeded in correctly categorizing transfer stimuli once the rule attribute was removed. This result was found to be consistent with those of the previous experiments, because participants were incidentally attending to color by classifying training stimuli using texture type. Otherwise, participants performed as randomly as those in the local conditions.

Globally, the results eliminate the need to postulate a multiple-system of categorization with a powerful similarity-based component as proposed by Brooks and his colleagues. Not only was this view opposed to other models of categorizations such as the GCM, which includes a parameter for attribute saliency or diagnostic value (see Medin and Schaffer, 1978; Nosofsky, 1986), it also went against findings in the implicit learning literature (Goschke, 1997). This is not to say however that absolutely no information about attributes excluded from the rule is gained in analytically driven categorization once attention-based explanations have been accounted for. For instance, it was shown in all the rule paradigm experiments that response times for training phase stimuli were faster than those of the transfer stimuli. This indicated that participants had acquired a perceptually based representation of the training stimuli, which facilitated responding in the transfer phases. Also, in Experiment 4, it was shown that participants for whom the non-diagnostic attributes were maintained in the training and transfer phases of the experiment could easily discriminate these attributes from the new ones in the memory test. Only, this information is not stored in a way to make participants deviate from correct category decisions. Nonetheless, we believe these results support a multiple-system view.

First, on its own, a rule-based model cannot account for the data. In experiments such as the rule paradigm, it would predict near flawless performances because the categorization rule is known. Yet, the naming condition of Experiment 1, and Experiments 3 and 4 showed that exemplar-based information could influence category decisions. This data add to the large quantity of research showing that categorization is not strictly analytical.

However, we do not believe that a similarity-based approach would be sufficient either. The GCM has been very successful in explaining exemplar-based effects (Nosofsky, 1984; 1986; 1988). In a recent article, Nosofsky and Johansen (2000) have also showed that the newest implementation of the GCM could account for the variability in strategies used by participants in induction tasks. Thus, it is quite possible that the GCM could predict most of the data obtained in our experiments. Although this single mechanism explanation would have the merit of being parsimonious, we believe that our data showed that a complete model of categorization must include a rule-based component. Central to this claim is the following observation, all exemplar-based effects observed were obtained through the application of a categorization rule, whether given by the experimenters or self-made. Participants used rules to infer and categorize exemplars. In doing so, they incidentally acquired knowledge concerning the variability between exemplars. As expertise for given categories developed, this implicit knowledge about the exemplars influenced the classification process. In sum, rule application drove exemplar learning. However, the participants' behavior showed that this explanation, though simpler, is incorrect. Thus, a one mechanism exemplar-based model, such as the GCM, could be sufficient to account for the data.

We believe viewing concepts as the result of a dual-system process can provide an explanation for two important problems on categorization. First, the fact that people often show analytical behavior in classifying objects although classical definitions for concepts are not available. Second, the fact that similarity-based views explain much categorization data, but that in the absence of constraints on what information is processed in similarity calculations, they simply cannot account for the meaning and coherence of everyday concepts. If concept acquisition is viewed as a rule-driven process, then these rules provide the constraint on similarity-based processes, as only a small number of attributes are typically included in the rules. Moreover, if exemplar memory is incidentally acquired and implicitly influences categorization, then this information will be part of the concept while being difficult to verbalize. Thus, this dual mechanism view of categorization expects that concepts will be rule-based and yet, hard to define. This characterization seems to succinctly describe what the human concept has been found to be.

References

Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, Massachusetts: Havard University Press.

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

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120, 3-19.*

Ashby, G. F., Alfonso-Reese, L. A., Turken, U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological review. Vol. 105, No. 3,* 442-481.

Ashby, G. F., & Maddox, T. (1993). Relations between prototypes, exemplar, and decision bound models of categorization. *Journal of mathematical psychology, 37*, 372-400.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211-227.

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

Bruner, J., Goodnow, J., & Austin, G. (1956). *A Study of Thinking*. New York: Wiley.

Capaldi, E. J., & Neath, I. (1995). Remembering and forgetting as context discrimination. *Learning and Memory, 2*, 107-132, Carey, S. (1985). *Conceptual change in children.* Cambridge, MA: MIT Press.

Carlson, R. A. & Dulany, D. E. (1985). Conscious attention and abstraction in concept learning. *Journal of experimental psychology: learning, memory and cognition, 11*, 45-58.

Cleeremans, A., Destrebecqz, A. & Boyer, M. (1998). Implicit learning: news from the front. *Trends in cognitive sciences, Vol. 2, No. 10*, 406-416.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time form semantic memory. *Journal of verbal learning and verbal behavior, 8,* 432-438.

Coombs, C. H., Dawes, R. M, & Tversky, A. (1970). Mathematical Psychology: an elementary introduction. Englewood Cliffs, N. Y.: Prentice-Hall.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of experimental psychology: General, Vol. 127, No. 2*, 107-140.

Fodor, J. A. (1998). *Concepts where cognitive science went wrong.* Clarendon Press: Oxford.

Fodor, J. A. (1975). *The language of thought.* New York: Crowell.

Fodor, J. A., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition, 28*, 3-71.

Goldstone, R. L. & Barsalou, L. W. (1998). Reuniting perception and cognition. In S. A. Sloman & L. J. Rips (Eds.), *Similarity and symbols in human thinking* (pp 145-176). Cambridge, MA: MIT Press.

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition, 52*, 125-157.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman. *Problems and Projects* (pp . 437-447). Indianapolis: Bobbs-Merrill.

Gopnik, A. (1988). Conceptual and semantic development as theory change. *Mind and language, 3*, 197-216.

Goschke, T. (1997). Implicit learning and unconscious knowledge: mental representation, computational mechanisms, and brain structures. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories*. (pp 247-334). Cambridge, MA: MIT Press.

Hahn, U. & Chater, N. (1996). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories*. (pp 247-334). Cambridge, MA: MIT Press.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

Hitzmann. D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological review, 93,* 411-428.

Hume, D. (1985). *Of miracles.* La Salle, Illinois: Open Court publishing company. (Original work published in 1748).

Katz, J. (1972). *Semantic Theory.* New York: Harper and Row.

Keil, F. C. (1989). *Concepts, kinds, and cognitive development.* Cambridge, MA: MIT Press.

Kersten, A. W., Goldstone, R. L., & Schaffert, A. (1998). Two competing attentional mechanisms in category learning. *Journal of experimental psychology: learning, memory and cognition, Vol.24, No. 6*, 1437-1458.

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychological bulletin, Vol. 112*, No. 3, 500-526.

Kruschke, J. K. (In press). Toward a unified model of attention in associative learning. *Journal of mathematical psychology*, 1-28.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, Vol. 99, No. 1*, 22-44.

Kuhn, Thomas (1962). *The structure of scientific revolutions, second edition, enlarged*. Chicago: The University of Chicago Press.

Lacroix, Guy L., & Larochelle, S. (2001). Learning family resemblance attributes in a one-dimension rule classification task. Unpublished manuscript.

Lassaline, M. E., & Murphy, G. L. (1996). Induction and category coherence. *Psychonomic bulletin & review. 3 (1)*, 95-99.

Laurence, S., & Margolis, E. (1999). Concepts and Cognitive Science. In S. Laurence & E. Margolis (Eds.). *Concepts Core Readings*. 3-81. The MIT Press: Cambridge, Massachusetts.

Lewicki, P., Hill, T. & Bizot, E. (1988). Unconscious acquisition of complex procedural knowledge. *Cognitive psychology. 20*, 24-37.

Lewicki, P. (1986). Processing information about covarations that cannot be articulated. *Journal of experimental psychology: learning, memory and cognition*, 12, 135-146.

Luce. R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103-189). New York: Wiley.

Macnamara, J. (1999). *Through the rearview mirror historical reflections on psychology*. The MIT Press: Cambridge, Massachusetts.

Martin, R. C. & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: evidence for common processing strategies. *Journal of experimental psychology: general. Vol. 109, No. 3*, 320-353.

Mayr, U. (1996). Spatial attention and implicit sequence learning: evidence for independent learning of spatial and non-spatial sequences. *Journal of experimental psychology: learning, memory and cognition*, 22, 350-364.

Medin, D. L. (1989). Concepts and conceptual structure. *American psychologist, 44*, 1469-1481.

Medin, D. L., & Schaffer, M. M. (1978). 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, Vol. 10, No. 1*, 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of experimental psychology: General. Vol. 115, No. 1*, 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of experimental psychology: learning, memory and cognition, 14*, 700-708.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of experimental psychology: learning, memory and cognition, Vol. 14, No. 1*, 54-65.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review, 7(3)*, 375-402.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 211-233.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review. Vol. 101, No. 1*, 53-79.

Posner, M., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology, 77*, 353-363.

Quine, W. (1960). Word and Object. Cambridge, MA: MIT Press.

Ratcliff, R. and McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*, 385-408.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of experimental psychology: General, 118*, 219-235.

Reber, A. S. (1976). Implicit learning of synthetic languages: the role of instructional set. *Journal of experimental psychology: human learning and memory, 2*, 88-94.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior, 6*, 855-863.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect test: determining what is learned about sequence structure. *Journal of experimental psychology: learning, memory and cognition, 20,* 585-594.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive psychology, 3,* 382-407.

Regehr, G. & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: the priority of holistic individuation. *Journal of experimental psychology: general, Vol. 122, No. 1,* 92-114.

Regehr, G. & Brooks, L. R. (1995). Category Organization in free classification: the organizating effect of an array of stimuli. *Journal of experimental psychology: learning, memory and cognition, Vol. 21, No. 2,* 347-363.

Rips, L. J. (1989). Similarity, Typicality, and Categorization. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21-59). Cambridge: Cambridge University Press.

Rips, L. J., Shoben, E. J., and Smith, E. E. (1973). Semantic Distance and the

Verification of Semantic Relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1-20.

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

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika, 22,* 325-345.

Schneider, W. (1989). Micro experimental laboratory: an integrated system for IBM PC compatibles. *Behavior research methods, instruments, & computers. Vol 20(2),* 206-217.

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

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of experimental psychology: learning, memory and cognition, Vol. 26, No. 1,* 3-27.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: the early epochs of category learning. *Journal of experimental psychology: learning, memory and cognition, Vol. 24, No. 6,* 1411-1436.

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. In S. A. Sloman & L. J. Rips (Eds.), *Similarity and symbols in human thinking* (pp 145-176). Cambridge, MA: MIT Press.

Smith, E. E., & Sloman, S. A. (1994). Similarity- versus rule-based categorization. *Memory & Cognition, 22* (4), 377-386.

Smith, L. B. & Samuelson, L. K. (1998). Perceiving and remembering: category stability, variability and development. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories.* (pp 161-196). Cambridge, MA: MIT Press.

Solomon, K. O., Medin, D. L., & Lynch, E. (1999). Concepts do more than categorize. *Trends in Cognitive Sciences, Vol. 3, No. 3,* 99-105.

Stadler, M. A. (1995). Role of attention in implicit learning. *Journal of experimental psychology: learning, memory and cognition, 15*, 1061-1069.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford University Press.

Wagner, J. & Allen, S. W. (1995). Analogies and rules in children's category learning. unpublished manuscript.

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

Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of experimental psychology: learning, memory and cognition, 15*, 1047-1060.

Wittgenstein, L. (1953). *Philosophical Investigations*. (G. E. M. Anscombe, Trans.). Oxford: Blackwell.

Appendix A

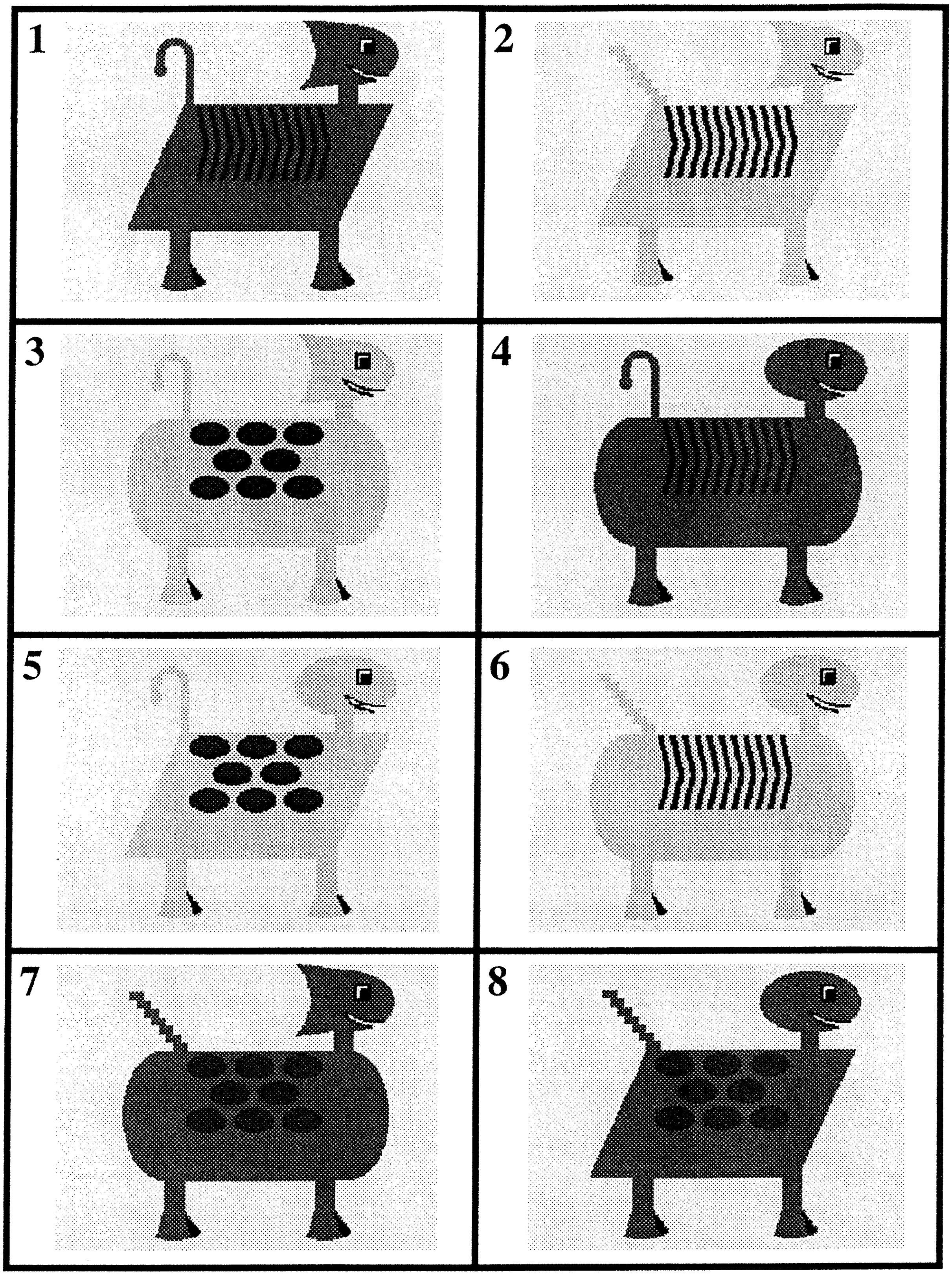
Numéro de participant:_________ **Version:___________**

Veuillez inscrire le prénom de chaque créature dans la case correspondante.

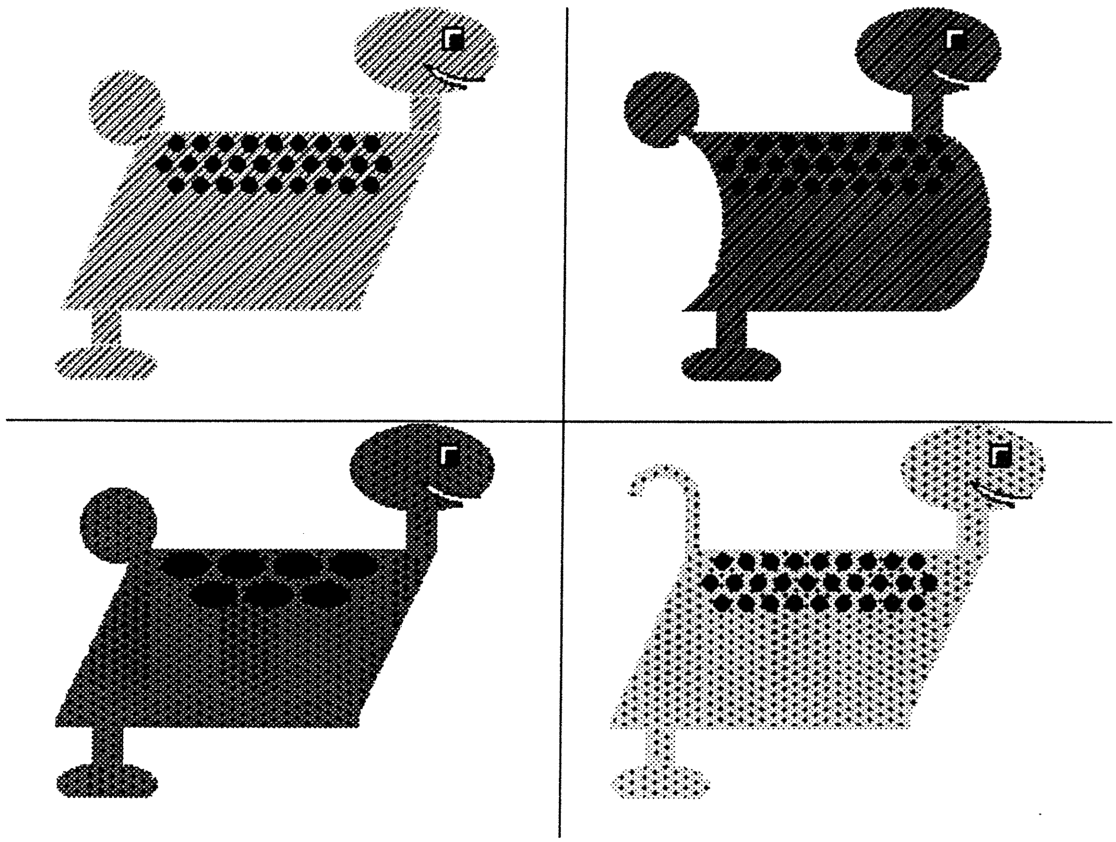
1	**2**
3	**4**
5	**6**
7	**8**

Appendix B

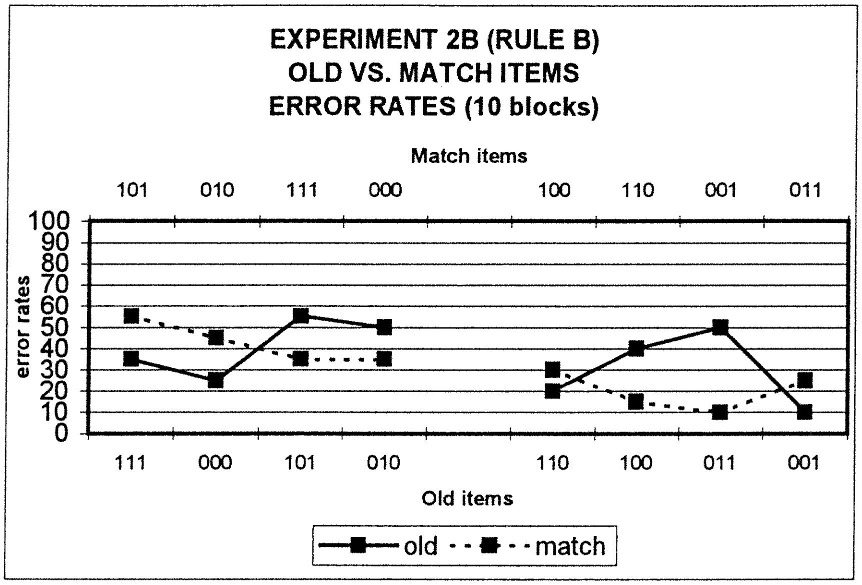
VERSION A

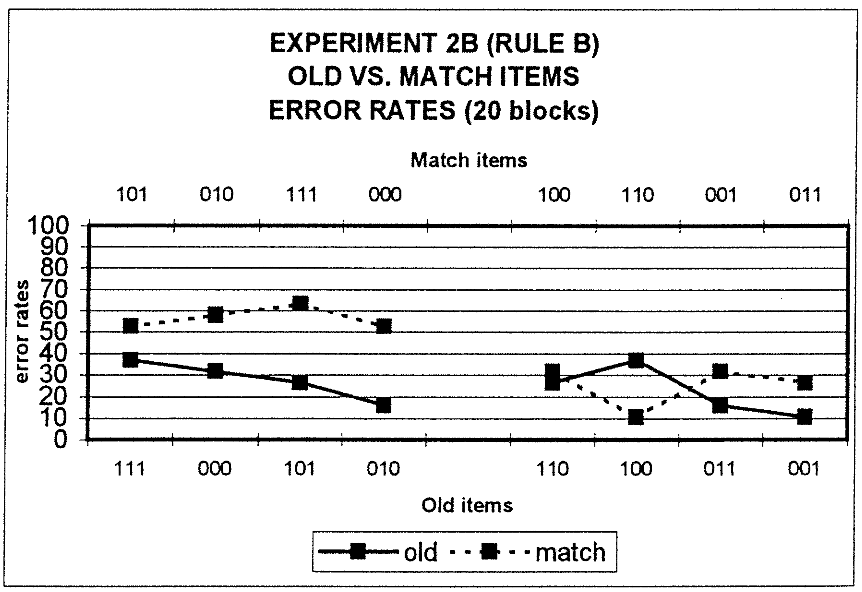


Appendix C



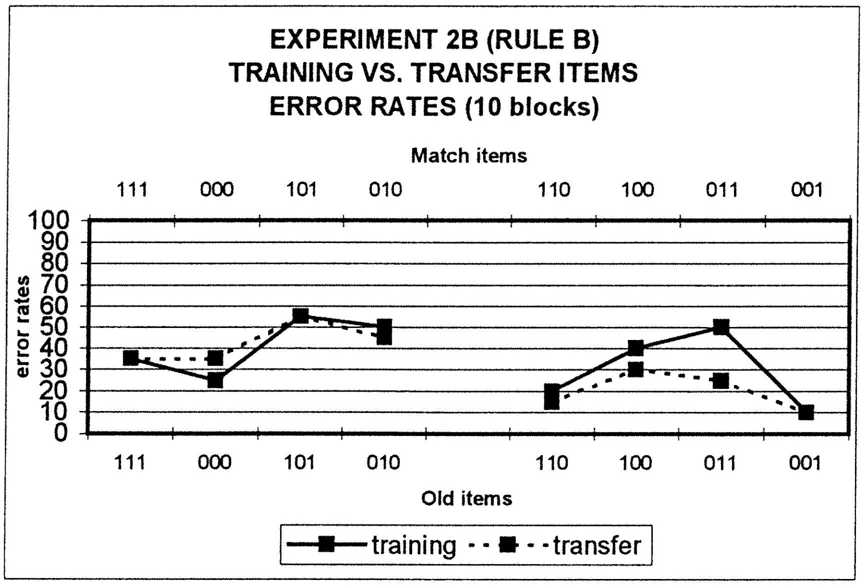
Appendix D

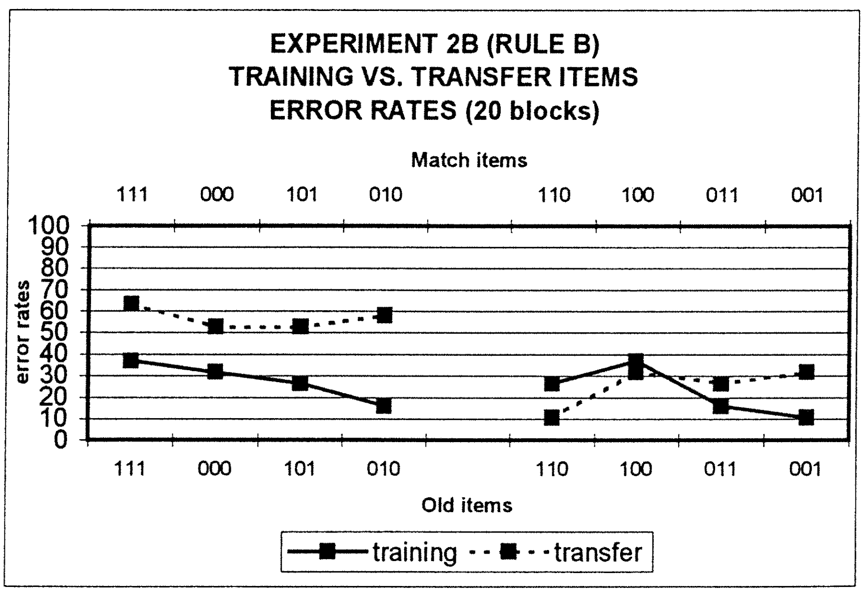




Note. The positive items are on the right-hand side of the figure and the negative items are on the left-hand side.

Appendix E





Note. The positive items are on the right-hand side of the figure and the negative items are on the left-hand side.

Appendix F

	Item type			
	pos.old	neg.old	pos.mat	neg.mat
participants				
10 blocks				
201	100%	100%	0%	50%
117	50%	100%	50%	50%
356	50%	100%	50%	25%
126	75%	50%	0%	50%
198	75%	50%	25%	25%
202	75%	50%	75%	50%
350	50%	75%	25%	50%
373	50%	75%	0%	50%
105	25%	75%	75%	75%
114	50%	50%	75%	25%
115	50%	50%	75%	25%
122	25%	75%	50%	75%
199	25%	75%	100%	25%
200	25%	75%	0%	50%
360	50%	50%	25%	0%
361	50%	50%	50%	75%
362	50%	50%	50%	50%
376	25%	75%	0%	50%
113	25%	50%	50%	0%
351	25%	50%	0%	25%
353	50%	25%	0%	50%
355	50%	25%	25%	0%
363	25%	50%	50%	75%
382	50%	25%	50%	0%
107	0%	50%	75%	0%
108	0%	50%	0%	75%
123	0%	50%	0%	75%
125	0%	50%	0%	50%
133	25%	25%	75%	25%
366	25%	25%	0%	50%
369	0%	50%	25%	25%
379	25%	25%	0%	75%
381	25%	25%	25%	75%
127	25%	0%	0%	50%
211	0%	25%	75%	50%
368	0%	25%	0%	50%
384	0%	25%	25%	50%
106	0%	0%	50%	25%
372	0%	0%	0%	0%
375	0%	0%	0%	75%

	Item type			
	pos.old	neg.old	pos.mat	neg.mat
participants				
20 blocks				
370	50%	75%	0%	50%
104	25%	75%	25%	50%
112	25%	75%	0%	75%
128	50%	50%	100%	50%
205	75%	25%	0%	75%
209	50%	50%	100%	50%
359	50%	50%	25%	50%
377	75%	25%	50%	25%
380	50%	50%	25%	50%
390	25%	75%	25%	50%
119	0%	75%	0%	50%
121	0%	75%	25%	50%
102	0%	50%	25%	100%
109	50%	0%	50%	0%
206	25%	25%	50%	25%
208	0%	50%	0%	50%
352	25%	25%	50%	100%
358	25%	25%	75%	25%
365	25%	25%	0%	50%
371	0%	50%	0%	50%
374	0%	50%	0%	50%
378	25%	25%	25%	25%
387	25%	25%	50%	75%
388	25%	25%	0%	75%
101	0%	25%	0%	50%
103	25%	0%	25%	100%
111	0%	25%	0%	100%
118	0%	25%	50%	50%
130	0%	25%	0%	75%
383	25%	0%	0%	75%
110	0%	0%	0%	100%
203	0%	0%	0%	100%
207	0%	0%	25%	100%
210	0%	0%	50%	25%
354	0%	0%	0%	100%
364	0%	0%	100%	0%
367	0%	0%	0%	100%
385	0%	0%	50%	25%
391	0%	0%	0%	100%

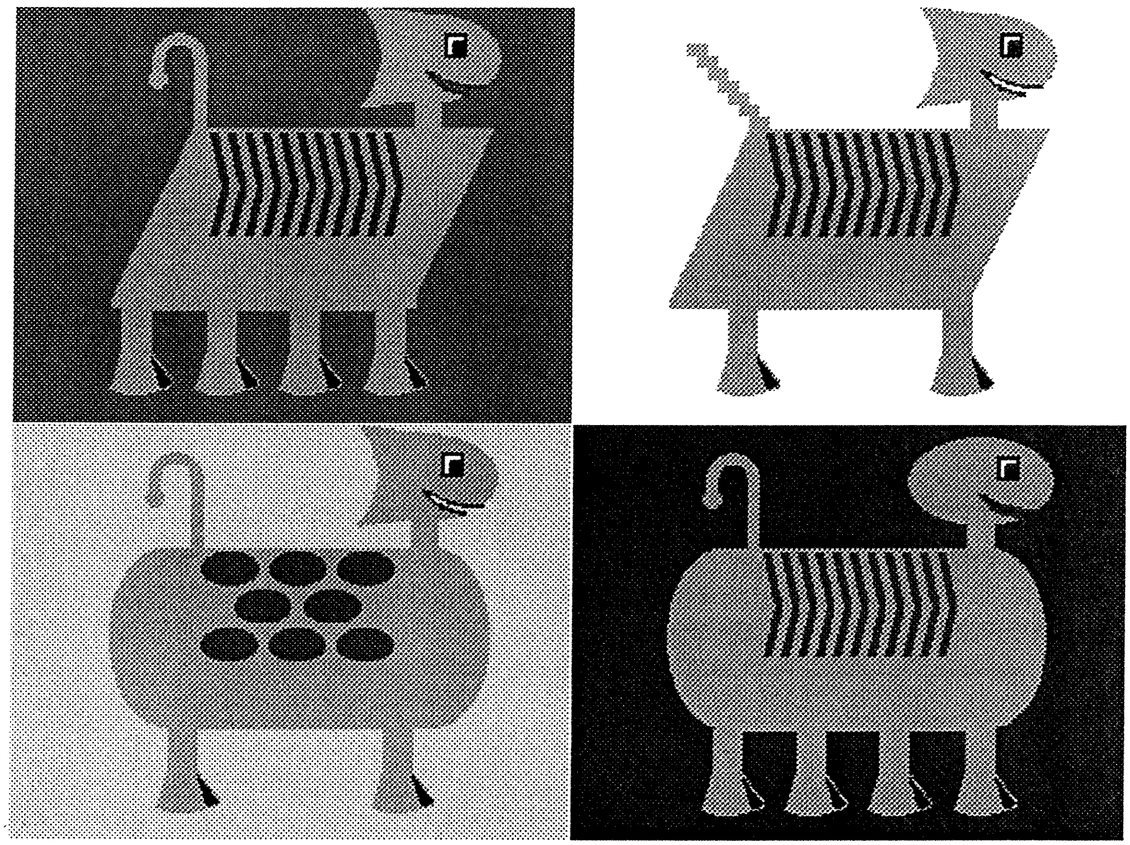
Appendix G

Item number	Old items (training phase) Tail type	Back pattern	Head shape	Body type	legs	Match items (test phase) Item number	Tail type	Back pattern	Head shape	Body type	legs	Background
Positive												
1	1	1	1	0	0	10	1	0	1	0	0	4
3	1	0	1	1	1	9	1	1	1	1	1	2
6	0	1	0	1	1	15	0	0	0	1	1	1
8	0	0	0	0	0	13	0	1	0	0	0	3
Negative												
2	0	1	1	0	1	14	0	0	1	0	1	1
4	1	1	0	1	0	16	1	0	0	1	0	3
5	1	0	0	0	1	11	1	1	0	0	1	2
7	0	0	1	1	0	12	0	1	1	1	0	4

New items (memory phase)

Item number	Tail type	Back pattern	Head shape	Body type	legs	Item number	Tail type	Back pattern	Head shape	Body type	legs	Background
Positive												
17	1	1	1	0	1	21	1	0	1	0	1	3
18	1	0	1	1	0	22	1	1	1	1	0	1
19	0	1	0	0	1	23	0	0	0	0	1	2
20	0	0	0	1	0	24	0	1	0	1	0	4
Negative												
25	0	1	1	0	0	29	0	0	1	1	1	2
26	1	1	0	1	1	30	1	0	0	0	0	4
27	1	0	0	1	1	31	1	1	0	0	0	1
28	0	0	1	0	0	32	0	1	1	1	1	3

Appendix H



Remerciements

Je voudrais premièrement remercier Serge Larochelle qui a été un directeur de thèse extraordinaire. Il a été disponible et franc. Il a su me laisser suivre mes idées tout en me prodiguant de bons conseils. Il m'a permis d'apprendre énormément. Je voudrais ensuite remercier mes collègues et amis Denis Cousineau et Christine Lefebvre avec qui j'ai eu plusieurs conversations stimulantes. Je voudrais également remercier Gyslain Giguère, ainsi que Sébastien Hélie et Marie-Andrée Laberge pour leurs contributions en tant qu'auxiliaires de recherche. Finalement, je voudrais remercier tous mes étudiants des cours *laboratoire de psychologie cognitive* qui ont contribué à la collecte de mes données.