Journal of Experimental Psychology:
Learning, Memory, and Cognition
2006, Vol. 32, No. 6, 1403–1415

0278-7393/06/$12.00 DOI: 10.1037/0278-7393.32.6.1403

Exemplar Effects in the Context of a Categorization Rule: Featural and Holistic Influences

Jean-Pierre Thibaut
University of Poitiers

Sabine Gelaes
University of Liège

Brooks and colleagues (S. W. Allen & L. R. Brooks, 1991; G. Regehr & L. R. Brooks, 1993) have shown that the classification of transfer stimuli is influenced by their similarity to training stimuli, even when a perfect classification rule is available. It is argued that the original effect obtained by Brooks and colleagues might have resulted from two potential confounding variables. Once these confounds were controlled, the current authors did not replicate Brooks and colleagues' results in Experiment 1. Exemplar effects appeared in Experiment 2 when transfer stimuli were perceptually more similar to training stimuli than in Experiment 1. In Experiment 3, the authors obtained exemplar effects with separated stimuli, a finding that was not predicted by Brooks and colleagues' model. The authors suggest that a close perceptual match between training and transfer stimuli is necessary for the effect to occur, for both integrated and separated stimuli. The nature of this perceptual match, holistic or featural, is discussed.

Keywords: categorization, exemplar effects, rule, concept learning

How do we, as human beings, classify novel stimuli belonging to a category? This question has received much attention in psychology. In many circumstances, we rely on a limited subset of component dimensions that can be used as a classification rule (e.g., Murphy, 2002; Smith & Medin, 1981). In other cases, we might classify a new stimulus in a category X because it reminds us of a very similar or almost identical item we have met before and that belonged to this category X. In this case, these remindings are helpful even though they are based on superficial features (i.e., idiosyncratic characteristics that describe a particular stimulus or a small subset of category members—e.g., the particular shape or color of a flower pot in one's garden) (see Brooks, 1978).

Exemplar Effects: Principles and Results

In the present contribution, our central idea is that even when participants can rely on a perfect classification rule, their performance with transfer stimuli might be influenced by the similarity between these transfer exemplars and previously seen training exemplars. This is a particularly interesting possibility for stimuli that are very similar to other items even though they belong to different categories (e.g., a bird and a bat). This topic is central to the study of categorization, as many rule-based and exemplar-based models of categorization have been proposed in the literature (see, e.g., Murphy, 2002, for an overview).

Brooks and colleagues (Allen & Brooks, 1991; Regehr & Brooks, 1993) devised an elegant paradigm, which we use here, to study how rules and exemplar similarity might interact in subjects' classifications. They showed that classification of transfer stimuli according to a simple explicit rule was influenced by their similarity to the training instances (see also Brooks, 1978; Brooks, Norman, & Allen, 1991). To illustrate, in Allen and Brooks (1991), one rule was if an animal has at least two of the following three critical (attribute) values—six legs, angular body, spots on the body—it is a Builder; otherwise, it is a Digger. Participants were trained in applying this rule to eight training stimuli. In the transfer phase, eight new stimuli were introduced. Each transfer stimulus was identical to one training item on four dimensions but differed on the fifth dimension (namely, the dimension of spots; e.g., if the training item had spots, its twin transfer item was identical except that it had no spots). This manipulation resulted in *good* transfer items (i.e., items similar to training items and belonging to the same category according to the rule) and *bad* transfer items (i.e., items similar to training items but belonging to the other category). According to Table 1 (see Figure 1 for the stimuli), following the above rule, Items 1, 2, 5, and 8 are Diggers, and Items 3, 4, 6, and 7 are Builders; Transfer Items 10, 12, 13, and 14 are Diggers, and Transfer Items 9, 11, 15, and 16 are Builders. Transfer Items 9, 12, 14, and 16 do not belong to the same category as their training twin and thus are bad transfer (BT) items. Because the items could be categorized perfectly by attending to the features defining the rule, one did not predict more errors or longer reaction times for the BT

Jean-Pierre Thibaut, Language, Mémoire et Développment cognitif (LMDC) Laboratory, Université de Poitiers, Poitiers, France; Sabine Gelaes, Department of Clinical Sciences, Université de Liège, Liège, Belgium.

Part of this work was presented at the Annual Meeting of the Belgian Psychological Society, Liège, Belgium, May 2000. Most of the research was completed when both authors were at the Université de Liège. This research benefited from a Fonds Spéciaux de Recherche (Special Research Funds) grant from the Université de Liège, awarded to Jean-Pierre Thibaut, and from a grant awarded by le Ministère de l'Education et de la Recherche de la Communauté Française de Belgique (Ministry of Education and Research to the French Community of Belgium) to Sabine Gelaes for her doctoral thesis.

We thank Maria Chuy, Julia Wronowski, and Marie Wiart for their help with data collection and Ludovic Le Bigot and Etienne Quertemont for their help with data analysis.

Correspondence concerning this article should be addressed to Jean-Pierre Thibaut, University of Poitiers, LMDC Laboratory, CNRS UMR 6215, 99, Avenue du Recteur Pineau, 86000, Poitiers, France. E-mail: jean-pierre.thibaut@univ-poitiers.fr

Table 1
Logical Description of the Stimuli Used in Experiments 1, 2, and 3

Item	Number of legs	Body shape[a]	Spots[a]	Neck length[b]	Tail length[b]
Training stimuli					
1	1	0	0	0	1
2	0	0	1	1	1
3	1	1	0	1	1
4	0	1	1	0	1
5	0	0	0	0	0
6	1	0	1	1	0
7	1	1	1	0	0
8	0	1	0	1	0
Transfer stimuli in Experiment 1 and Experiment 3A					
9	1	0	1	0	1
10	0	0	0	1	1
11	1	1	1	1	1
12	0	1	0	0	1
13	0	0	1	0	0
14	1	0	0	1	0
15	1	1	0	0	0
16	0	1	1	1	0
Transfer stimuli in Experiment 2					
9	1	1	0	0	1
10	0	1	1	1	1
11	1	0	0	1	1
12	0	0	1	0	1
13	0	1	0	0	0
14	1	1	1	1	0
15	1	0	1	0	0
16	0	0	0	1	0
Transfer stimuli in Experiments 3B and 3C[b]					
9	1	0	1		
10	0	0	0		
11	1	1	1		
12	0	1	0		
13	0	0	1		
14	1	0	0		
15	1	1	0		
16	0	1	1		

Note. For each dimension, zeros and ones represent the following values: number of legs, 1 = six legs, 0 = two legs; body shape, 1 = angular, 0 = round; spots, 1 = present, 0 = absent; neck length, 1 = long, 0 = short; tail length, 1 = long, 0 = short. The eight transfer stimuli were transformations of Training Stimuli 1–8 on the dimension of spots in Experiment 1 and body shape in Experiment 2. In Experiments 3A–3C, the transfer stimuli were transformed on the spots dimension as in Experiment 1, except that the value no spots was replaced by the value stripes.

[a] Note that in Experiments 1, 3A, 3B, and 3C, the dimension of spots is stated in the second position and body shape in the third position in the rule. In Experiment 2, body shape is stated in the second position and spots in the third position of the rule.

[b] In Experiments 3B–3C, the nonrule dimensions neck length and tail length were not used in the stimuli.

items than for the good transfer (GT) items. However, Allen and Brooks showed that GT items were categorized faster and/or more accurately than BT items (a phenomenon hereafter referred to as the *BT-GT effect*).

Regehr and Brooks (1993) showed that the BT-GT effect described by Allen and Brooks (1991) was obtained with "holistically individuated" stimuli (i.e., items composed of features that cohere into an individuated, distinctive whole). By comparison, when stimuli were not holistically individuated, the similarity-to-exemplar effect disappeared. This corresponds to the widely accepted view (see Smith, Patalano, & Jonides, 1998) that in order to obtain exemplar effects with this sort of stimuli, each item should be a coherent individual rather than a collection of separate properties. This is because an item holistically similar to another one is more likely to bring that other item to mind than a collection of separate properties similar to another collection of properties. This is also consistent with the view that categorization based on exemplar similarity is associated with holistic and automatic processing, equal weighting of attributes, and matching of concrete information (Goldstone, 1994; Pothos, 2005; Smith et al., 1998) or that the attributes are not weighted with respect to their categorical value (Regehr & Brooks, 1993). Murphy (2002) also interpreted Regehr and Brooks's study in terms of holistic similarity. In our experiments, using Regehr and Brooks's paradigm, we show that holistic similarity is not sufficient to produce an exemplar simi-

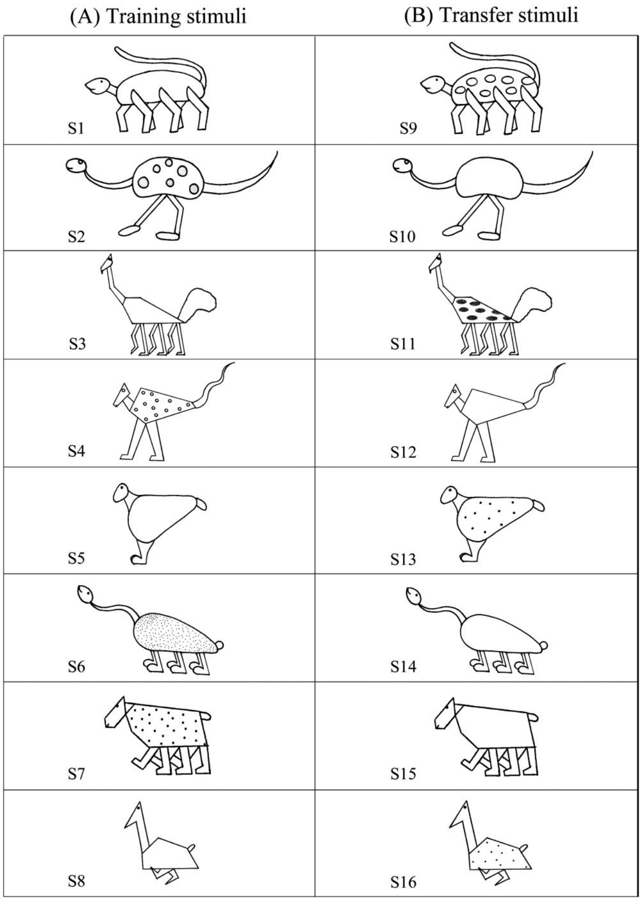

Figure 1. Training items (A) and transfer items (B) used in Experiment 1. The stimuli are from "Perceptual manifestations of an analytic structure: The priority of holistic individuation," by G. Regehr & L. R. Brooks, 1993, *Journal of Experimental Psychology: General, 122,* p. 94. Copyright 1993 by the American Psychological Association.

larity effect and that the effect can be obtained with collections of separate features.

In many cases, each dimension value has its own perceptual implementation that distinguishes it from other instances of the same feature (e.g., all instances of "round head" differ from the other "round heads" in the set), what Regehr and Brooks (1993) call "perceptually individuated features." Indeed, a number of authors have argued that these perceptual features play a role in categorization (see Barsalou, 1999; Goldstone & Barsalou, 1998; Schyns, Goldstone, & Thibaut, 1998; Solomon & Barsalou, 2001). Regehr and Brooks (1993) did not find any BT-GT effect with stimuli composed of distinctive features that did not constitute distinctive wholes, a result they took as evidence favoring the role of holistic similarity. In the present experiments, we want to suggest that these perceptually instantiated features (Brooks & Hannah, 2006) might play a central role in the effect. We hypothesize that a perfect perceptual match at the featural level between a training item and its transfer twin (i.e., if the stimuli share a subset of perceptually identical features) might lead to a BT-GT effect, especially for stimuli that are not integrated.

Rules and Procedural Differences Between Stimuli

Regehr and Brooks (1993) showed an exemplar effect when a perfect rule was available, that is, in a context in which there was no need to use exemplar similarity to solve the task. However to be sure that exemplar similarity influenced the results, it is fundamental to establish that all of the stimuli were procedurally equivalent in terms of the rule. Hereafter we describe two potential procedural differences between the stimuli used by Allen and Brooks (1991) and Regehr and Brooks (1993); the first one is related to a difference between prototypes and the other items, the second one to the order in which the dimensions are given in the rule. These differences might have contributed to the BT-GT effect described by the authors and might undermine their conclusions regarding the role of exemplar similarity.

Regarding prototypes, they have the three dimension values associated with their category, and any test on two rule-defining dimensions leads to a correct answer. By contrast, the six remaining stimuli have two attribute values of the rule pointing to one category and the third value pointing to the opposite category. This conflict might contribute to an increase in the decision times and the number of errors as compared with the prototypes. We call this difference between prototypes and the other stimuli the *prototype advantage*.[1] Because prototypes are always positive training and transfer items, this advantage results in a procedural imbalance between positive and negative items (addressed below).

Lacroix, Hélie, and Larochelle (2000) have described another procedural difference between the stimuli associated with the order of presentation of the features defining the rule. If participants checked the three dimensions in the order stated by the rule, what Regehr and Brooks (1993) called positive stimuli (i.e., the GT items and their corresponding training stimuli, called positive training items) and negative stimuli (i.e., the BT items and the corresponding negative training stimuli) were not equivalent. For the positive items, a test on the first two dimensions was always sufficient for classification. For the negative items, the first two dimension values always pointed to opposite categories, so that a test on the third dimension was requested to classify them. Referring to the abstract description given in Table 1 and the above rule, Training Items 2, 3, 5, and 7 and their corresponding transfers could be categorized by attending to the two first dimensions (i.e., number of legs and body shape). They are positive items. We call this the *serial order* parameter. The cause of this imbalance is that transfer items result from a change on the last rule dimension value (i.e., spots, in our example). If we put the transformed diagnostic dimension in the second position instead of the third, then the number of items that can be categorized with two tests or with three tests, if subjects test each dimension in the order specified by the rule, is equivalent for both positive and negative stimuli.

Overview of the Experiments

The purpose of Experiment 1 was to check whether the BT-GT effect would appear when the serial order parameter was controlled. A second purpose was to compare the analyses of the data when prototypes were included in or removed from the data set in the context of the paradigm introduced by Regehr and Brooks (1993). We checked whether the BT-GT effect would appear when the prototype advantage was controlled. Experiment 1 revealed that no BT-GT effect appeared when the influence of prototypes was controlled.

The third purpose was to study the relationships between the amplitude of the BT-GT effect and the overall similarity between holistically individuated stimuli. To do this, in Experiment 2, we used transfer stimuli that were more similar to their training twin than the transfer stimuli used by Regehr and Brooks (1993). We hypothesized that more similar transfer stimuli might elicit the BT-GT effect absent in Experiment 1.

A fourth general purpose was to show that a BT-GT effect could be obtained with separated stimuli (i.e., stimuli split into their components). In Experiments 3A to 3C, we investigated whether a BT-GT effect might appear with these separated stimuli. If stimulus integration is necessary for the BT-GT effect, it should not appear with separated stimuli. Moreover, separated stimuli should promote analytical processing: Dimensions should be tested one by one until the evidence is sufficient to ground an unambiguous answer.

Experiment 1

This first experiment was a replication of Regehr and Brooks's (1993) Experiment 3A using the same set of training and transfer stimuli and the same rules. The difference was that the serial order confound was controlled. Regarding this confound, note that in the following sections, "verifiable in two (or three) tests" should be understood to mean verifiable in two (or three) tests if one follows the order in which the dimensions are given in the rule. Thus, "verifiable in two (or three) tests" refers to the number of steps necessary to classify this stimulus unambiguously if one checks the stimulus features in the order given by the rule and does *not*

---

[1] In fact, this potential difference was brought to our mind by Allen and Brooks (1991) and Regehr and Brooks (1993, p. 100, footnote 2), who noticed this confound. The authors did not mention that this was also the case for training stimuli; the training phase prototypes are always positive training items.

refer to the number of feature values in a stimulus that belong to the set of feature values given by the rule.

Critical comparisons will refer to the difference between GT and BT items, on the one hand, and between positive and negative training items, on the other. Of note, two types of analyses will be compared: analyses with prototypes removed from the data set and analyses with prototypes included in the data set. If we obtain a GT-BT effect with prototypes included in the data set, as was the case in Allen and Brooks (1991), in Regehr and Brooks (1993), and in Lacroix, Giguère, and Larochelle (2005), but not with prototypes excluded, this will strongly suggest that the BT-GT effect obtained in other studies was caused by this imbalance between prototypes and the other items.

Method

Participants

Twenty-four undergraduates from the University of Liège served as unpaid volunteers in this experiment.

Materials

We used the same training and transfer sets of stimuli as in Regehr and Brooks (1993). The training set contained the eight original stimuli (line drawings of imaginary animals) designed by Regehr and Brooks in their Experiment 3A (see Figure 1A). These animals were made up of five dimensions, each taking two values: number of legs (two vs. six), body shape (round vs. angular), spots (present vs. absent), tail length (short vs. long), and neck length (short vs. long). Participants had to categorize the stimuli into two categories (Builders or Diggers) on the basis of a three-feature additive rule. The rule was based on the three following dimensions: number of legs, body shape, and spots. A Builder possessed at least two of the three dimension values given by the rule. All other animals were deemed to be Diggers. Each value of the two nonrule dimensions, tail length and neck length, appeared equally often in both categories, and thus these dimensions were irrelevant for categorization (the expressions *nonrule* and *irrelevant* dimensions are used synonymously throughout the text, as are the terms *attributes*, *features*, and *dimensions*). (See Table 1 for a logical description of the stimuli.) Recall that all of the dimension values, including the nonrule dimension values, were individuated in Regehr and Brooks's sense (see Figure 1A).

Four rules were used to counterbalance the stimuli. The four rules defined below are associated with Builders; the serial order confound was controlled in the rules.

> *Rule 1*: Six legs, spots present, and angular body.
>
> *Rule 2*: Two legs, spots present, and angular body.
>
> *Rule 3*: Six legs, spots absent, and round body.
>
> *Rule 4*: Two legs, spots absent, and round body.

The eight transfer stimuli were the ones used by Regehr and Brooks (1993, Experiment 3A). Each transfer stimulus was very similar to a single twin training item. The difference between a transfer stimulus and its twin was always on the spots dimension, which took the opposite value (e.g., if spots were present on a training item, spots were absent on the transfer twin) (see Figure 1B). This resulted in two types of transfer items:

> *GT item*: A stimulus seen for the first time in the transfer phase that belongs, according to the rule, to the same category as its twin training stimulus.
>
> *BT item*: A stimulus seen for the first time in the transfer phase that belongs, according to the rule, to the category opposite to its twin training stimulus.

Two types of training items were associated with these two types of transfer items:

> *Positive training item*: A stimulus introduced during the training phase, a training twin of a GT item.
>
> *Negative training item*: A stimulus presented during the training phase, a training twin of a BT item.

It is important to note that a priori, the two classes of training stimuli get their name, positive or negative, by reference to their transfer twin. For example, in Table 1, with the rule "Builders have at least two of the three values—six legs, spots on the body, angular body" (note that the dimensions are mentioned in a different order in Table 1—i.e., number of legs, body shape, and spots), Items 1, 2, 5, and 8 in the training set and Items 10, 12, 13, and 14 in the transfer set are Diggers, and Items 3, 4, 6, and 7 in the training set and Items 9, 11, 15, and 16 in the transfer sets are Builders. Transfer Items 9, 12, 14, and 16 do not belong to the same category as their training twin and are BT items corresponding to Negative Training Items 1, 4, 6, and 8, whereas GT items (10, 11, 13, and 15) correspond to Positive Training Items 2, 3, 5, and 7. Recall that positive and negative training items are not equivalent, as prototypes are always positive items (e.g., according to the above rule, Items 5 and 7 are prototypes of their respective category). The two other positive items are verifiable in three tests. For the negative items, there are two items verifiable in two tests and two items verifiable in three tests (according to the stated rule, Items 6 and 8 are verifiable in two tests).

Procedure

Participants were tested individually. They were seated about 70 cm from the screen of an Apple Macintosh LC 630 or Power Macintosh G3. Stimuli were displayed on the computer screen. The program Superlab (Cedrus Corporation, 1989) was used to display the instructions and stimuli and to record the answers. Participants had to press one of two keys on the keyboard (Builder = *4* and Digger = *5* on the numeric pad) to make their categorical decision. The reaction time was the interval of time between the onset of stimulus presentation and the participant's response. The stimuli were displayed until an answer was given. The experiment was composed of two phases: a training phase and a transfer phase.

Training phase. Participants were told that they had to learn to classify line drawings of imaginary animals into two categories according to a rule. They were asked to categorize the stimuli as quickly and as accurately as possible. Each training stimulus was presented five times (yielding 40 trials) in a random order. The classification rule was written on a sheet of paper left in view during the entire experiment. Feedback was provided after each classification.

Transfer phase. The eight transfer stimuli were presented randomly. By contrast with Regehr and Brooks (1993), the eight training items were not included in the transfer phase. Indeed, the mixture of training and transfer items might enhance the BT-GT effect because of the temporal closeness between training and transfer items. However, four stimuli from the training set were presented before the transfer items for familiarization with the absence of feedback. Participants had to classify the animals according to the rule as quickly as possible. No feedback was given.

Results

Introduction to the Analyses

We compared BT items with GT items, on the one hand, and positive training items with negative training items, on the other

Table 2
Median Response Times (in ms) and Proportions of Errors for Positive and Negative Training Stimuli and GT and BT Items in Experiment 1

	Pos		Neg				GT		BT			
Data	*Mdn*	*SD*	*Mdn*	*SD*	Neg minus Pos	Overall training	*Mdn*	*SD*	*Mdn*	*SD*	BT minus GT	Overall transfer
RTs[a] ($n = 20$)	1,374	699	1,433	745	59	1,404	1,519	513	1,576	584	57	1,548
RTs[b] ($n = 24$)	1,214	552	1,405	834	191	1,310	1,370	451	1,672	716	302	1,521
Errors[a] ($N = 24$)	.13	.27	.21	.33	.083	.167	.06	.17	.10	.25	.04	.08
Errors[b] ($N = 24$)	.09	.18	.17	.25	.073	.132	.03	.08	.16	.22	.13	.09

Note. GT = good transfer; BT = bad transfer; Pos = positive training stimuli; Neg = negative training stimuli; RTs = response times.
[a] Without prototypes and negative items that could be verified in two tests included in the data set. [b] With all data included in the data set. There are no prototypes in the negative items.

hand. To be compelling, a difference between BT and GT items (the BT-GT effect) must appear together with *no* difference between positive and negative training items. A significant difference between BT and GT items together with a significant difference between positive and negative training items would reveal a general bias in favor of positive items.

Allen and Brooks (1991) and Regehr and Brooks (1993) did not use this definition of the BT-GT effect. In their Experiment 3A, Regehr and Brooks compared the training items (called "olds") taken as a class (i.e., positive and negative training items being merged into one data set) with GT items on the one hand and GT items with BT items on the other hand. The authors' reasoning was that the absence of a difference between training items taken as a class and GT items, on the one hand, and a difference between GT and BT items, on the other hand, would be an adequate test of the influence of previously seen exemplars. This comparison, however, assumes that positive and negative *training* items are equivalent, which is not the case—all the prototypes are positive items (as discussed in the introduction).

Allen and Brooks (1991) proposed another analysis in which they compared positive training items ("positive olds" in their terminology) with GT items and negative training items ("negative olds") with BT items. These analyses were performed on the entire set of stimuli. The BT-GT effect was obtained when the second difference was larger than the first one, that is, when there was a significant interaction between type of item (positive vs. negative) and phase (training vs. transfer). However, within this pattern of results, there might be a significant difference between BT and GT items and a smaller, but significant, difference between positive and negative training items. This would mean that the class of negative items is "special" compared with positive items, which should be avoided.

In our case, because the serial order confound was controlled in the rule, to achieve computational equivalence between positive and negative items, we removed the positive items verifiable in two tests (i.e., prototypes) and the negative items verifiable in two tests. We also kept the positive and the negative items verifiable in three tests (see *Materials*). Further, we performed an analysis on the entire set of data to test the contribution of prototypes. The comparisons between positive and negative training items and between GT and BT items were performed with planned comparisons. In all of the experiments, we set the significance level at .05.

Response Times

When only items verifiable in three tests were included in the analysis, 4 participants were removed from the analyses owing to empty cells ($n = 20$). The 2 × 2 ANOVA with phase (training vs. transfer) and stimulus type (positive vs. negative) as repeated factors revealed no main effect of phase, $F(1, 19) = 1.67$, MSE = 249,857 ($\eta^2 = .081$); no main effect of stimulus type, $F(1, 19) = 0.58$, MSE = 116,396 ($\eta^2 = .030$); and no interaction, $F(1, 19) = 0.001$, MSE = 51,976 ($\eta^2 = .008$). Planned comparisons showed no significant difference between positive and negative training stimuli, $F(1, 19) < 1$, MSE = 71,872 ($\eta^2 = .025$), or between GT and BT items, $F(1, 23) < 1$, MSE = 96,500 ($\eta^2 = .017$; with an expected mean difference of 300 ms, the power was .86[2]; see Table 2).

With all of the data included in the analysis (i.e., with prototypes and negative items verifiable in two tests), the 2 × 2 ANOVA with phase (training vs. transfer) and stimulus type (positive vs. negative) ($N = 24$) revealed no main effect of phase and no significant interaction. However, there was a main effect of stimulus type, $F(1, 23) = 9.29$, MSE = 157,333. Planned comparisons revealed a significant difference between positive and negative training stimuli, $F(1, 23) = 4.82$, MSE = 91,039, and a difference between GT and BT items, $F(1, 23) = 9.63$, MSE = 114,392. Thus, of note, we found a difference between positive and negative training items, a difference between GT and BT items, and no interaction (see Table 2).

---

[2] Given that we did not obtain any BT-GT effect, it is important to assess the power of our statistical analyses. Thus, in the experiments in which we did not obtain the BT-GT effect, we report power estimates based on an effect size equal to significant BT-GT effects that were found in the other experiments. Our first estimation was based on the BT-GT effects obtained in Experiment 2—that is, a difference of 626 ms for the response times and .425 for the proportion of errors. With these effect sizes, the power estimates were always superior to .95, close to 1. The power estimates reported in the text were calculated with more conservative effect sizes—that is, 300 ms for response times and .20 for the proportion of errors, which are closer to the significant BT-GT effects obtained in Experiments 3A and 3B.

Proportion of Errors

With items verifiable in three tests, the 2 × 2 ANOVA ($N = 24$) showed no significant effect of phase, $F(1, 23) = 2.63$, $MSE = 0.063$ ($\eta^2 = .103$), or of stimulus type, $F(1, 23) = 1.68$, $MSE = 0.056$ ($\eta^2 = .068$), and no significant interaction, $F(1, 23) = 0.18$, $MSE = 0.059$ ($\eta^2 = .008$). Planned comparisons showed no significant difference between positive and negative training stimuli, $F(1, 23) = 1.35$, $MSE = 0.062$ ($\eta^2 = .056$), or between GT and BT items, $F(1, 23) = 0.39$, $MSE = 0.053$ ($\eta^2 = .017$; with an expected mean difference of .20, the power was .855; see Table 2).

With prototypes and items verifiable in two tests included in the data set, the same ANOVA revealed no significant effect of phase and no significant interaction. However, of interesting, there was a significant main effect of stimulus type, $F(1, 23) = 10.80$, $MSE = 0.02$, with more errors for negative items ($M = 0.161$) than for positive items ($M = 0.063$). Planned comparisons showed no significant difference between positive and negative training stimuli, $F(1, 23) = 3.13$, $MSE = 0.020$, and a significant difference between BT and GT items, $F(1, 23) = 5.06$, $MSE = 0.027$. The significant effect of stimulus type and the absence of a significant interaction between phase and stimulus type show that positive items as a class are special compared with negative items. The BT-GT effect with prototypes included in the data set and its absence when prototypes were removed show that prototypes contribute to the BT-GT effect (see Table 2).

Discussion

The main purpose of this experiment was to replicate Regehr and Brooks's (1993) Experiment 3A with the serial order parameter controlled and with the effect of prototypes either controlled or not. Several results are worth stressing. First, most important, there was no BT-GT effect when the data included in the analyses were properly controlled. Second, when the positive and the negative training items were separated in the analyses and when the prototypes were included in the data set, as in Allen and Brooks (1991) and in Lacroix et al. (2005), we obtained a significant difference between positive and negative items for both response times and errors. Note that we also ran the same experiment as in Regehr and Brooks (1993)—that is, with no control of the serial order effect. There was a BT-GT effect when we performed their analyses (see *Introduction to the Analyses*). It disappeared when the prototypes were removed from the data set. However, there is a remaining difference between the design used in the latter experiment and Regehr and Brooks's experiment: Their transfer phase contained the eight training and eight transfer items, which was not the case in our experiment. Regehr and Brooks's idea was to make the retrieval context more similar to the training context. It is possible that within our framework, the BT-GT effect would have occurred with the eight training items shown in the transfer phase. However, this would restrict the original Regehr and Brooks BT-GT effect to the specific context of a very close temporal relation between training and transfer items. Moreover, we also introduced four training items at the beginning of the transfer phase, and so the difference between our methodology and Regehr and Brooks's methodology might be only quantitative, not qualitative. What we can say is that with our methodology, when the prototype advantage and the serial order parameter were controlled, there was no evidence of a BT-GT effect.

Experiment 2: Similarity Between Training and Transfer Stimuli

As mentioned in the introduction, according to Pothos (2005), Regehr and Brooks (1993), and Smith et al. (1998), similarity-to-exemplar effects result from holistic similarities between items. For integrated stimuli, a transformation on a local dimension contributes to two modifications, one at the local dimensional level and one at the holistic level. Depending on the transformed dimension, the holistic similarity between a transfer stimulus and its training twin might be more or less important, perceptually.

Regehr and Brooks (1993), in their Experiments 3A and 3B, manipulated the size of the transformation (i.e., the similarity) between the training phase stimuli and the transfer phase stimuli. In their Experiment 3B, they used the same stimuli as in their Experiment 3A (or in our Experiment 1) except that they also transformed the external, irrelevant dimensions, neck length and tail length (see *Materials*), in the transfer stimuli. This resulted in a quite distinct holistic appearance between each training stimulus and its transfer twin. They did not observe any BT-GT effect.

In the present experiment, by comparison with Experiment 1, the idea was to increase the similarity between each training item and its twin transfer. The absence of a similarity-to-exemplar effect in Experiment 1 might have been due to the fact that each learning exemplar and its twin were not similar enough holistically (even though they differed on only one perceptual dimension out of five and were quite similar).

In the present experiment, we replicated Experiment 1 except that the transformed transfer dimension was body shape (rounded vs. angular) instead of spots. The training phase stimuli were the same as in Experiment 1. We first show that the transfer stimuli that were transformed on body shape are more similar holistically to their training twins than the corresponding transfer stimuli that were transformed on spots.

Method

Participants

Thirty unpaid undergraduates from the University of Liège participated in the experiment, 10 in a similarity-rating task and 20 in the classification task itself.

Materials

We used the same set of training exemplars as in Experiment 1. A new set of eight transfer items was created. They were transformed on the dimension body shape (instead of spots as in Experiment 1). When a training stimulus had a round body, its transfer twin had an angular body, and vice versa (see Figure 2). The rules were constructed as in Experiment 1. The transformed dimension (i.e., body shape) was mentioned in the middle of the rule for a proper control of the serial order parameter (i.e., "number of legs, body shape, and spots").

Procedure

Similarity judgment task. We used a forced-choice similarity judgment task. Participants were requested to choose between the two types of transfer stimuli (transformed on the spots dimension or on the body shape dimension) the one that was the most similar to the corresponding training stimuli (the standard). The stimuli were displayed via a computer screen. For each trial, one of the eight training stimuli was displayed at the top of

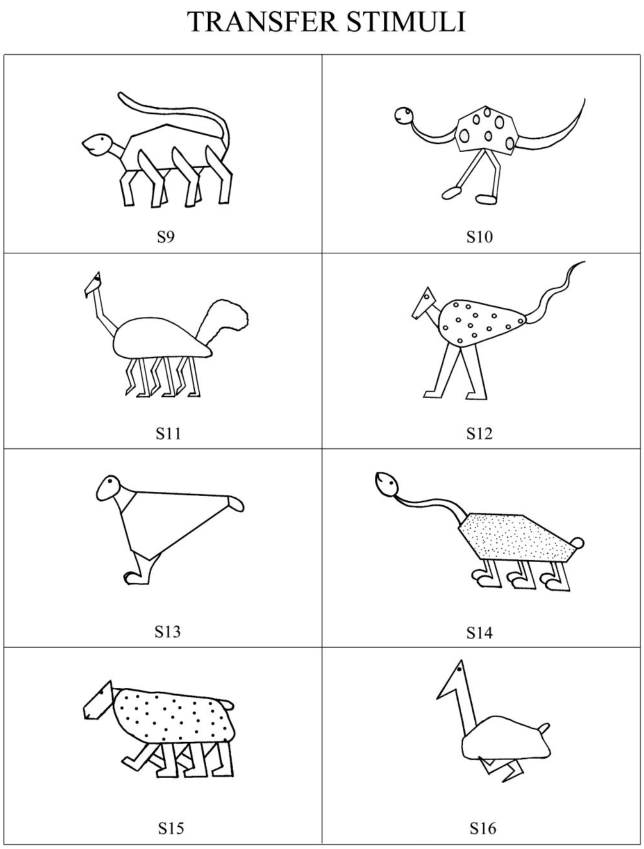

Figure 2. Transfer items used in Experiment 2.

the computer screen together with its two transfer twins, displayed below in a row. The position of each type of transformed stimulus was counterbalanced across trials. Each triad of stimuli was displayed for 300 ms. A short presentation time was chosen to avoid analytically based ratings (Ward, 1983). Participants had to answer by pushing the numerical key corresponding to their choice as quickly as possible: *4* for the stimulus on the left and *5* for the stimulus on the right. Eight filler trials were introduced in which the two transformed stimuli differed from the standard on other dimensions (i.e., tail length, neck length, and number of legs). This was done to avoid any systematic answer in favor of one dimension.

Categorization task. The procedure was the same as in Experiment 1. The same eight training stimuli were presented five times (yielding 40 trials). In the transfer phase, the transfer items transformed on the dimension of body shape were introduced as in Experiment 1.

Results

Similarity Rating Task

The purpose was to compare the similarity between each type of transformed stimuli (i.e., on body shape or on spots) and the training stimuli. We computed the percentage of body shape choices per participant. A mean score of 50% would correspond to an absence of discrimination between the two types of stimuli, whereas a score significantly beyond 50% would reveal a preference in favor of body shape choices. The mean percentage of body shape choices was 76.25%, which differed significantly from 50%, $t(9) = 3.8$, $p < .005$. Thus, the transfer stimuli transformed on body shape were judged more similar to their training twins than the equivalent transfer items transformed on spots.

Analysis of the Data for the Classification Task

A 2 × 2 ANOVA with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors was performed on participants' median response times for correct classifications and proportion of errors. Again, the analyses were performed on the items verifiable in three tests (see Table 3). Planned comparisons were also performed to investigate the difference between positive and negative training items and between GT and BT items. We also conducted an analysis on the entire data set.

Response Times

Six participants were lost in the analysis owing to empty cells.[3] The 2 × 2 ANOVA ($n = 14$) showed a main effect of phase, *F*(1, 13) = 25.52, *MSE* = 209,724 ($\eta^2 = .663$), and stimulus type, *F*(1, 13) = 16.55, *MSE* = 97,049 ($\eta^2 = .560$). Of note, the interaction was significant, *F*(1, 13) = 11.51, *MSE* = 100,580 ($\eta^2 = .470$). Planned comparisons revealed no significant difference between positive and negative training stimuli, *F*(1, 13) = 0.58, *MSE* = 31,866 ($\eta^2 = .042$), but a significant difference between GT and BT items, *F*(1, 13) = 16.56, *MSE* = 165,763, $\eta^2 = .56$.

In the analysis with all of the data included in the data set, the 2 × 2 ANOVA with phase (training vs. transfer) and stimulus type (positive vs. negative) ($n = 17$) revealed a main effect of phase, *F*(1, 16) = 45.51, *MSE* = 158,102; a main effect of stimulus type, *F*(1, 16) = 31.49, *MSE* = 97,321; and a significant interaction, *F*(1, 16) = 12.19, *MSE* = 119,899. Planned comparisons revealed no significant difference between positive and negative training stimuli, *F*(1, 16) = 0.71, *MSE* = 52,961, but a significant difference between GT and BT items, *F*(1, 16) = 26.97, *MSE* = 164,259.

Proportion of Errors

The 2 × 2 ANOVA ($N = 20$) showed no significant main effect of phase, *F*(1, 19) = 3.35, *MSE* = 0.075 ($\eta^2 = .150$), but a significant effect of stimulus type *F*(1, 19) = 10.82, *MSE* = 0.065 ($\eta^2 = .363$). There was a reliable Phase × Stimulus Type interaction, *F*(1, 19) = 16.37, *MSE* = 0.069 ($\eta^2 = .463$). Planned comparisons revealed no significant difference between positive and negative training items, *F*(1, 19) = 0.49, *MSE* = 0.051 ($\eta^2 = .025$), but a significant difference between GT and BT items, *F*(1, 19) = 21.88, *MSE* = 0.083 ($\eta^2 = .535$).

As for the analysis with all of the data included in the data set, the 2 × 2 ANOVA with phase (training vs. transfer) and stimulus type (positive vs. negative) ($N = 20$) gave a main effect of phase, *F*(1, 19) = 13.90, *MSE* = 0.04; a main effect of stimulus type, *F*(1, 19) = 15.22, *MSE* = 0.037; and a significant interaction *F*(1, 19) = 26.48, *MSE* = 0.042. Planned comparisons revealed no

---

[3] We lost many participants because of the important number of errors, which results in a larger number of empty cells. If participants categorized uniquely on the basis of exemplar similarity, all of the BT items would be misclassified and there would be no data left for the analysis of response times.

Table 3

Median Response Times (in ms) and Proportion of Errors for Positive and Negative Training Stimuli and GT and BT Items in Experiment 2, Body Shape Condition

Data	Pos		Neg		Neg minus Pos	Overall training	GT		BT		BT minus GT	Overall transfer
	Mdn.	*SD*	Mdn.	*SD*			Mdn.	*SD*	Mdn.	*SD*		
RTs[a] ($n = 14$)	1,313	619	1,306	459	−7	1,310	1,644	688	2,270	656	626	1,957
RTs[b] ($n = 17$)	1,170	481	1,287	342	117	1,228	1,505	527	2,227	628	722	1,866
Errors[a] ($N = 20$)	.15	.24	.10	.26	−.05	.13	.03	.11	.45	.39	.43	.24
Errors[b] ($N = 20$)	.10	.15	.03	.08	−.07	.07	.03	.08	.43	.36	.40	.23

Note. GT = good transfer; BT = bad transfer; Pos = positive training stimuli; Neg = negative training stimuli; RTs = response times.
[a] Without prototypes and negative items that could be verified in two tests included in the data set. [b] With all data included in the data set.

significant difference between positive and negative training stimuli, $F(1, 19) = 1.36$, $MSE = 0.018$, but a significant difference between GT and BT items, $F(1, 19) = 26.43$, $MSE = 0.061$.

Discussion

By contrast with Experiment 1, there was a significant BT-GT effect for both the response times and the errors and no difference between positive and negative training stimuli. The similarity judgments suggest that this difference between the two experiments results from the larger training–transfer similarity in the present experiment than in the first experiment. However, recall that in both experiments, training and transfer stimuli shared four perceptual dimension values, but this similarity was not sufficient to give an exemplar effect in the first experiment. The difference between the two experiments could be the result of the perceptual status of the two dimensions. We chose the dimension body shape because, a priori, we thought that changing round shapes into angular shapes without modifying the size or proportion of the body shapes would make transfer stimuli very similar to their parent stimuli. On the other hand, removing or adding spots is a salient all-or-none transformation. We come back to this issue in the General Discussion.

Experiment 3: The BT-GT Effect and Separated Stimuli

We mentioned in the introduction that similarity-based categorization is often associated with holistic processing and the matching of perceptual information (Goldstone, 1994; Pothos, 2005; Smith et al., 1998). Exemplar effects should not appear with stimuli introduced in a piecemeal manner. Allen and Brooks (1991) obtained no BT-GT effect using separated stimuli presented as lists of written verbal features instead of as drawings. However, this finding can be explained by the absence of perceptual individuation of the verbal features used. Indeed, by contrast, with our stimuli, lists of words are composed of a finite number of letters, are written with the same font, and have more or less the same perceptual appearance.

By contrast, Regehr and Brooks (1993) obtained a BT-GT effect when they introduced integrated stimuli in the training phase and the corresponding features scattered randomly around the screen in the transfer phase. However, the experimenter provided no rule in the training phase. In the absence of any classification rule, participants probably encoded the training items in terms of their dimensions and classified BT items in the same category as their training twins because BT items shared four features out of five with their twins (see also Modigliani, 1971; Modigliani & Rizza, 1971).

In the following experiments, we seek a BT-GT effect with stimuli split into their constituent dimensions in the training and the transfer phase. The identification of a stimulus part as an abstract dimension value (e.g., this is the "spots present" value) requires that the part be encoded with its idiosyncratic characteristics (e.g., the particular shape of the spots). In the transfer phase, these encoded dimensions should evoke their training counterparts and might give rise to the BT-GT effect. By contrast, if one conceives similarity-to-exemplar effects as a matter of holistic similarity, these inspections of individuated features should weaken the BT-GT effect. Indeed, features displayed as individual components should promote the iterative sequence-based testing that is the hallmark of rule-based processing (Smith et al., 1998).

In Experiments 3A and 3B, the location of each dimension was varied from one presentation of a stimulus to the next. We chose this option because in a preliminary experiment, there was no BT-GT effect when each dimension had the same location across trials. The four dimensions head, body shape, mark on body (spots vs. stripes), and tail were arranged in a row, and the legs were placed below this row. Each dimension was clearly spatially separated from the neighboring parts. We speculated that the BT-GT effect did not appear because each feature location was predictable and the nonrule features might have been ignored during training and, as a result, did not contribute to the training–transfer similarity. We also speculated that though the features were disconnected, an overall shape was still apparent and could have blocked the BT-GT effect, as in Experiment 1.

Experiment 3A

In the present experiment, we used stimuli composed of the same five dimensions as in Experiment 1. We predicted that if the location of each dimension changed from one stimulus to the next, participants would have to process the five dimensions to test the rule-defining dimensions. Thus, all dimensions (and their idiosyncratic characteristics) should be encoded, which might increase the likelihood of a BT-GT effect.

Method

Participants. Forty-seven undergraduates from the University of Liège served as unpaid volunteers in this experiment.

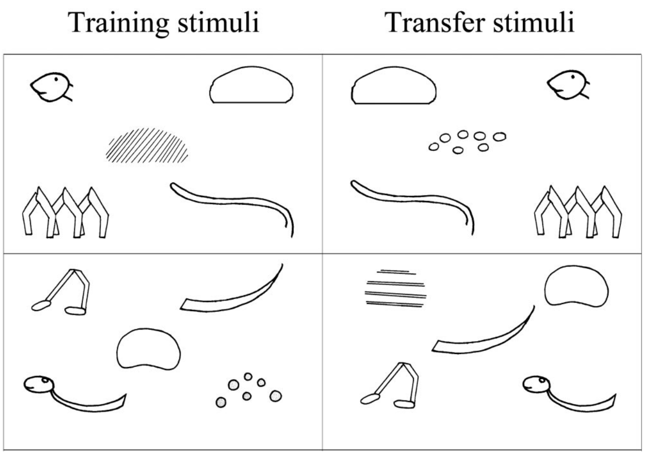

Figure 3. Two training–transfer pairs of stimuli for Experiment 3A. Training items are on the left, and their transfer twins on the right.

Materials and procedure. Two sets of stimuli were designed for this condition. The set of training stimuli was created by breaking the eight line drawings used in Experiment 1 into their five constitutive parts (i.e., head with neck, body, spots, legs, and tail). Because it was difficult to represent the dimension value "spots absent" used in the original stimuli as a separated component, this value was replaced by the value "stripes." The eight transfer stimuli were built up in the same way as the training stimuli. They differed from their training twins on the dimension of body marks: The value "spots present" on a training stimulus was replaced by the value "stripes present" on its transfer twin, and vice versa (see Figure 3).

The individual features were displayed on a 12-cm-wide × 8-cm-high rectangle in the center of the computer screen. The rectangle was divided into five positions: the four corners and the center (see Figure 3). Each of the five dimensions of a given stimulus was randomly assigned to one of the five positions. Because in the training phase, the set of stimuli was presented 12 times, there were 12 different instances of the same stimulus. The training and the test phases were the same as in Experiment 1 (except for the 12 presentations). Participants were told to classify the stimuli into two categories according to a given rule and to do so as quickly and as accurately as possible. The name of each part was given: "You will see a body, angular or rounded; legs, two or six [etc.]. These five dimensions will appear for each stimulus. Follow the rule to classify them."

Results and Discussion

We performed the same analyses as in Experiment 1 on participants' median response times for correct classifications and proportions of errors for items verifiable in three tests.

Response times.[4] We lost 12 participants in the ANOVA owing to missing cells (see footnote 3; $n = 35$). A 2 × 2 ANOVA on the response times, with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors, showed a main effect of phase, $F(1, 34) = 11.14$, $MSE = 460{,}030$ ($\eta^2 = .247$), and no main effect of stimulus type, $F(1, 34) = 0.45$, $MSE = 177{,}412$ ($\eta^2 = .013$). The interaction did not reach significance, $F(1, 34) = 0.92$, $MSE = 157{,}028$ ($\eta^2 = .026$) (see Table 4). A priori planned comparisons revealed no significant difference between positive and negative training stimuli, $F(1, 34) = 0.06$, $MSE = 74{,}518$ ($\eta^2 = .002$), or between GT and BT items, $F(1, 34) = 0.84$, $MSE = 259{,}921$ ($\eta^2 = .024$; with an expected mean difference of 300 ms, the power was .684).

Proportion of errors. A 2 × 2 ANOVA on the proportion of errors, with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors ($N = 47$), showed no main effect of phase, $F(1, 46) = 2.73$, $MSE = 0.049$ ($\eta^2 = .056$), or stimulus type, $F(1, 46) = 2.14$, $MSE = 0.122$ ($\eta^2 = .044$). The interaction did not reach significance, $F(1, 46) = 3.00$, $MSE = 0.064$ ($\eta^2 = .061$). A priori planned comparisons showed no significant difference between positive and negative training stimuli, $F(1, 46) = 0.03$, $MSE = 0.079$ ($\eta^2 = .001$), but a reliable difference between the proportions of errors for GT and BT items, $F(1, 46) = 4.20$, $MSE = 0.107$ ($\eta^2 = .084$). In sum, by contrast with Experiment 1, we obtained a BT-GT effect for errors, together with no difference between positive and negative training items. This important result shows that BT-GT effects can be obtained with separated stimuli.

Experiment 3B: Separated Stimuli, No Irrelevant Dimension

In Experiment 3A, both relevant and irrelevant dimensions might have contributed to the exemplar effect. However, it might also be that the similarity based on the rule-defining features would be sufficient to "produce" the BT-GT effect. In the present experiment, we test this possibility by removing the irrelevant dimensions. A BT-GT effect would show that similarity-to-exemplar influence, with separated stimuli, can result from the sole rule-defining features. It would also suggest that similarity does not necessarily exert its influence on a broader set of features than the rule-defining ones (Lacroix et al., 2005), whereas the absence of a BT-GT effect would mean that the similarity between training and transfer stimuli conveyed by the nonrule features plays a central role in the effect.

Method

Participants. Twenty-three undergraduates from the University of Liège served as unpaid volunteers in this experiment.

Materials and procedure. The stimuli were the same as in Experiment 3A except that we removed the nonrule dimensions (i.e., tail and head). The dimensions of legs, body shape, and body marks defined a triangle (20 cm base × 10 cm height), displayed at the center of the computer screen (see Figure 4). As in Experiment 3A, each stimulus dimension was randomly assigned to one position, and this location differed from one presentation of a stimulus to the next. However, because there are only six permutations of three features, each stimulus was presented twice with the same spatial configuration. The procedure was the same as in Experiment 3A.

Results and Discussion

As in previous experiments, we performed the analyses on participants' median response times for correct classifications and on proportions of errors for items verifiable in three tests.

---

[4] Regarding the importance of prototypes in the BT-GT effect, we obtained the BT-GT effect for both response times and errors in the three experiments when prototypes were included in the analyses. In the same way, there was a significant effect of stimulus type with prototypes included in the data set, for response times and errors in the three experiments, confirming that the asymmetry between the two classes of stimuli in the first case was due to the fact that prototypes differed computationally from the other stimuli.

Table 4

Median Response Times (in ms) and Proportions of Errors for Positive and Negative Training Items and for GT and BT Items in Experiments 3A, 3B, and 3C

	Pos		Neg		Neg minus Pos	Overall training	GT		BT		BT minus GT	Overall transfer
Condition and data	Mdn.	*SD*	Mdn.	*SD*			Mdn.	*SD*	Mdn.	*SD*		
3A: Random separated												
RTs (n = 35)	1,201	455	1,177	372	−24	1,189	1,505	644	1,647	783	142	1,576
Errors (N = 47)	.19	.34	.20	.30	.011	.197	.17	.26	.32	.34	.15	.25
3B: Random separated[a]												
RTs (n = 22)	949	237	965	241	16	957	1,239	254	1,514	512	275	1,377
Errors (N = 23)	.02	.10	.11	.21	.088	.066	.15	.28	.24	.30	.09	.20
3C: Integrated[a]												
RTs (N = 24)	825	328	866	307	41	846	1,609	746	1,531	563	−78	1,570
Errors (N = 24)	.02	.10	.021	.102	0	.021	.06	.17	.10	.21	.04	.08

Note. GT = good transfer; BT = bad transfer; Pos = positive training items; Neg = negative training items; RTs = response times.
[a] Stimuli without the nonrule dimensions.

Response times. We lost 1 participant owing to empty cells. The 2 × 2 ANOVA (n = 22) with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors performed on participants' median response times for correct classifications revealed a main effect of phase and stimulus type, F(1, 21) = 31.90, MSE = 95,731 (η^2 = .603), and F(1, 21) = 9.05, MSE = 91,038 (η^2 = .301), respectively. The interaction was not significant, F(1, 21) = 2.84, MSE = 63,512 (η^2 = .119). The response times in the training phase were shorter than in the transfer phase (see Table 4). Planned comparisons revealed a reliable difference between GT and BT items F(1, 21) = 8.40, MSE = 105,702 (η^2 = .286), but no significant difference between positive and negative training stimuli, F(1, 21) = 2.39, MSE = 48,848 (η^2 = .102).

Proportion of errors. The 2 × 2 ANOVA (N = 23) with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors performed on errors revealed a main effect of phase, with fewer errors in the training phase than in the transfer phase, F(1, 22) = 11.73, MSE = 0.033 (η^2 = .348). There was no significant effect of stimulus type, F(1, 22) = 2.63, MSE = 0.058 (η^2 = .071), and no reliable interaction, F(1, 22) = 0, MSE = 0.04 (η^2 = 0). Planned comparisons showed no significant difference between positive and negative training stimuli, F(1, 22) = 1.87, MSE = 0.026 (η^2 = .078), or between GT and BT items, F(1, 22) = 0.68, MSE = 0.072 (η^2 = .030; with an expected mean difference of .200, the power was .696). In sum, we obtained a BT-GT effect for response times with separated stimuli even when the nonrule dimensions were removed from the stimuli.

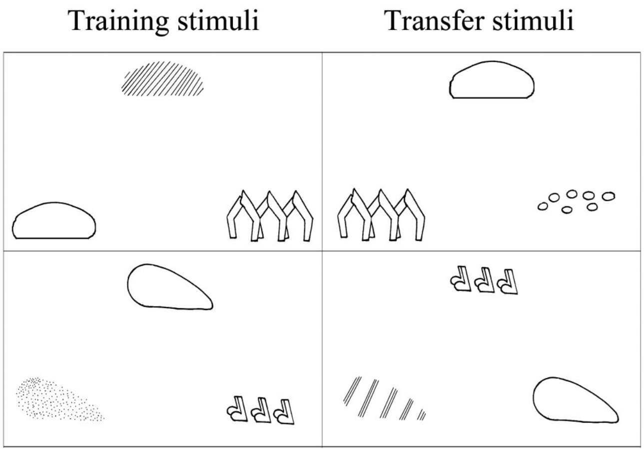

Figure 4. Two training–transfer pairs of stimuli for Experiment 3B. Training items are on the left, and their transfer twins on the right.

Experiment 3C: Holistic Stimuli, Rule-Defining Dimensions

There was no exemplar effect with integrated stimuli in Experiment 1, whereas a BT-GT effect was obtained for errors in the corresponding separated stimuli in Experiment 3A. In Experiment 3B, there was a BT-GT effect for response times. We now verify whether the asymmetry between Experiments 1 and 3A also holds for Experiment 3B and the corresponding integrated condition. If we obtain a BT-GT effect with integrated stimuli, this would mean that the irrelevant dimensions are not necessary for the effect. This would be consistent with Lacroix et al.'s (2005) idea that the BT-GT effect stemmed from the rule-defining dimensions. The absence of a BT-GT effect, by contrast, would suggest that the holistic difference between training and transfer stimuli is sufficient to decrease the influence of exemplar similarity, as in Experiment 1.

Method

Participants. Twenty-four unpaid undergraduates participated in this experiment.

Materials and procedure. A new set of training stimuli and transfer items was created. We erased the nonrule dimensions from the original stimuli (i.e., tail and head) so that only the rule-defining dimensions remained, and we replaced the "no spots" value with the stripes used in Experiment 3B. The procedure was the same as in Experiment 3A.

Results and Discussion

The same analyses as in the previous experiments were run on participants' median response times for correct classifications and proportion of errors for items that must be checked on the three relevant dimensions (see Table 4).

Median response times. The 2 × 2 ANOVA (N = 24) with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors performed on participants' median response times for correct classifications revealed a reliable effect of phase, $F(1, 23) = 46.94$, MSE = 256,575 ($\eta^2 = .671$). The effect of stimulus type was not significant, $F(1, 23) = 0.001$, MSE = 150,774 ($\eta^2 = .000$). The interaction did not reach significance, $F(1, 23) = 1.60$, MSE = 86,873 ($\eta^2 = .065$), and there was no significant difference between positive and negative training stimuli, $F(1, 23) = 3.80$, MSE = 17,149 ($\eta^2 < .142$), or between GT and BT items, $F(1, 23) < 1$, MSE = 220,500 ($\eta^2 < .014$; with an expected mean difference of 300 ms, the power was .577).

Proportion of errors. The 2 × 2 ANOVA (N = 24) with phase (training vs. transfer) and stimulus type (positive vs. negative) as within-subject factors performed on errors revealed no significant effect of phase, $F(1, 23) = 3.29$, $MSE = 0.029$ ($\eta^2 = .125$), or stimulus type, $F(1, 23) = 0.49$, $MSE = 0.021$ ($\eta^2 = .021$). The interaction failed to reach significance, $F(1, 23) = 0.49$, $MSE = 0.021$ ($\eta^2 = .021$). No significant difference appeared between positive and negative training stimuli, $F(1, 23) = 0$, $MSE = 0.011$ ($\eta^2 = .00$), or between GT and BT items, $F(1, 23) = 0.66$, $MSE = 0.032$ ($\eta^2 = .00$; with an expected mean difference of .200, the power was .977). In sum, with these new simplified integrated stimuli, the exemplar effect was absent for both response times and errors.

Summary of Experiments 3A to 3C

The purpose of Experiments 3A to 3C was to study in which conditions the BT-GT effect could be obtained with separated stimuli. The holistic hypothesis predicted that the BT-GT effect should appear mainly when stimuli cohere into individuated wholes and should not appear with separated stimuli (Regehr & Brooks, 1993). By contrast with this view, we obtained a BT-GT effect in separated conditions (Experiments 3A and 3B), whereas the effect was absent in the corresponding integrated conditions (Experiments 1 and 3C).

General Discussion

In the present article we have considered the influence of similarity to previously seen exemplars in a classification task in which a perfect rule was available (Regehr & Brooks, 1993). We assessed the respective contributions of rule-based processing considered as an attention-based mechanism focusing on a subset of rule-defining features and of similarity-to-exemplar influence conceptualized by several authors (see introduction) as automatic, operating incidentally, requiring no (or little) selective attention. Regehr and Brooks (1993) associated the BT-GT effect with *holistically* integrated stimuli.

In the present experiments, we controlled two potential confounds. First, the prototype advantage referred to the fact that the three rule-defining features point to the same category in the case of prototypes, which was not the case for the other items, and that prototypes were always positive items. Second, the serial order parameter referred to the fact that some stimuli could be verified in two tests when participants followed the order given by the rule whereas other stimuli required three tests. We have suggested that the BT-GT effect observed by Regehr and Brooks (1993) and Lacroix et al. (2005) in the same experimental context as our Experiment 1 could have resulted from these confounds. The present experimental contribution revealed that using the same paradigm and stimuli as Regehr and Brooks, we obtained the same BT-GT effect when the analyses included prototypes. The effect disappeared when the two confounds were under control, suggesting that at the end of the training phase, computational differences between prototypes and the other stimuli influenced classifications. Our contribution also revealed that holistic individuation did not always produce the BT-GT effect, that the effect could be obtained with separated stimuli,[5] and that nonrule features were not necessary to obtain the effect.

Second, we investigated the role of the holistic similarity between each training stimulus and its corresponding transfer items on the BT-GT effect. There was no effect in Experiment 1, whereas it appeared in Experiment 2 with transfer stimuli that were more similar to their training twin than in Experiment 1. Third, the effect appeared with stimuli composed of perceptually individuated, spatially separated features in Experiments 3A and 3B—that is, with stimuli in which holistic features, if any, were not very salient.

Encoding Individual Features and Integrated Stimuli As a By-Product of Rule Verification

In the case of integrated stimuli, we have to discuss the BT-GT effect obtained in Experiment 2 and its absence in Experiments 1 and 3C. To explain this pattern of results, we posit, first, that each exemplar is encoded in memory as a result of processing. Each verification of the rule is an opportunity to encode the stimulus to be classified. In this vein, Logan (1988) claimed that encoding and retrieval from memory of all information associated with a stimulus are unavoidable consequences of attention to this stimulus. Second, as a result of the repetition of the stimuli, this encoding can be described as stimulus imprinting (Goldstone, 1998). One can think of the effect as the development of specific detection routines that are specialized in stimuli and their parts. Goldstone

---

[5] With the large number of repetitions of the training stimuli (12 per item), one could argue that a given transfer stimulus is likely to have a training stimulus that it resembles as a whole (i.e., including feature position). So, the description of these stimuli as "separated" does not necessarily imply that they would (or could) not be processed holistically. However, this is false for Experiment 3A, in which there were, by definition, 120 possible permutations of the five features that composed a stimulus. Each of the 13 permutations (12 training stimuli + 1 transfer stimulus) chosen for a given instance of a stimulus differed from the other permutations taken for the same stimulus. However, this is possible for Experiment 3B, in which there were only 6 different permutations, so that each spatial configuration of the three features that composed a stimulus was introduced twice in the training phase and a third time in the corresponding transfer stimulus. Thus, we cannot dismiss the possibility that the transfer stimuli were, to some extent, classified on the basis of some holistic spatial resemblance to their transfer twin. However, the major problem with this reasoning is that it does not explain the absence of the BT-GT effect in Experiment 3C, which used holistically integrated stimuli made up of the same features as in Experiment 3B. Indeed, if we admit that this holistic spatial resemblance played a role in the BT-GT effect in Experiment 3B, it should have had a massive effect in Experiment 3C with items that share a perfect holistic spatial resemblance with their twin training stimuli.

(1998) interpreted the exemplar effect obtained by Allen and Brooks (1991) in these terms: “even in situations where one might think . . . rule-based processes are used, there is good evidence that observers become tuned to the particular instances to which they are exposed” (pp. 591–592). At this point, it is important to note that participants might become tuned to both local and holistic features of the stimuli. It is also likely that, depending on the structure of the stimuli and the requirements of the task (Schyns, 1998), this tuning process might result in holistic features more salient than local features, and vice and versa. Along these lines, Murphy (2002) gave an implicit memory interpretation of Brooks and colleagues’ work, in which perceptual learning was thought to be tied to the specific physical properties of the objects (Schacter, 1994). Participants who have categorized a given item many times in Category A have practiced a certain way of making Category A categorizations. When a new item quite similar to an old one appears, the practiced categorization takes over. Third, as suggested by Schacter (1994), the perceptual match between the stimulus shown and the evoked stimulus must be important in order to obtain a priming of the evoked stimulus.

To understand the difference between Experiments 1 and 2, recall that the transfer items were rated as more similar to their training twins in Experiment 2 than in Experiment 1. Thus, the absence of the BT-GT effect in Experiment 1 might have resulted from a lack of holistic similarity between training and transfer stimuli. The perceptual difference between training and transfer stimuli on the spots dimension resulted in a holistic difference that was sufficient to limit the influence of the remaining perceptual similarity on classification of the transfer stimuli. For the “spots” transfer stimuli, the spots give a texture replacing an empty area. This results in a transfer stimulus that appears holistically more dissimilar than an angular stimulus compared with a rounded stimulus. Moreover, the “no spots” stimuli in Regehr and Brooks (1993) had only one perceptual implementation and, thus, were perceptually more similar to the other “no spots” stimuli, including stimuli in the other category, than stimuli with spots. This lack of distinctiveness might have limited the role of exemplar similarity. The same reasoning holds for the stimuli in Experiment 3C. Stripes and dots are perceptually dissimilar textures, resulting in perceptually dissimilar training–transfer pairs. Moreover, in this experiment, we removed the external irrelevant features that contributed to the holistic individuation of the stimuli in Experiments 1 and 2. Again, a BT-GT effect is less likely with less distinctive stimuli than with distinctive stimuli, as already shown by Regehr and Brooks. By contrast, for the angular–rounded distinction, the resulting holistic perceptual difference between training and transfer items was minimal (angles were “smoothed” into a more rounded shape, and rounded shapes were slightly “squared”). In sum, we interpret our results as suggesting that for the effect to appear, the holistic similarity between training and transfer stimuli must be virtually perfect, a bit like the photograph-like images of objects postulated by template models (see Goldstone, 1998).

Our experiments also questioned the putative link between exemplar similarity and the notion of a *holistic* matching process of *concrete* information (Regehr & Brooks, 1993, p. 110; see also Brooks, 1978; Kemler Nelson, 1989; Pothos, 2005). We obtained the BT-GT effect with separated stimuli in Experiments 3A and 3B. This is surprising because the situation was one assumed to promote a strong analytic rule-based processing of the stimuli (i.e., sequential processing of the isolated parts; Smith et al., 1998).

In separated conditions, we think that the “close perceptual match” hypothesis also holds. We believe that in these conditions, the BT-GT effect is due to the identical features that are shared by each training item and its transfer twin. This means that for a BT item, by definition, the shared features were associated with a training twin belonging to the opposite category—that is to say, with the other category name. Thus, there was a perfect perceptual match between training and transfer items on two (Experiment 3B) or four (Experiment 3A) features, which was sufficient to obtain a BT-GT effect. Again, these results are consistent with the idea of a concrete matching process, as put forward by Smith et al. (1998) and Regehr and Brooks (1993); however, we describe it at the holistic level and/or at the featural level.

Last, we have to explain why we obtained a BT-GT effect with separated stimuli in Experiments 3A and 3B but not with the equivalent integrated stimuli in Experiments 1 and 3C. Our interpretation is connected with the notion of stimulus imprinting mentioned above. We believe that the two types of stimuli differ in the respective perceptual saliency of the featural and the holistic levels. For separated stimuli, the holistic level, if any, is not salient at all (see footnote 4) because the location of a feature changes from one trial to the next, whereas this level is very salient in integrated stimuli. This difference between the two types of stimuli is due to the training phase. With integrated stimuli, as training proceeds, the stimuli are recognized on the basis of the holistic features and less on the basis of individual features. At the end of the training phase, there is a strong perceptual imbalance in favor of holistic features, which become the main classification cues (in a similar way, Schyns & Rodet, 1997, showed that once participants have learned that a configural cue is relevant for classification, they do not use the individual cues this configuration is made of in their later classifications and, most likely, do not *see* the individual cues in the configural cue) (see also Goldstone, 1998; Kemler Nelson, 1989; Palmer, 1978; Regehr & Brooks, 1993, for a discussion). In terms of holistic features, the difference between Experiments 1 and 2 showed that the holistic similarity between training and transfer stimuli had to be virtually perfect in order to obtain the BT-GT effect. For separated stimuli, as training proceeds, the individual features are encoded in memory. The identical features that appear in the corresponding transfer stimuli are then recognized in the transfer phase. The perceptually identical features automatically evoke their training twins and the associated category name. To summarize, at the end of the training phase, we have different salient perceptual features for the two types of stimuli. Holistic features and, thus, holistic similarities win the contest for integrated stimuli, whereas individual features and featural similarities are more salient for separated stimuli. Note that this does not mean that featural similarities play no role in the case of integrated stimuli.

The BT-GT Effect and Rule-Defining Features

We also want to contrast our view with a proposal introduced very recently by Lacroix et al. (2005). In their interpretation of the results obtained by Regehr and Brooks (1993), they maintained that the BT-GT effect depends on the rule-defining attributes, selectively attended, and that the nonrule dimensions (i.e., tail length and neck length) do not contribute to the effect. This is in contrast with Regehr and Brooks, who argued that nonrule dimensions contributing to the holistic similarity between training and transfer stimuli played a role in the BT-GT effect. Lacroix et al., replicating Regehr and Brooks's Experiment 3B, showed that when the two nonrule dimensions of the training stimuli were replaced by perceptually different dimension values on their transfer twins, there was a BT-GT effect when the training stimuli were repeated 20 times, instead of 5 in Regehr and Brooks. With sufficient training, the BT-GT effect appears when training and transfer stimuli are similar at the level of rule-defining features and dissimilar at the level of nonrule features.

Unfortunately, in their current form, Lacroix et al.'s (2005) results are not conclusive because the authors did not remove the prototypes from their analyses and did not control the serial order parameter. Thibaut and Gelaes (2006) also varied the number of presentations. They replicated Regehr and Brooks's Experiment 3A with 30 presentations of the training set and found no BT-GT effect when the prototype advantage and the serial order parameters were controlled. Thus, we did not obtain the BT-GT effect with transfer stimuli that were more similar to their training twins than in Lacroix et al. As in Lacroix et al., the effect appeared when prototypes were included in the data set. The number of presentations (20 vs. 5) advocated by Lacroix et al. was not, in itself, the critical parameter explaining the BT-GT effect when training and transfer stimuli were not holistically similar. Thibaut and Gelaes's data do not fit with Lacroix et al.'s claim that the BT-GT effect depends solely on the defining attributes. By contrast, the absence of effect in Experiment 1 and its occurrence in Experiment 2 strongly suggest that the BT-GT effect requires a close perceptual match between training and transfer items, at least within the present experimental conditions.

To summarize, our results suggest that exemplar-based mechanisms are involved in a classification task in which a perfect classification rule is available. However, these exemplar-based mechanisms might not be restricted to holistically individuated integrated stimuli and, on the other hand, do not always appear with integrated stimuli such as the ones used by Brooks and colleagues. The emerging picture suggests that exemplar-based effects also incorporate similarities—identities—at the featural level (Brooks & Hannah, 2006). In our experiments, holistic and featural individuations, when they influenced classifications, required an important perceptual match between training and transfer items.

References

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

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–660.

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.

Brooks, L. R., & Hannah, S. D. (2006). Instantiated features and the use of "rules." *Journal of Experimental Psychology: General, 135,* 133–151.

Brooks, L. R., Norman, G. R., & Allen, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General, 120,* 278–287.

Cedrus Corporation. (1989). Superlab Pro (Version 1.5.9) [Computer software]. San Pedro, CA: Author.

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

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

Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition, 65,* 231–262.

Kemler Nelson, D. G. (1989). The nature and occurrence of holistic processing. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 357–386). Hillsdale, NJ: Erlbaum.

Lacroix, G. L., Giguère, G., & Larochelle, S. (2005). The origin of exemplar effects in rule-driven categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 272–288.

Lacroix, G. L., Hélie, S., & Larochelle, S. (2000, June–July). *Testing the unconscious influence of exemplar memory on conscious category decisions.* Poster presented at the meeting of the Association for the Scientific Study of Consciousness, Brussels, Belgium.

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

Modigliani, V. (1971). On the conservation of simple concepts: Generality of the affirmation rule. *Journal of Experimental Psychology, 87,* 234–240.

Modigliani, V., & Rizza, J. (1971). Conservation of simple concepts as a function of deletion of irrelevant attributes. *Journal of Experimental Psychology, 90,* 280–286.

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

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. L. Lloyd (Eds.), *Cognition and categorization* (pp. 259–302). Hillsdale, NJ: Erlbaum.

Pothos, E. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences, 28,* 1–14.

Regehr, G., & Brooks, L. R. (1993). Perceptual manifestations of an analytic structure: The priority of holistic individuation. *Journal of Experimental Psychology: General, 122,* 92–114.

Schacter, D. L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 233–268). Cambridge, MA: MIT Press.

Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition, 67,* 147–179.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21,* 1–53.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 681–696.

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

Smith, E. E., Patalano, A. L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65,* 167–196.

Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43,* 129–169.

Thibaut, J. P., & Gelaes, S. (2006). *Learning and automating a new categorization rule: Effect of similarity and practice.* Unpublished manuscript.

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

Received December 24, 2005
Revision received April 24, 2006
Accepted May 6, 2006 ■