ARTICLE IN PRESS

Cognition xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The development of category learning strategies: What makes the difference?

Rubi Hammer [a,b,*], Gil Diesendruck [d], Daphna Weinshall [a,c], Shaul Hochstein [a,b]

[a] The Interdisciplinary Center for Neural Computation, Hebrew University, Jerusalem, Israel
[b] Neurobiology Department, Institute of Life Sciences, Hebrew University, Jerusalem, Israel
[c] School of Computer Sciences and Engineering, Hebrew University, Jerusalem, Israel
[d] Gonda Brain Research Center and Department of Psychology, Bar-Ilan University, Ramat-Gan, Israel

ARTICLE INFO

Article history:
Received 17 January 2008
Revised 26 March 2009
Accepted 31 March 2009
Available online xxxx

Keywords:
Category learning
Categorization
Perceived similarity
Cognitive development
Conceptual development
Comparison
Logical inference
Structural alignment
Learning to learn
Concept hierarchy
Taxonomic organization

ABSTRACT

Category learning can be achieved by identifying common features among category members, distinctive features among non-members, or both. These processes are psychologically and computationally distinct, and may have implications for the acquisition of categories at different hierarchical levels. The present study examines an account of children's difficulty in acquiring categories at the subordinate level grounded on these distinct comparison processes. Adults and children performed category learning tasks in which they were exposed either to pairs of objects from the same novel category or pairs of objects from different categories. The objects were designed so that for each category learning task, two features determined category membership whereas two other features were task irrelevant. In the learning stage participants compared pairs of objects noted to be either from the same category or from different categories. Object pairs were chosen so that the objective amount of information provided to the participants was identical in the two learning conditions. We found that when presented only with object pairs noted to be from the same category, young children ($6 \leqslant YO \leqslant 9.5$) learned the novel categories just as well as older children ($10 \leqslant YO \leqslant 14$) and adults. However, when presented only with object pairs known to be from different categories, unlike older children and adults, young children failed to learn the novel categories. We discuss cognitive and computational factors that may give rise to this comparison bias, as well as its expected outcomes.

1. Introduction

Categorization enables generalization from a few experiences to novel conditions while reducing dramatically the computational complexity of perceived objects or events. For categories to have this capability, objects within the same category must share some attributes, and at the same time they need to differ on other attributes from objects belonging to different categories. Thus, logically, category learning can be achieved either by identifying the attributes shared by some exemplars[1] known to be from the same category, by identifying the attributes discriminating some exemplars known to be from different categories, or – perhaps most probably – by both identifications.

Crucially, however, a number of researchers have argued that psychologically, the processing and computation of similarities and differences are not necessarily equally usable. For instance, Tversky (1977) noted that the weight of exemplars' common features (i.e., similarities) vs. distinctive features (differences) varies across tasks, and Medin, Goldstone, and Gentner (1990) concluded that the type

* Corresponding author. Address: The Interdisciplinary Center for Neural Computation, Edmond Safra Campus, Hebrew University, Givat Ram, Jerusalem 91904, Israel. Tel.: +972 54 653 1642.
E-mail address: rubih@alice.nc.huji.ac.il (R. Hammer).

[1] The term exemplar refers to an actual category member, not the mental representation of one.

0010-0277/$ - see front matter © 2009 Published by Elsevier B.V.
doi:10.1016/j.cognition.2009.03.012

of features on which adults focus varies between tasks which require attention to similarities vs. differences. More recently, a number of researchers suggested domain differences due to the processing of correlated vs. distinctive features. For instance, Randall and colleagues showed that while adults' capacity to process correlated features (i.e., similarities) among category exemplars does not depend on category domain (e.g., living vs. non-living), the capacity to process distinctive features does (Randall, Moss, Rodd, Greer, & Tyler, 2004). Specifically, the semantic representations of living compared to non-living exemplars, are more slowly activated by their respective distinctive properties. Similarly, Cree and McRae (2003) suggested that while correlated features are at the core of all concept representations, the usability of distinctive features differs dramatically across domains. These accounts imply that "domain-specific deficits" do not necessarily derive from poor representation of a specific category domain, but rather from a more general computational malfunction; namely, an inability to process distinctive features, exacerbated in the case of living things due to their typical greater visual complexity (Randall et al., 2004).

An additional central aspect of human categorical knowledge that may also be susceptible to the differential processing of similarities and differences, is the hierarchical organization of categories. Objects or events are not only grouped into categories, but categories are also organized in an inclusive structure, with different levels of abstraction (Murphy, 2002; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In particular, highly abstract – superordinate – categories (e.g., "furniture", "animals"), comprise more specific – basic-level – categories (e.g., "chair", "dog"), which in turn include even more specific – subordinate – categories (e.g., "rocking chair", "poodle").

It has been suggested that the hierarchical representation of concepts results from the "objective" structure of object categories. Specifically, it is argued that basic-level categories can be easily learned because, on the one hand, they are quite homogenous, while on the other, they are fairly distinct from each other (Malt, 1995; Rosch et al., 1976). The representation of superordinate-level categories might be relatively more complex since they are poorly homogenous. Finally, more specific, subordinate-level categories may be more difficult to acquire since, despite their being even more homogenous than basic-level categories, they are not as distinct from other subordinate-level categories associated with the same basic-level category (Markman & Wisniewski, 1997; Murphy & Brownell, 1985).

The above description implies differences in the modal computations of similarities and differences typically required for acquiring categories from the different hierarchical levels (see for instance, Markman & Wisniewski, 1997). Specifically, basic-level categories may require a learner to identify both within-category similarities and between-category differences. In fact, one may argue that in many cases, the former kind of computation will suffice to give rise to basic-level categories, because of the high-degree of between category distinctiveness, together with only fairly good within category cohesiveness at this level. Categories at the superordinate-level, in turn, demand that the learner ignore somewhat striking within-category exemplar differences, focusing instead on relatively abstract similarities correlated with a few perceptual similarities. Finally, subordinate-level categories, in which there is high similarity both within-category and between-categories, place the heaviest weight on detecting subtle yet important differences between exemplars from different categories.

The above analysis suggests that when deciding whether two items belong to the same category, the kinds of computations that learners have to undertake vary according to the level of abstraction at which the learner is categorizing. Thus, the computations regarding between-objects similarities and differences that are available to or favored by a learner, should impact the level of abstraction at which the learner is proficient at categorizing items. Or putting it differently, the level of abstraction at which learners are proficient at categorizing, may be indicative of, or result from, the kinds of computations they are capable of performing.

Developmental psychologists have for a long time been interested in children's acquisition of a hierarchical organization of categories. One view argues that the first categories acquired by children are at the basic-level (Brown, 1958; Malt, 1995; Rosch et al., 1976) – e.g., children prefer labeling an animal as a "dog", rather than referring to it with a superordinate ("animal") or subordinate ("poodle") label. Lately, a number of researchers have suggested that both perceptually (Quinn, 2004; Younger & Fearing, 2000) and conceptually (Keil, 2008; Mandler, 2008), children start off with broad categories, and gradually move down to more specific levels of abstraction. For example, Quinn and Johnson (2000) showed that 2-month-old infants are capable of discriminating between mixtures of different mammals and furniture, but not between cats and a mixture of other basic-level categories of mammals including elephants, rabbits, and dogs. While these accounts seem to disagree about the developmental starting point, they concur that subordinate-level categories are the last to be acquired (see also Furrer & Younger, 2005; Horton & Markman, 1980; Mervis & Crisafi, 1982; Waxman, Lynch, Casey, & Baer, 1997, for supportive findings).

The goal of the present study is to test an account of this developmental phenomenon, grounded on processes of computing similarities and differences in the context of exemplar comparison. The hypothesis guiding the present study is that if category learning at the various hierarchical levels is related to the differences noted above in the computation of similarities and differences, then the late emergence of subordinate level categories in children may be indicative of a difficulty in the computational processes required for learning subordinate categories – namely, identifying or computing between-category differences.

For the last two decades, scholars have emphasized the importance of comparison processes for category learning (e.g., Gentner & Markman, 1994; Gentner & Namy, 2006; Kurtz & Boukrina, 2004; Markman & Gentner, 1993; Namy & Gentner, 2002; Spalding & Ross, 1994). One important conclusion deriving from these studies is that comparison may differentially stress similarities and differences between compared items. For instance, in their studies on the role of structural alignment and comparison, Markman

Please cite this article in press as: Hammer, R., et al. The development of category learning strategies: What makes the difference?. *Cognition* (2009), doi:10.1016/j.cognition.2009.03.012

and Gentner (1993) showed that when comparing pairs of similar words (i.e., words representing similar concepts), adults were capable of listing more similarities than when comparing pairs of dissimilar words. Curiously, the reverse was not true – when asked to list differences, subjects listed more differences for the compared similar pairs than for the dissimilar pairs. Furthermore, differences were specified mostly when they could be aligned (e.g., having two legs vs. having four legs). When differences could not be aligned (e.g., having wings vs. having horns), they were more likely to be ignored (Gentner & Markman, 1994). Consistent with these ideas, Boroditsky (2007) found that comparison of two objects highlighted to adults the similarities between the objects, even when participants were encouraged to address the differences between them. This comparison bias increased the perceived similarity between objects. In their review of this literature, Doumas, Hummel, and Sandhofer (2008) suggested that when two objects are compared, similar properties are represented twice, and as a result similarities receive twice the input as do differences. They concluded that this may give rise to the reported "attention bias", in which similarities overshadow differences.

Shifting to developmental studies, this comparison bias seems to be even more evident: Gentner and Namy (1999) found that comparing two perceptually similar category members increased 4-year-olds' tendency to categorize the objects taxonomically (rather than thematically, for instance). Furthermore, they showed that providing children with a common label for objects encouraged comparison, whereas providing conflicting labels deterred it (Namy & Gentner, 2002). Findings with 12-month-olds suggest that this comparison bias is present already at the earliest stages of word learning (Waxman & Braun, 2005). As a number of developmental researchers have concluded, a common label seems to foster children's acquisition of a category because it implies that commonalities among the referents of the label must exist (Gentner & Namy, 2006; Waxman & Lidz, 2006).

The findings described above with adults, and especially with children, intimate that in the process of comparing objects, similarities among category members eclipse differences. The goal of the present study is to test the inverse implication, namely, how this comparison bias favoring the processing of similarities may affect *category learning*. To recap, for global categories, category learning by comparing objects that share the *same* label or function (i.e., objects from the same category) can be appropriate and perhaps even sufficient, since the main challenge here is to identify the few features that are *common* to category members (within-category similarities). In contrast, for learning highly specific categories, comparing objects with *different* labels or functions (i.e., objects from different categories) may have greater value, since there are not only many common features within each subordinate-level category, rather there is also high between-category similarity. Since this is the case, the challenge is to identify the few features that *distinguish* between members of different categories.

The implication of the above analysis of the processing of similarities and differences is that, while superficially, *same-class exemplars* comparison (i.e., comparing objects from the same category) and *different-class exemplars* comparison (i.e., comparing objects from different categories) seem to be equally useful for category learning, these two comparison processes differ fundamentally. In particular, comparing same-class exemplars is useful for highlighting possibly informative *within-category similarities*, and informative within-category differences. In turn, comparing different-class exemplars is useful for highlighting possibly informative *between-category differences*, and uninformative between-category similarities. That is, same-class and different-class comparisons may both involve the processing of similarities and differences, but for each comparison type, similarities and differences have different meanings in category learning. If as intimated by the literature, the processing of similarities is cognitively favored and available developmentally earlier than the processing of differences, then children might have an easier time acquiring categories via comparison of same-class exemplars, than via comparison of different-class exemplars. Given the analyses described earlier about the differences in the kinds of computations involved in categorization at different hierarchical levels, support for this hypothesis would partly explain why categories at hierarchical levels whose acquisition rests heavily on the processing of differences – namely, subordinate-level categories – pose more difficulty for acquisition than categories that rest heavily on the processing of similarities.

To the best of our knowledge, no developmental study has systematically investigated the differential contributions of comparison of same-class exemplars vs. comparison of different-class exemplars for category learning (but see Hammer, Bar-Hillel, Hertz, Weinshall, & Hochstein, 2008; Hammer, Hertz, Hochstein, & Weinshall, 2005, 2007, 2009, for computer simulations and findings with adults). The current study is designed for this purpose, testing both children and adults. Unlike previous studies on comparison processes, we systematically dissociate the two comparison types. Furthermore, we test the process of category *learning* by comparison, rather than how comparison is used when referring to already *familiar* categories.

In order to ensure that the hypothesized condition differences would most likely result from the operation of a cognitive bias, rather than category-specific prior knowledge, perceptual properties of the objects, or other aspects associated with the pairing of the compared objects, we undertook a series of methodological precautions: First, we equated the objective amount of information provided to the participants in the two conditions (see description in Section 2). Second, in the two experimental conditions participants learned the same categorization rules, one group by same-class comparison and the other group by different-class comparison. Participants in the two conditions were then tested on exactly the same task, using exactly the same stimuli. Third, we used novel stimuli, and counterbalanced the common vs. distinctive features, so as to exclude any possible interference of previous domain-specific knowledge or feature salience. Furthermore, we also encouraged participants to attend to both the similarities and the differences between the compared stimuli during the learning stage in both conditions. Finally, in order to exclude the possibility that the comparison bias is associ-

Table 1
Mean and standard deviation (*SD*) of participants' ages in the different experimental conditions and age groups.

Condition/age-group	$6 \leqslant$ Age $\leqslant 9.5$	$10 \leqslant$ Age $\leqslant 14$	Adults
Same-class exemplars	$M = 7.70$	$M = 11.10$	$M = 24.71$
	$SD = 1.06$	$SD = 1.29$	$SD = 5.15$
	$n = 10$	$n = 10$	$n = 20$
Different-class exemplars	$M = 7.55$	$M = 11.20$	$M = 25.45$
	$SD = 1.07$	$SD = 1.48$	$SD = 3.61$
	$n = 10$	$n = 10$	$n = 20$

ated with the processing of labels, we also avoided the use of labels as category relation identifiers (see Gelman & Waxman, 2007; Sloutsky, Kloos, & Fisher, 2007a, 2007b; Sloutsky, Lo, & Fisher, 2001, for a discussion on the possible developmental changes in label processing).

Based on the adult, but especially the developmental literature suggesting an advantage of comparison by similarity over comparison by difference, we expected early elementary-school aged children to be better at learning categories from comparison of similar-class exemplars than from different-class exemplars, but older children and adults to perform equivalently in these two conditions. Alternatively, if under such controlled conditions, young children would show similar proficiencies in learning from different-class exemplar comparison as from same-class exemplar comparison, then this may suggest that the previously reported comparison bias is more specific. That is, it might be reasonable to conclude that the bias results from other factors associated with everyday life learning conditions of natural categories rather than from a general cognitive bias as we postulate. Children's age range was selected so as to maximize the possibility of observing variance in performance across development, while at the same time ensuring that children were mature enough to complete the categorization tasks with minimal intervention by the experimenter.

2. Methods

2.1. Participants

Forty adults ($19 \leqslant$ years $\leqslant 36$), 20 early elementary school aged children ($6 \leqslant$ years $\leqslant 9.5$), and 20 older children ($10 \leqslant$ years $\leqslant 14$), participated in the experiment (similar numbers of males and females). In the statistical analysis we refer to participants' age using both a rational scale and a categorical scale in which children's age categories are determined by their median age. We obtained written consent from adult participants and parental consent for participating children. See Table 1 for further information on participants' ages.

2.2. Materials

Five sets of computer generated color images of "alien creatures" were used as stimuli. Each set was characterized by four binary feature-dimensions that could differ and determine creature categories within the given set (that is, for each set, 16 creatures were created). Stimuli were designed so that the differences between the creatures in all varying feature-dimensions were highly distinctive (see Fig. 1).

The experiment was conducted using a laptop computer with a 15-in. screen, set to a resolution of 1280×1024 pixels. Stimulus presentation was done using software specially designed for the experiment. In each experimental trial, a pair of stimuli was simultaneously presented in the center of the computer screen. Each stimulus occupied 320×320 pixels, and the two stimuli were separated by a gap of 320 pixels. Both children and adults responded directly, using the two keys of a mini-sized computer mouse. The left key was marked with a green[2] smiley sticker, and the right key was marked with a red smiley sticker.

2.3. Design and procedure

Participants in each age group were randomly assigned to one of two experimental conditions – the learning from *same-class exemplars* condition and the learning from *different-class exemplars* condition. There were no significant differences in terms of mean ages of participants between the two experimental conditions in each of the age groups (all $p > .3$; see Table 1).

The experimental task was a simple same/different task in which participants decided whether two simultaneously presented creatures were of the same creature kind, or of two different kinds. Each participant performed the category learning task for five stimulus sets (identical tasks for children and adults). Each categorization task had three blocks: the *pre-learning test block* consisted of eight test trials, the *learning block* consisted of four learning trials, and the *post-learning test block* consisted of eight additional test trials. In each one of the test blocks, half of the test trials presented pairs of creatures of the same kind (identical in the two pre-selected relevant dimensions), and half were of different kinds (different in one of the pre-selected relevant dimensions, and one of the irrelevant dimensions). The overall similarity between the paired creatures, with respect to the varying features, was always the same whether the two creatures were of the same kind or from different kinds, making a strategy based on overall similarity judgment inadequate. Taken together, in each one of the test blocks there were four trials in which the paired creatures were identical in the two relevant features, two trials in which they differed in the first relevant and one of the irrelevant features, and two trials in which they dif-

[2] For interpretation to color in Figs. 1–6, the reader is referred to the web version of this article.

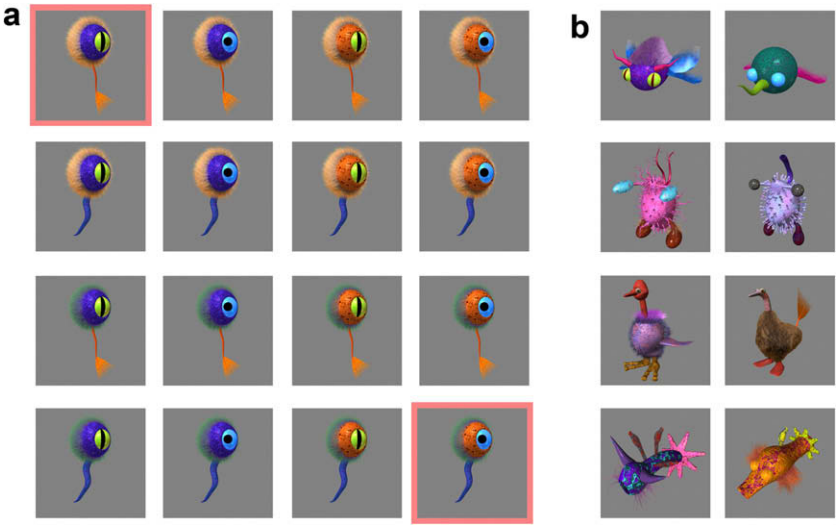


Fig. 1. (a) The stimuli created for set 1. Each creature can be identified by the unique combination of its eye, skin color, fur color, and tail. Marked with pink frames are two orthogonal exemplars, i.e., two creatures that differ in all four dimensions. (b) Two "orthogonal" exemplars from sets 2–5 (from top to bottom).

fered in the second relevant and one of the irrelevant features (see more details below). We used identical stimulus pairs for the test blocks in the two experimental conditions. Participants concluded the three-block categorization task for one stimulus set, and then moved on to the next set.

Before starting the experiment, participants were told that they were going to play a game in which they would learn about different creatures living on a remote planet. Participants were further instructed that they would have to decide whether each two creatures presented together are of the same kind (pressing the left mouse key) or two different kinds (pressing the right mouse key). Participants were then told that, "Creatures of the same kind do not need to be identical, as two different dogs are not totally identical although they are of the same kind. Similarly, two creatures from two different kinds do not have to be totally dissimilar, as a dog and a cat also share many properties although not being of the same kind".

Participants in the *same-class exemplars* condition were then instructed that when two creatures appear inside a green frame, it means that the two creatures are necessarily of the same kind. Similarly, participants in the *different-class exemplars* condition were instructed that when two creatures appear inside two separate red frames, it means that the two creatures are necessarily of different kinds. Participants were further instructed that when such a "clue" is provided, they should respond by pressing the left/right key. In addition, they were told to look for both similarities and differences in order to try and identify what is important to know about these creatures, and so as to decide later whether other paired creatures are of the same kind or not.

Following the instructions, participants performed one warm-up categorization task that was similar to the experimental tasks but had no time limit. While performing the warm-up task, the experimenter repeated the instructions to ensure that the participant knew which keys to use for "same" vs. "different" responses, and that the participant understood the meaning of the clues. After performing the warm-up task, the participant started the experimental task with the first stimulus set without further intervention by the experimenter (except the verbal encouragement given to children at the end of each category learning task). *No feedback* was provided for error or success.

Pre-learning test block: In the *pre-learning test block* of the experiment, trial duration was four seconds, and participants had to respond within this period of time. The time interval between trials was half a second. In each of the test trials, the two presented creatures were identical in exactly two out of the four possible feature-dimensions, and differed in the other two. Thus, the amount of similarity vs. dissimilarity with respect to the four varying features was roughly balanced, reducing possible response bias. To further reduce the possibility of response bias, in each experimental condition there were actually two sub-conditions, which differed in the selected relevant feature-dimensions. The *pre-learning test block* provided an indication of participants' baseline performance, enabling an estimation of the contribution of unsupervised learning to performance of the task. It also allowed participants to become familiarized with the particular dimensions in which features varied for a given set. This phase of the experiment was identical in the two conditions.

Learning block: The only difference between the two experimental conditions was in the *learning block*, which included different stimulus pairs within the colored frames indicating their relation (same kind/different kinds). At the beginning of each learning block, a slide stating "be prepared for the clues" indicated to the participant the beginning of the learning phase (it was also verbally announced by the experimenter). After this slide disappeared, four pairs of creatures appeared, one after the other, each with a designating "clue", as follows; in the *same-class exemplars*

condition the two creatures appeared inside a green frame indicating that these two creatures are of the same kind. In the *different-class exemplars* condition, the two creatures appeared in two separate red frames indicating that these two creatures are of two different kinds. Each pair of creatures differed by only one irrelevant feature in the *same-class exemplars* condition, or by only one relevant feature in the *different-class exemplars* condition. This pairing enabled a decisive indication concerning the irrelevance or relevance of the specified dimension, respectively. Although the relation between the two presented creatures in the learning block was obvious, in order to verify that participants understood the clues, they identified the categorical relation between the creatures by pressing the relevant mouse key. Each one of the four trials in the learning phase lasted 6 s (separated by a half second interval).

The quantity of information provided to the participants in the *learning block* was equalized and maximized (given the limitations of each one of the comparison type) between the two experimental conditions. Fig. 2a illustrates an example of the hypothesis space that the participants could have worked out while performing the *pre-learning test block* for the specific experimental tasks. The table presents the four different binary feature-dimensions in which this set of creatures varied: tail, eye, fur, and skin dimensions. H1–H16 represent the $2^4 = 16$ possible combinations for irrelevant (marked with "0") and relevant (marked with "1") dimensions. At one extreme, Hypothesis 1 suggests that no dimension is relevant for categorization – that is, all creatures can be treated as if they are from the same category. At the other extreme, Hypothesis 16 suggests that all dimensions are relevant for categorization – that is, each creature can be treated as if it is from a different category. For this example, the two feature-dimensions selected to be relevant for categorizing these particular creatures are the tail and the eye dimensions. The partici-

a

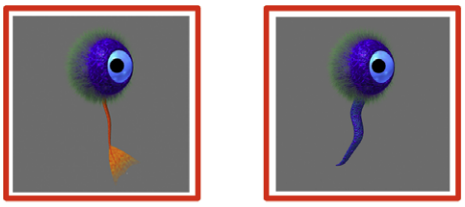
	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
Tail	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Eye	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Fur	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Skin	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

b

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
Tail	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Eye	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Fur	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Skin	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
Tail	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Eye	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Fur	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Skin	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

c



	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
Tail	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Eye	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Fur	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Skin	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
Tail	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Eye	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Fur	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Skin	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

Fig. 2. An illustration of the measures that were taken for equalizing the information quantity in the two learning-by-comparison conditions. (a) A table illustrating the initial hypothesis space (16 possible hypotheses): all the possible combinations for relevant dimensions (marked as "1") and irrelevant dimensions (marked as "0"). (b) Two same-class exemplars indications (paired creatures in a green frame). The table below the upper same-class exemplars represents the remaining hypotheses after being provided with the same-class indication suggesting that fur color is irrelevant (H1, H2, H5, H6, H9, H10, H13, and H14). The table below the lower same-class exemplars represents the remaining hypotheses after being *also* provided with the indication that skin color is irrelevant (H1, H5, H9, and H13). (c) Two different-class exemplars indications (paired creatures in two red frames). The table below the upper different-class exemplars represents the remaining hypotheses after being provided with the indication that the tail is relevant (H9–H16). The table below the lower different-class exemplars represents the remaining hypotheses after being *also* provided with the indication that the eye is relevant (H13–H16).

Please cite this article in press as: Hammer, R., et al. The development of category learning strategies: What makes the difference?. *Cognition* (2009), doi:10.1016/j.cognition.2009.03.012

pants were asked to deduce this when provided with either same-class or different-class indications.

Fig. 2b illustrates the learning in the *same-class exemplars* condition. The upper two exemplars are from the same-class (as indicated to the participants by the single green frame) and they differ only in their fur color. This same-class "clue" indicates that fur color is not relevant for categorization since the within category variation in this dimension is similar to its overall variation. This eliminates all the hypotheses in which fur color is relevant (the hypotheses marked in green in the table below the stimulus pair). By reducing the hypothesis space by half, this same-class indication provides $-\log_2 8/16 = 1$ bit of information. The same-class exemplars of the lower pair differ only in their skin color. This same-class clue indicates that skin color is also not relevant for categorization. This eliminates all the remaining hypotheses in which skin color is relevant leaving the participants with the four hypotheses in which both fur and skin color are not relevant for categorization. This same-class indication also provides $-\log_2 4/8 = 1$ bit of information. Taken together, the two same-class exemplars indication provided 2 bits of information by eliminating all the hypotheses in which either one of the irrelevant dimensions is marked as relevant (leaving only H1, H5, H9, and H13).

Additional same-class exemplar indications cannot provide any further information since all the irrelevant features are already specified. Nevertheless, in each category learning task we provided four same-class indications (using four different pairs) so that each irrelevant feature was specified twice. This was done to ensure that participants had sufficient opportunity to identify the task relevant features.

Fig. 2c illustrates the learning in the *different-class exemplars* condition. The upper two exemplars are from different-classes (as indicated to the participants by the two red frames) and they differ only in their tails. This different-class "clue" indicates that the tail is relevant for categorization since this is the only feature discriminating two creatures from two different kinds. This eliminates all the hypotheses in which tails are irrelevant (marked in red in the table below the stimulus pair). By reducing the hypothesis space by half, this different-class clue provides $-\log_2 8/16 = 1$ bit of information. The lower two different-class exemplars differ only in their eyes. This different-class indication provides an additional $-\log_2 4/8 = 1$ bit of information by suggesting that eyes are also relevant for categorization, leaving only the four hypotheses in which both tails and eyes are relevant (H13–H16). Additional different-class exemplar clues cannot provide any further information since all the relevant features are already specified. Here we also provided four different-class indications (using four different pairs) so that each relevant feature was specified twice.

To sum up, the quantity of information of each of the same-class or different-class clues that were used is 1 bit, and for each task in either condition participants received a total of 2 bits of information (and received them twice, redundantly) for each category learning task with each creature set. Nevertheless, it is also obvious that even

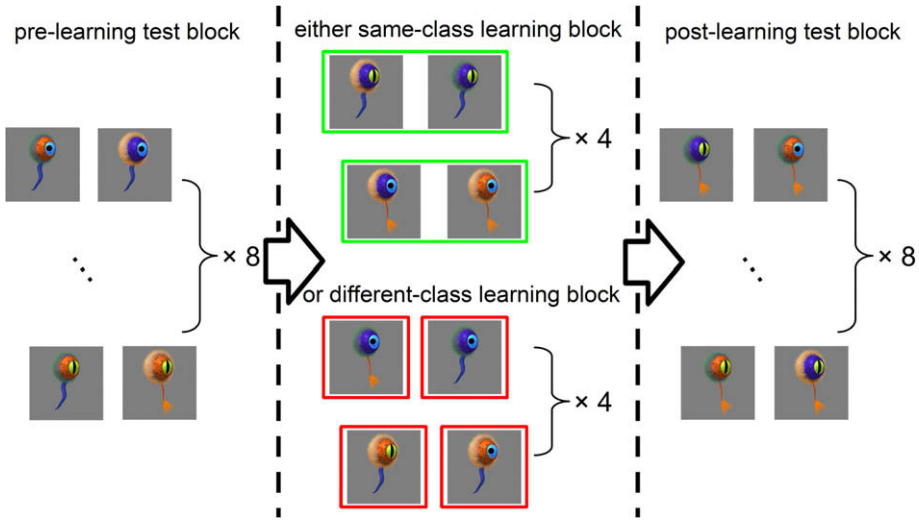


Fig. 3. Schematic illustration of the categorization task for one set. The task-relevant dimensions are tails and eyes. In the *pre-learning test block*, participants could only guess whether the two presented creatures were from the same kind. Nevertheless, this test phase enabled participants to identify the *potentially* relevant features for this set (features in which these creatures can differ). In the *same-class exemplars learning block*, participants could learn from the first clue – that the two presented creatures are of the same kind despite differing in fur color – that fur color is not relevant for category membership. From the second clue they could have learned that the creatures' skin color is also irrelevant. The two additional clues provided the same insights concerning the creatures' fur and skin colors (using different stimulus pairs), leaving the creatures' eyes and tails as the only possible relevant features for establishing category membership. From the first clue in the *different-class exemplars learning block*, participants could learn that the creatures' tail is important since it is the only feature discriminating two creatures noted to be of different kinds. Similarly, from the second clue participants could have learned about the importance of the creatures' eyes. In the *post-learning test block*, participants had to perform the task according to what they had just learned in the learning block. In the examples illustrated here, for the upper pair participants should have responded that the two are not of the same kind since they differ in their eyes. For the lower pair participants should have responded that the two are of the same kind since they have identical eyes and tails. Note that the creature pairs used in the test phases were identical for both conditions.

Please cite this article in press as: Hammer, R., et al. The development of category learning strategies: What makes the difference?. *Cognition* (2009), doi:10.1016/j.cognition.2009.03.012

Table 2
Number of participants excluded from the analysis in each age group and experimental condition.

Condition/age-group	$6 \leqslant$ Age $\leqslant 9.5$	$10 \leqslant$ Age $\leqslant 14$	Adults
Same-class exemplars	3	1	5
Different-class exemplars	2	1	3

when optimally used, each type of clue leaves a few alternative hypotheses in addition to the correct one (the "true hypothesis" is H13 in which the tail and eye dimensions are both specified as relevant, and the fur and skin colors are specified as irrelevant). Alternative not-disproved hypotheses either exclude also tail and/or eye as irrelevant (for the same-class exemplars condition) or include fur and/or skin color as also relevant (for the different-class exemplars condition).

Post-learning test block: Immediately after the learning block, the *post-learning test block* started. This test block was identical for the two conditions, and was similar in format and stimuli to the *pre-learning test block*. In this phase, however, participants were instructed to make their decisions according to what they had learned during the learning block. After the categorization task for one set was completed, there was a five second interval before the task with the next stimulus set started. Fig. 3 presents a schematic illustration of the experimental task paradigm.

3. Results

Our main hypothesis was that there would be no significant effect of *condition* among older children and adults, but there would be a significant effect among younger children in favor of the *same-class exemplars* condition. Complementarily, the hypothesis was that there would be no effect of *age* on performance in the *same-class exemplars* condition, but that there would be such an effect in the *different-class exemplars* condition. We measured participants' ability to learn the new categories by using the non-parametric sensitivity measure A' (Grier, 1971), calculated from participants' hits (correctly identifying two creatures as belonging to the same category) and false-alarms (incorrectly identifying two creatures as belonging to the same category). $A' = 0.5$ represents chance performance, $A' = 1$ represents perfect performance, and $0 < A' < 0.5$ represents response confusion. For each participant we calculated his or her average performance in all five sets. Participants with $A' < 0.5$ or more than 12% of missed trials in the *post-learning test block* were excluded from the analysis (see Table 2). For the analysis, we used as dependent measures both participants' A' in the *post-learning test block* (denoted as post-A'), and the difference between this and their measured A' in the *pre-learning test block* (post-A' minus pre-A'). The latter measure (denoted as A'-difference) is intended to "filter out" participants' guessing strategy.

In order to evaluate the effect of age on sensitivity, we first calculated the Pearson correlation between participants' age (in years) and their post-A' score, for each experimental condition separately. In order to reduce the relative weight of age differences among adults (which is less relevant), we calculated the correlations between performance and the natural log of age. We found no significant correlation between *ln* age and post-A' in the *same-class exemplars* condition, $r(38) = .12$, $p = .47$, but the correlation between *ln* age and post-A' in the *different-class exemplars* condition was highly significant, $r(38) = .66$, $p < .0001$. This result supports our hypothesis that the capacity to learn from different-class exemplars develops with age, whereas the capacity to learn from same-class exemplars is available even for young children.

An ANOVA with post-A' as dependent variable, age group (young children, older children, and adults), and experimental condition (*same-class exemplars* vs. *different-class exemplars*) as between-subject factors, revealed no main effect of condition, $F(2, 74) = .53$, $p = .47$, but a significant effect of age group, $F(2, 74) = 10.74$, $p < .0001$, $\eta_p^2 = .23$. Importantly, there was a significant interaction between condition and age, $F(2, 74) = 4.38$, $p < .02$, $\eta_p^2 = .11$ (see Fig. 4 – left). Independent samples *t*-tests on the effect of condition within each age group showed that young children's post-A' score was significantly higher when they were trained with same-class exemplars ($M = .74$; $SD = .12$) than when they were trained with different-class exemplars ($M = .62$; $SD = .12$), $t(18) = 2.24$, $p < .05$, $d = 1.05$. Older children performance when they were trained with same-class exemplars ($M = .81$; $SD = .14$) was not significantly different from their performance when trained with different-class exemplars ($M = .80$; $SD = .11$), $t(18) = .26$, $p = .80$, and adults' performance was somewhat better when they were trained with different-class exemplars ($M = .87$; $SD = .09$) than when they were trained with same-class exemplars ($M = .80$; $SD = .14$), $t(38) = -1.99$, $p = .054$, $d = .65$, though this difference was not statistically significant.

Moreover, one-way ANOVAs showed a significant effect of age on the post-A' score only in the *different-class exemplars* condition $F(2, 37) = 18.39$, $p < .001$, but not in the *same-class exemplars* condition $F(2, 37) = .79$. Post-hoc Scheffe tests showed that in the *different-class exemplars* condition, young children's performance ($M = .62$; $SD = .12$) was significantly lower than that of older children ($M = .80$; $SD = .11$) and adults ($M = .87$; $SD = .09$) ($p < .005$ in both cases). There was no significant difference in the post-A' score between older children and adults, $p = .23$.

Similarly, an ANOVA with A'-difference as the dependent variable revealed no main effect of condition, $F(2, 74) = .89$, $p = .35$, but a significant effect of age, $F(2, 74) = 5.18$, $p < .01$, $\eta_p^2 = .12$. Again there was a significant interaction between condition and age, $F(2, 74) = 7.47$, $p < .002$, $\eta_p^2 = .17$ (see Fig. 4 – right). *t*-Tests on the effect of condition within each age group showed that for young children, improvement was significantly greater when they were trained with same-class exemplars ($M = .32$; $SD = .16$) than with different-class exemplars ($M = .12$; $SD = .14$), $t(18) = 2.86$, $p < .02$, $d = 1.35$. An opposite pattern was found in the other age groups, such that improvement was greater when provided with different-class exemplars than when provided with same-class exemplars – though only for adults the condition effect was statistically signif-

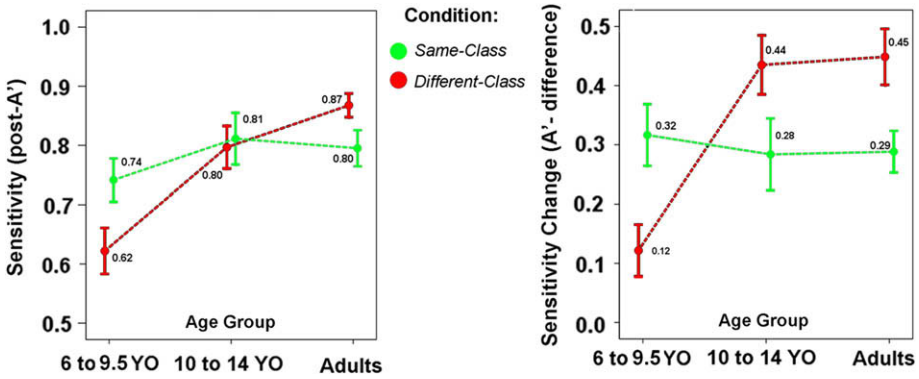


Fig. 4. *Left:* mean (and standard error of) sensitivity A' following training, i.e., post-A', by condition and age. *Right:* mean (and standard error of) sensitivity change from pre-training to post-training test, i.e., A'-difference, by condition and age. Note the superior sensitivity and greater sensitivity change for same-class exemplars in young children, and the opposite effect in adults (as well as for older children in regard to sensitivity change). Note also the increase in both measures with age for different-class exemplars and nearly no dependence on age for same-class exemplars.

icant – older children, $t(18) = -1.92$, $p = .071$, $d = .91$; adults, $t(38) = -2.71$, $p < .02$, $d = .88$.

One-way ANOVAs showed a significant effect of age on the A'-difference only in the *different-class exemplars* condition $F(2, 37) = 11.54$, $p < .001$, but not in the *same-class exemplars* condition $F(2, 37) = .12$. Post-hoc Scheffe tests showed that in the *different-class exemplars* condition, young children's improvement ($M = .12$; $SD = .14$) was significantly smaller than that of older children ($M = .43$; $SD = .16$) and adults ($M = .45$; $SD = .21$) ($p < .005$ in both cases). There was no significant difference in the A'-difference between older children and adults, $p = .98$.

Taken together, the above analyses show that for young children, learning from same-class exemplars is more effective than learning from different-class exemplars. For older children, and especially for adults, the exact opposite is the case.

3.1. Participants' response-bias

In order to evaluate the effect of learning condition (*same-class* vs. *different-class*) on the response bias in each age group, we analyzed the changes in participants' hit and false-alarm rates separately. More specifically, we analyzed the difference between participants' hit rate in *the post-learning test block* and participants' hit rate in the *pre-learning test block* (denoted as hit-difference). Positive values represent improvement in performance (increase in the hit rate after learning; hit-difference = 0 represent no improvement). Similarly we analyzed the difference in participants' false-alarm rate (denoted as FA-difference). Negative values represent improvement in performance (reduction in the false-alarm rate after learning; FA-difference = 0 represent no improvement). Fig. 5 illustrates participants' false-alarm and hit rates plotted on a receiver operating characteristics (ROC) diagram.

An ANOVA with FA-difference as the dependent variable revealed both a significant effect of condition, $F(2, 74) = 9.23$, $p < .005$, $\eta_p^2 = .11$ (a larger reduction in the false-alarm rate in the *different-class exemplars* condition than in the *same-class exemplars* condition), and a significant effect of age, $F(2, 74) = 7.50$, $p < .002$, $\eta_p^2 = .17$, but no significant interaction between condition and age, $F(2, 74) = .37$, $p = .69$. One-way ANOVA with Post-hoc Scheffe tests showed that this latter main effect results from the lack of significant reduction in young children's false alarm rate in both experimental conditions ($M = -.06$; $SD = .16$), as compared to adults' false-alarm reduction ($M = -.24$; $SD = .17$), $p < .05$.

An ANOVA with hit-difference as the dependent variable revealed no effect of condition, $F(2, 74) = 2.93$, $p = .09$, but a significant effect of age, $F(2, 74) = 3.55$, $p < .05$, $\eta_p^2 = .09$. More importantly, there was a significant interaction between condition and age, $F(2, 74) = 8.22$, $p < .001$, $\eta_p^2 = .18$ (see Table 3 for means). Further investigation of this interaction using one sample t-tests (with test value = 0) for each condition in each age group separately, showed that learning from same-class exemplars helped increase the hit rate for young children,

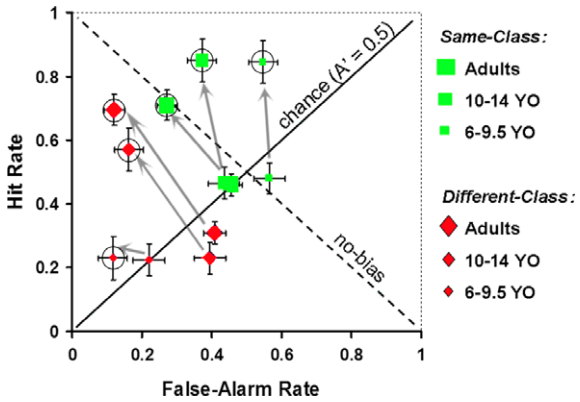


Fig. 5. ROC diagram presenting mean hit and false-alarm rates (error bars represent standard errors). Distance from the diagonal solid line represents sensitivity level (with points on the line representing chance performance, $A' = 0.5$). Distance from the diagonal dashed line represents response bias (points below this line represent "conservative" performance). Gray arrows roughly illustrate the performance change (magnitude and direction) from the *pre-learning test blocks* (not circled) to the *post-learning test blocks* (circled).

Please cite this article in press as: Hammer, R., et al. The development of category learning strategies: What makes the difference?. *Cognition* (2009), doi:10.1016/j.cognition.2009.03.012

Table 3
Mean (±*SD*) changes in hit and false-alarm rates (*post-learning–pre-learning*) in each age group and experimental condition.

Condition/age-group		$6 \leqslant$ Age $\leqslant 9.5$	$10 \leqslant$ Age $\leqslant 14$	Adults
Same-class exemplars	Hit rate change	0.36 ± 0.13	0.38 ± 0.18	0.25 ± 0.24
	FA rate change	−0.02 ± 0.17	−0.06 ± 0.21	−0.18 ± 0.36
Different-class exemplars	Hit rate change	0.00 ± 0.13	0.34 ± 0.26	0.38 ± 0.27
	FA rate change	−0.10 ± 0.13	−0.23 ± 0.15	−0.29 ± 0.19

$t(9) = 8.93$, $p < .001$, $d = 5.95$, older children, $t(9) = 6.78$, $p < .001$, $d = 4.52$, and adults, $t(19) = 4.75$, $p < .001$, $d = 2.18$. In turn, learning from different-class exemplars did not have a significant effect on young children's hit rate, $t(9) = .12$, $p = .91$, but it dramatically increased the hit rate of older children, $t(9) = 4.07$, $p < .005$, $d = 2.71$, and adults, $t(19) = 6.34$, $p < .001$, $d = 2.91$.

In summary, learning by same-class exemplars comparison is significantly helpful in increasing children's hit rate – i.e., for correctly identifying creatures that are of the same kind. But this comes at the cost of over generalization – i.e., occasionally identifying creatures of different kinds as if they are from the same kind. In contrast, for adults, learning from same-class exemplars resulted both in increased hit and reduced false-alarm rates. In the *different-class exemplars* condition, participants' pattern of behavior was quite different. Namely, learning from different-class exemplars was highly useful for older children and adults, enabling increased sensitivity without changing their response bias. However, for younger children, learning from different-class exemplars had only a minor contribution in reducing their false-alarm rate.

4. Discussion

In the current study, we tested the contribution of object comparison to category learning. Specifically, we tested the differential utility of comparing same-class exemplars vs. different-class exemplars. We hypothesized that if the bias to favor commonalities over differences in comparison processes indeed contributes to the late acquisition of subordinate categories, then both children and adults should be able to learn effectively new categorization principles by comparing same-class exemplars, but young children should have greater difficulty than adults when comparing different-class exemplars. Alternatively, if the bias has no relevance to the developmental findings regarding hierarchical structure, then no such age by condition interaction should occur. The results strongly support the former hypothesis.

Our findings show that elementary school aged children (6–9.5 years old), similar to adolescents (10 years old and older) and adults, were capable of learning a categorization principle after being presented with a few paired same-class exemplars. In contrast, when provided with paired different-class exemplars, young children, but not older children or adults, showed poor performance. In fact, while young children showed the greatest improvement in category learning performance when presented with same-class exemplars, older children and adults learned even better when presented with different-class exemplars.

A number of studies in the developmental literature have noted the importance of comparison for category learning. For instance, Gentner and Namy (1999) found that when young children are asked to categorize an object (e.g., a banana) in isolation, they often do so by using similarities that are either thematic (e.g., grouping it with a monkey) or perceptual (e.g., grouping it with a moon). However, when the same object is paired with another object from the same superordinate category (e.g., an apple), children switch back to sorting it taxonomically (e.g., grouping it with an orange). These researchers, and others, noted that the process of comparison invites children to perceive, attend to, or perhaps even actively search for, commonalities between the compared items (Gentner & Namy, 2006). Waxman has argued that this may be a major source of the finding that applying the same label to different objects facilitates categorization (Waxman, 1999). What the present findings reveal, however, is that this "consequence" of comparison is in fact a bias, especially for young children. In every comparison process, the observer can potentially detect *both* commonalities and differences between items, and the capacity to detect these could, *a priori*, be equivalent. Our findings are consistent with the idea that detecting commonalities is favored, and that this preference is significantly exaggerated in young children.

Our analysis of the error patterns supports the above conclusion. In particular, young children's high false-alarm rate (when presented with same-class exemplars) means that they were especially prone to over generalize, thus including in the relevant category objects that did not fit all of its defining features. There is indeed a vast developmental literature on young children's tendency to over generalize, ranging from overextensions in word learning (Gelman, Croft, Panfang, Clausner, & Gottfried, 1998), to over regularizations in rule learning (Marcus et al., 1992). In the present context, this finding fits current claims in the categorization literature that children start off with fairly global categories, and only later do they break these down into narrower classes (Mandler, 2008; Quinn, 2004).

The pattern of errors among the older participants, mainly the reduction of false-alarms, suggests a potential advantage for using different-class exemplars (see also Hammer et al., 2005, 2008, 2009) – an advantage not available to young children. Namely, comparing different-class exemplars that differ only in a single salient property, as was the case in the current experiment, is very useful for identifying a relevant dimension for categorization. Apparently, starting in late childhood, people become capable of

implementing this useful strategy for learning by different-class exemplars comparison.

Young children's relative difficulty in learning from comparison of different-class exemplars, even when the two learning conditions are objectively similarly useful as in the present study, can contribute to the late emergence of subordinate categories. Specifically, the argument is that learning global categories requires mainly detection of a few within-category commonalities while ignoring the many within-category differences, something that can be effectively achieved by same-class comparison. In turn, learning subordinate-level categories requires primarily detection of the few between-category differences, something that is most effectively achieved via different-class comparison. It thus follows that the differential usability of these two learning-by-comparison processes across ages documented here, may give rise to developmental changes in the hierarchical structure of categories.

An open issue underlined by the current findings relates to the origins of this comparison bias. Why does learning from same-class exemplars emerge earlier than learning from different-class exemplars? A number of motivational and/or cognitive causes are plausible. One motivational alternative is that early in development, children may lack the *need* or *interest* to learn highly specific categories, thus leading to less practice in different-class exemplars comparison. A cognitive possibility is that the two comparison processes place different demands on *working memory* (Halford, Wilson, & Phillips, 1998), which in turn might give rise to the developmental differences. A second cognitive possibility is that the two processes require different kinds of *inferences for learning* a categorization rule. In particular, a same-class exemplars comparison decisively indicates irrelevant feature dimensions (feature-dimensions in which the same-class exemplars differ). Nevertheless, same-class comparisons also invite the learner to use a strategy by which he or she directly assumes which are the relevant feature-dimensions (feature-dimensions in which the same-class exemplars are similar). In turn, when learning from different-class exemplars comparison, attending to similarities has no value. Here, learners can only infer that, taken together, the set of discriminating feature-dimensions is relevant for categorization. That is, if two objects do *not* share all of these features, then they are *not* of the same kind. When the objects differ in only one feature, then this feature must be a relevant one. This inference from negation, (which is useful only when the compared exemplars differ in only a single feature), may be harder, especially for younger children.

A related computational explanation for the observed comparison bias is that same-class indications are not only usable for forming a rule-based category representation, but they are also usable for a similarity based representation. In particular, a small set of same-class exemplars can either be directly used as an exemplar-based representation of a category (by mapping the permitted distribution of a category members within the feature-dimension space), or can be used for creating a prototype-based representation (the same-class exemplar set can be used for computing a weighted mean of the prototype properties). In contrast, different-class indications are poorly usable for this purpose. This is the case even when they are sufficiently informative for creating a rule-based representation as in the current experimental setting. Properly using different-class indications for a non-rule-based representation would require a different strategy than the one used with same-class indications (Hammer et al., 2007, 2008, 2009). It is possible that in our experiment, when young children were provided with same-class indications, they were able to form a similarity-based representation of categories, which enabled a similar performance level to that of the older participants. However, young children might not have an alternative strategy for computing different-class indications, as they are also not as capable as older participants in forming an explicit categorization rule using different-class indications.

A final possibility that we would like to propose, however, is that there are further, even more significant *objective* computational differences between the two comparison processes. As it will become clear, the strength of this account is that it makes the comparison bias inevitable, thus providing an ecological explanation for the development of the hierarchical structure of categories. Moreover, this explanation is consistent with the motivational and cognitive explanations listed above, and may, in fact, provide an account of their origin.

4.1. The information quantity of exemplars comparison

In the experiment reported here, we predefined the target categories by two feature-dimensions, e.g., the tail and the eye, and deliberately equated the information quantity in the two learning conditions by providing participants with either same-class indications or different-class indications with an information value of 1 bit. In recent studies, however, Hammer et al. (2007, 2008) showed that the qualitative differences between same-class and different-class comparisons are, in fact, typically associated with a quantitative difference in the information content of these two comparison types. Specifically, a typical same-class comparison is significantly more informative than a typical different-class comparison. This statement is demonstrated by the example portrayed in Fig. 6, which presents a scenario similar to the one illustrated in Fig. 2, wherein creatures differ on four possible feature-dimensions.

In terms of same-class comparisons, Fig. 6 reveals that, unlike the stimuli presented in the current experiment, not all same-class indications provide 1 bit of information. Being constrained only by the requirement that the two paired creatures will share the same tail and eye, the following same-class indications are possible. (1) If informed that a creature is of the same kind as itself (or another apparently identical creature; Fig. 6b(I)), then we are provided with no information. Such indication does not permit us to exclude any of the hypotheses presented in the hypotheses table (Fig. 6a), and thus, $-\log_2 16/16 = 0$ bits. (2) If two paired same-class creatures differ in a single feature (Fig. 6b(II)), then we can exclude all the hypotheses in which this feature is identified as relevant, i.e., $-\log_2 8/16 = 1$ bit (leaves H1, H2, H5, H6, H9, H10, H13, and H14). This is the type of pairing used in the learning

a

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16
Tail	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Eye	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Fur	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Skin	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

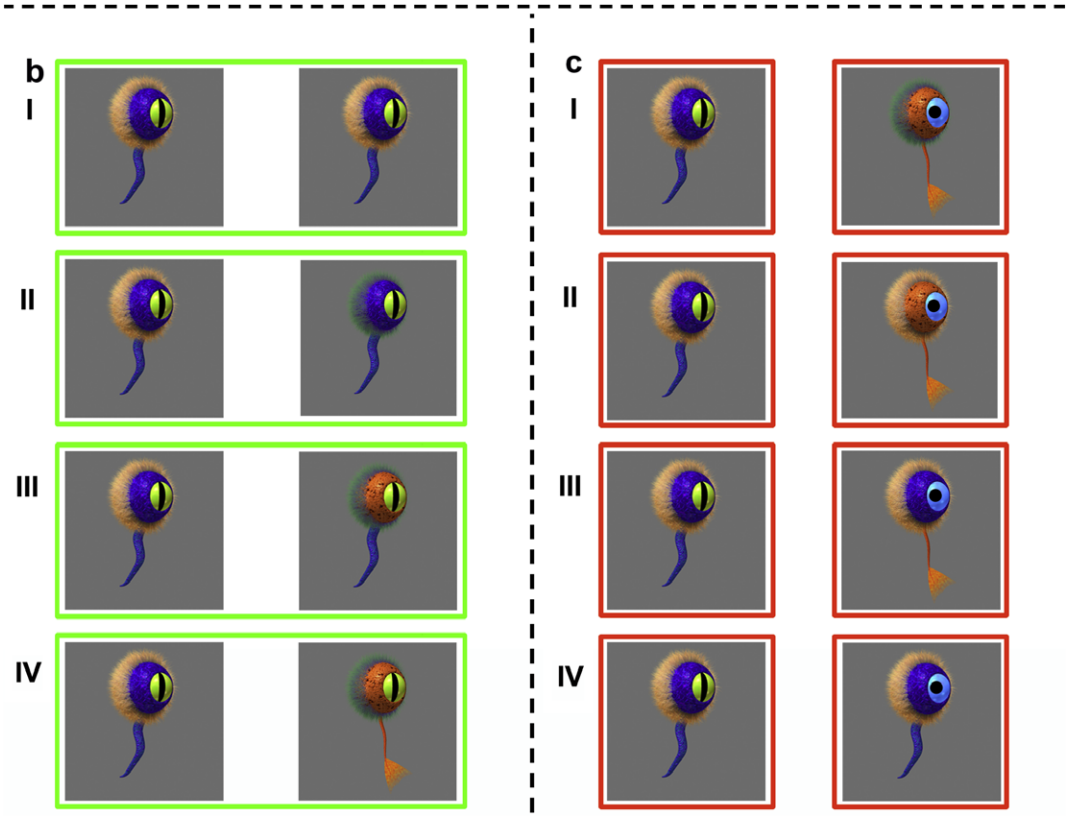


Fig. 6. The different information quantity possibilities for same-class and different-class comparisons for a four-dimensional feature space. (a) The hypothesis table. (b) Same-class exemplars pairs, from poorly informative (I) to highly informative (IV). (c) Different-class exemplars pairs, from poorly informative (I) to highly informative (IV).

phase in the *same-class exemplars* condition of the current experiment. (3) If two paired same-class creatures differ in two features (Fig. 6b(III)), then we can exclude all the hypotheses in which either one of these features is relevant, $-\log_2 4/16 = 2$ bits (leaves H1, H5, H9, and H13). (4) If we had more generalized categories than those used in the current experiment, such as categories in which same-class creatures could also differ in their tails (Fig. 6b(IV)), then even more informative same-class indications would have become available, $-\log_2 2/16 = 3$ bits (leaves only H1 and H5). As the within category variation increases, which is the case with general categories, the information content of a typical same-class indication also increases.

In contrast, different-class indications are constrained by the requirement that two paired creatures will differ at least in the tail or the eye, thus giving rise to the following possibilities: (1) if two paired different-class creatures differ in all four features (Fig. 6c(I)), then we can exclude only the hypothesis in which no dimension is relevant (H1) $-\log_2 15/16 = 0.093$ bit. (2) If the two different-class creatures differ in three features (Fig. 6c(II)), then we can exclude only the hypotheses in which none of the features differentiating the two creatures is relevant, $-\log_2 14/16 = 0.193$ bit (exclude only H1 and H3). (3) If the two different-class creatures differ in two features (Fig. 6c(III)), then we are provided with $-\log_2 12/16 = 0.415$ bit (exclude only H1, H2, H3 and H4). Finally, (4) if the two different-class creatures differ in only one feature (Fig. 6c(IV)) – as was the case in the *different-class exemplars* condition of the current experiment – then we are provided with $-\log_2 8/16 = 1$ bit (exclude H1, H2, H3, H4, H9, H10, H11, H12, and H13). From this analysis, we can see that the maximal information value of different-class indications will always be 1 bit. When the number of between category differences decreases, as is the case when referring

to subordinate-level categories, then the relative portion of the more informative different-class indications, compared to the poorly informative ones, increases.

In sum, excluding the null case in which we are informed that a creature is of the same kind as itself, the minimal information quantity of same-class indication is equal to the maximal information quantity of different-class indication. Furthermore, as the number of irrelevant feature-dimension increases, the information quantity of a typical same-class indication exponentially increases, while the information quantity of a typical different-class indication exponentially decreases (for a formal proof of this statement, see Hammer et al., 2008, Appendix 1).

As a result of this analysis, we suggest that even an ideal observer, who has no specific motivation for creating a particular hierarchical organization of categories, and no constraints on working memory or inferential capacities, will nonetheless face difficulties in learning from haphazard different-class indications simply because these are objectively poorly informative. In contrast, the information content of same-class indications is always high, enabling observers to identify irrelevant variations in almost all conditions (see Hammer et al., 2007, for a computer simulation supporting this statement). That is, it may be the case that everyday life experiences have motivated us to perceive different-class indications as worthless.

Moreover, the information quantity of a typical same-class indication is expected to be higher than that of a typical different-class indication due to the fact that same-class indications are transitive (if $A = B$, and $B = C$, then $A = C$), but different-class indications are not (if $D \neq E$, and $E \neq F$, then the relation between D and F cannot be inferred). Although transitivity is not relevant to the category learning task we tested here, we suggest that its usability in everyday life scenarios may have also contributed to the expectation of receiving valuable information only from same-class comparisons.

In summary, this theoretical survey suggests that: (1) the differential usability of same-class and different-class indications is an *objective* fact; (2) same-class indications are always highly informative for identifying irrelevant within category differences, and their information value increases as one shifts to more global categories where the within category variation is large; and (3) different-class indications will be informative only when the compared different-class exemplars differ in very few features. The postulated causality that can be derived from these conclusions is that the hierarchical structure of categories may emerge from the computational limitations of using same-class and different-class indications, especially when the learner has limited computational resources. In particular, subordinate-level category learning will require information that can be gained mainly from different-class comparisons. However, given that collecting pieces of coherent information from haphazard different-class comparisons is computationally very demanding, this learning process is unlikely to be rewarding. Consequently, young children are likely to typically rely either on same-class comparison processes, or on unsupervised category learning strategies affected by bottom-up factors such as global similarity judgment biased by the distinctiveness of object features (e.g., Hammer & Diesendruck, 2005; Samuelson & Smith, 2000; Sloutsky, 2003). This, in turn, will make the learning of more global categories commonplace, and the learning of subordinate categories infrequent.

Over-generalized category representations may also result from the fact that although same-class indications have high information content, learning only from same-class comparisons may end up with a potentially increased number of false-alarms (but not misses). This is so because the constraints imposed by same-class indications always leave alternative hypotheses, in addition to the correct one (H13 in the example described in Fig. 2), in which some of the relevant feature-dimensions are suspected as irrelevant (as is the case with H1, H5, and H9). If people implement a heuristic by which they look for the simplest representation possible consistent with the constraints imposed by the provided same-class indications, then there is high likelihood that they will select an over simplified representation such as the one suggested by H9 or H5, in which only one of the relevant features is indeed treated as relevant, or perhaps even H1, in which none of the features is taken to be relevant (see Pothos & Close, 2008 for similar thoughts). At the same time, using this "simplification of representation" heuristic when trained with informative different-class indications is likely to lead learners to select the correct hypothesis (H13), since this will always be the simplest representation suggested by these indications. The response bias analyses reported here support the idea that people, at least partially, follow this heuristic (see Hammer et al., 2009, for further discussion of previous similar results with adults).

Later in development, perhaps in order to meet further everyday life demands, people adopt tools appropriate for learning more specific categories, and become able of extracting insights by comparing objects from different categories. But this requires further effort since informative different-class indications are not likely to become available by sheer chance. In fact, in many circumstances, even adult participants need to be "pushed" in order to correctly execute learning by different-class exemplars comparison (Hammer et al., 2005, 2008, 2009). Moreover, such learning may require the availability of an "expert supervisor", knowingly providing the learner with different-class exemplars that are similar enough to be informative. This may not only drive the learner to reconsider a few easily perceived different-class exemplar differences as important, but it may also boost perceptual learning by forcing the learner to identify subtle differences between apparently identical exemplars. For example, training physicians by contrasting two highly similar X-ray images, one of a patient with an early tumor and one of a healthy person, may help them detect the few subtle diagnostic features associated with the pathology, even without any further guidance (for similar ideas, see Allen & Brooks, 1991; Brooks, 1987; Brooks, Norman, & Allen, 1991). As Cree and colleagues argue, when deliberately selecting stimuli from low-level concept nodes (e.g., a subordinate-level category), it becomes easy to associate a highly distinctive feature to a concept, and, as a result, differences can be processed even more easily than similarities regardless of the domain of knowledge (i.e., living vs.

non-living kinds) (Cree, McNorgan, & McRae, 2005; see also Randall et al., 2004, for a comparison of domain differences in exemplar complexity). This is consistent with the findings presented here regarding a possible advantage of learning via different-class comparison when exemplars are selected to maximize the information provided.

Although there is a relatively heavy cost in searching for more informative different-class indications or extracting information from relatively poorly informative different-class indications, different-class exemplars comparison may become crucial at a later developmental stage in order to refine conceptual knowledge. At the other extreme, comparison of highly dissimilar same-class exemplars may force the learner to consider very few subtle similarities as important. This may be essential for learning abstract or highly generalized categories. Eventually, the two learning-by-comparison processes are needed for learners to shift from shallow categorization driven by overall similarities between objects, to categories defined by networks of core properties and the visually accessible properties that are most strongly associated with them. It is this flexibility and depth in categorical representation that are the trademarks of human conceptual knowledge.

Acknowledgements

This study was supported by Grants from the Israel Science Foundation, the US–Israel Binational Science Foundation, and the European Union under the DIRAC integrated project IST-027787.

References

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

Boroditsky, L. (2007). Comparison and the development of knowledge. *Cognition, 102*(1), 118–128.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 141–174). Cambridge, England: Cambridge University Press.

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

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65*, 14–21.

Cree, G. S., McNorgan, C., & McRae, K. (2005). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Language, Memory and Cognition, 32*(4), 643–658.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experemental Psychology: General, 132*(2), 163–201.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*, 1–43.

Furrer, S. D., & Younger, B. A. (2005). A developmental investigation of asymmetry in infants' categorization of cats and dogs. *Developmental Science, 8*(6), 544–550.

Gelman, S., Croft, W., Panfang, F., Clausner, T., & Gottfried, G. (1998). Why is a pomegranate an apple? The role of shape, taxonomic relatedness, and prior lexical knowledge in children's overextensions of apple and dog. *Journal of Child Language, 25*, 267–291.

Gelman, S. A., & Waxman, S. R. (2007). Looking beyond looks: Comments on Sloutsky, Kloos, and Fisher (2007). *Psychological Science, 18*, 554–555.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science, 5*(3), 152–158.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.

Gentner, D., & Namy, L. L. (2006). Analogical processes in language learning. *Current Direction in Psychological Science, 15*(6), 297–301.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424–429.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 723–802.

Hammer, R., Bar-Hillel, A., Hertz, T., Weinshall, D., & Hochstein, S. (2008). Comparison processes in category learning: From theory to behavior. *Brain Research, 1225*, 102–118.

Hammer, R., & Diesendruck, G. (2005). The role of dimensional distinctiveness in children and adults' artifact categorization. *Psychological Science, 16*(2), 137–144.

Hammer, R., Hertz, T., Hochstein, S., & Weinshall, D. (2005). Category learning from equivalence constraints. In *Proceedings of the 27th annual conference of the cognitive science society*.

Hammer, R., Hertz, T., Hochstein, S., & Weinshall, D. (2009). Category learning from equivalence constraints. *Cognitive Processing*.

Hammer, R., Hertz, T., Hochstein, S., & Weinshall, D. (2007). Classification with positive and negative equivalence constraints: Theory, computation and human experiments. In F. Mele, G. Ramella, S. Santillo, & F. Ventriglia (Eds.), *Brain, vision, and artificial intelligence. Second international symposium, BVAI 2007. Lecture notes in computer science* (pp. 264–276). Berlin, Heidelberg: Springer-Verlag Press.

Horton, M., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development, 51*, 708–719.

Keil, F. C. (2008). Space – The primal frontier? Spatial cognition and the origins of concepts. *Philosophical Psychology, 21*(2), 241–250.

Kurtz, K. J., & Boukrina, O. (2004). Learning relational categories by comparison of paired examples. In *Proceedings of the 26th annual conference of the cognitive science society*.

Malt, B. C. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology, 29*, 85–148.

Mandler, J. M. (2008). On the birth and growth of concepts. *Philosophical Psychology, 21*(2), 207–230.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57* (4, Serial No. 228).

Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431–467.

Markman, A. B., & Wisniewski, E. J. (1997). Same and different: The differentiation of basic-level categories. *Journal of Experimental Psychology: Language, Memory and Cognition, 23*, 54–70.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1990). Similarity involving attributes and relations: Judgments of similarity and difference are not inverses. *Psychological Science, 1*(1), 64–69.

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development, 53*, 258–266.

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

Murphy, G., & Brownell, H. (1985). Category differentiation in object recognition: Typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 70–84.

Namy, L. L., & Gentner, D. (2002). Making a silk purse out of two sow's ears: Young children's use of comparison in category learning. *Journal of Experimental Psychology: General, 131*, 5–15.

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition, 107*, 581–602.

Quinn, P. C. (2004). Development of subordinate-level categorization in 3- to 7 month-old infants. *Child Development, 75*, 886–899.

Quinn, P. C., & Johnson, M. H. (2000). Global-before-basic object categorization in connectionist networks and 2-month-old infants. *Infancy, 1*, 31–46.

Randall, B., Moss, H. E., Rodd, J. M., Greer, M., & Tyler, L. K. (2004). Distinctiveness and correlation in conceptual structure: Behavioral and computational studies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(2), 393–406.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Samuelson, L. K., & Smith, L. B. (2000). Grounding development in cognitive processes. *Child Development, 71*, 98–106.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7*, 246–251.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007a). What's beyond looks? Reply to Gelman and Waxman. *Psychological Science, 18*(6), 556–557.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007b). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science, 18*(2), 179–185.

Sloutsky, V. M., Lo, Y.-F., & Fisher, A. (2001). How much does a shared name make things similar? Linguistic labels, similarity and the development of inductive inference. *Child Development, 72*(6), 1695–1709.

Spalding, T. L., & Ross, B. H. (1994). Comparison-based learning: Effects of comparing instances during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1251–1263.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.

Waxman, S. R. (1999). The dubbing ceremony revisited: Object naming and categorization in infancy and early childhood. In D. L. Medin & S. Atran (Eds.), *Folkbiology* (pp. 233–284). Cambridge, MA: MIT Press.

Waxman, S. R., & Braun, I. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition, 95*(3), 59–68.

Waxman, S. R., & Lidz, J. (2006). Early word learning (6 ed.. In D. Kuhn & R. Siegler (Eds.). *Handbook of child psychology* (Vol. 2, pp. 299–335). Hoboken, NJ: Wiley.

Waxman, S. R., Lynch, E. B., Casey, K. L., & Baer, L. (1997). Setters and samoyeds: The emergence of subordinate level categories as a basis for inductive inference in preschool-age children. *Developmental Psychology, 33*, 1074–1090.

Younger, B. A., & Fearing, D. D. (2000). A global-to-basic trend in early categorization: Evidence from a dual-category habituation task. *Infancy, 1*, 47–58.

Please cite this article in press as: Hammer, R., et al. The development of category learning strategies: What makes the difference?. *Cognition* (2009), doi:10.1016/j.cognition.2009.03.012