

ARTICLE

The blowfish effect: children and adults use atypical exemplars to infer more narrow categories during word learning

Lauren L. EMBERSON, Nicole LONCAR, Carolyn MAZZEI, Isaac TREVES,
and Adele E. GOLDBERG*

Princeton University, USA

*Corresponding author: Princeton University, Psychology Department, Princeton, NJ 08544, USA.

E-mail: adele@princeton.edu

(Received 13 August 2018; revised 12 February 2019; accepted 18 April 2019)

Abstract

Learners preferentially interpret novel nouns at the basic level ('dog') rather than at a more narrow level ('Labrador'). This 'basic-level bias' is mitigated by statistics: children and adults are more likely to interpret a novel noun at a more narrow label if they witness 'a suspicious coincidence' – the word applied to three exemplars of the same narrow category. Independent work has found that exemplar typicality influences learners' inferences and category learning. We bring these lines of work together to investigate whether the content (typicality) of a single exemplar affects the level of interpretation of words and whether an atypicality effect interacts with input statistics. Results demonstrate that both four- to five-year-olds and adults tend to assign a narrower interpretation to a word if it is exemplified by an atypical category member. This atypicality effect is roughly as strong as, and independent of, the suspicious coincidence effect, which is replicated.

Keywords: word learning; suspicious coincidence; atypicality; language

Introduction

Philosophers and psychologists have long marveled at how it is that children learn the meanings of new words so quickly and so well (Medina, Snedeker, Trueswell, & Gleitman, 2011; Quine, 1960). To be successful, children ultimately take a number of factors into account, including an entity's shape and function, and its linguistic and non-linguistic contexts. Perhaps the best-known factor is the tendency to interpret novel nouns as referring to a BASIC taxonomic level (Golinkoff, Mervis, & Hirsh-Pasek, 1994; Markman, 1989). For example, speakers tend to interpret a novel word used to refer to a Dalmatian dog as meaning 'dog' as opposed to 'Dalmatian' or 'animal'; they likewise tend to interpret a novel word applied to a Macintosh apple as an 'apple' and not a 'Macintosh apple' nor 'fruit' (Hall, 1993; Hall & Waxman, 1993; Markman, 1989; Rosch *et al.*, 1976; Taylor & Gelman, 1989;

© Cambridge University Press 2019

Waxman, 1990; Waxman, Shipley, & Shepperson, 1991; cf. Callanan, Repp, McCarthy, & Latzke, 1994). It is well established that basic-level terms tend to be learned earlier and used more frequently than more narrow (subordinate level) or more broad (superordinate level) terms.

It is worth considering why a privileged level of description exists. Murphy and Brownell (1985) make a compelling case that the basic level conveys the appropriate amount of information in most contexts. For example, knowing that a thing is an apple tells us a great deal of relevant information, including what kind of shape and texture it has, what it tastes like, and how to eat it. Knowing that a thing is more specifically a Macintosh apple only adds a small amount of additional information and that added information is often not directly relevant to communicative demands; it simply doesn't usually matter if one is holding or has eaten a Macintosh or a Fuji apple. At the other end of the spectrum, knowing that something is a fruit is often insufficient, since it tells us little about its size, color, taste, or how it is to be eaten. That is, the level of description that corresponds to the basic level is one that determines the category's overall shape, function, and affordances, and it is therefore the most appropriate term to use in the majority of contexts.

If learners treat basic-level interpretations as a default, as prior work suggests, then the question arises as to when and why they ever decide to assign a more narrow interpretation to a novel word. It is easy to see how witnessing multiple exemplars from distinct categories can encourage learners to generalize to a higher level. For example, if Macintosh, Fuji, and Granny Smith apples are all labeled the same way, then the label cannot refer to any of these subtypes and is instead more likely to mean 'apple'. Likewise, if an apple, a banana, and a peach are all labeled with the same word (e.g., *fruit*), then the word must refer to an even higher level of generalization. Learners' ability to interpret words more narrowly than the basic level, however, is not accounted for as simply (Jenkins, Samuelson, Smith, & Spencer, 2015). The tendency to interpret words at the basic level helps us learn the words *dog*, *apple*, and *table*, and witnessing a word applying to a variety of exemplars encourages more broad generalizations, but these factors are decidedly unhelpful for learning words such as *Dalmatian*, *Granny Smith*, or *coffee table*. One recognized situation in which a more narrow interpretation is encouraged is when a new term properly includes a basic-level term: a *coffee table* is a more narrow type of *table* and a *Granny Smith apple* is a special type of *apple*, and both children and adults are sensitive to this (Clark, Gelman, & Lane, 1985; Waxman & Hatch, 1992).

Recent work has also found that the statistics of the input results in more narrow interpretations. Specifically, Xu and Tenenbaum (2007) found that when adults and three- to five-year-old children were shown a single exemplar of a category (e.g., a picture of a Dalmatian dog labeled a *fep*), they exhibited the basic-level bias and were generally willing to extend the label to any instance of the corresponding basic level category (*fep* = 'dog'). But after witnessing three different *feps*, each of which referred to a different exemplar of a Dalmatian, participants were much more likely to apply the term only to other Dalmatians (i.e., a more narrow category) and NOT to other types of dogs. Xu and Tenenbaum suggest that children and adults are aware that witnessing three exemplars of a narrow category presents the learner with a 'suspicious coincidence', because in ostensive contexts people assume exemplars are chosen purposely to be representative of the intended category (an assumption of STRONG SAMPLING). The coincidence is resolved by assuming that the label only refers

to members of the narrower category 'Dalmatian'. Thus different statistics of the input (i.e., witnessing multiple exemplars of a novel word) appear to play a role in determining which level of categorization a novel term applies to (see also Gweon, Tenenbaum, & Schulz, 2010; Lawson, 2014; Xu & Denison, 2009; Lewis & Frank, 2018; but cf. Spencer, Perone, Smith, & Samuelson, 2011).

In this paper, we investigate a different way in which learners may be led to a more narrow interpretation of a novel word. Specifically, we investigate whether the CONTENT of a single exemplar plays a role in whether learners interpret a novel word at a basic or more narrow taxonomic level. Specifically, we investigate whether the typicality of the illustrating exemplar leads to the application of a more narrow interpretation. Importantly, the rationale described above for preferring basic-level descriptions does not necessarily hold for ATYPICAL exemplars of a category (Murphy & Brownell, 1985). Atypical exemplars of basic-level categories (e.g., *bowling ball*, *race car*) often have highly salient or relevant properties that distinguish them from other members of the basic-level category. For example, unlike other balls, bowling balls are heavy and cannot be thrown; unlike other cars, race cars are usually found on race tracks, are driven by specially trained drivers, and don't have car seats for children. Therefore, it is often pragmatically appropriate to refer to these entities with more specific labels in order to convey highly relevant information. If the reason basic-level terms are used most frequently stems from the fact that they provide the appropriate amount of information, we predict that a novel label for an atypical exemplar should be more likely to be interpreted as referring to a more narrow taxonomic level, since the narrower interpretation is more relevant in the case of atypical exemplars. Our hypothesis can also be construed as resolving a different type of 'suspicious coincidence' as follows. Bayesian inference of a word's meaning involves comparing hypotheses about the word's distribution, based on the likelihood of the exemplar(s) being generated from each distribution, given prior knowledge. Our prior knowledge tells us that the distribution of category members is not uniform. Some types of exemplars are more common than others, and atypical exemplars tend to be rare. Given this, the selection of an atypical exemplar to illustrate the meaning of a word that refers to an entire basic-level category is unlikely, presenting a type of suspicious coincidence based on the content of the exemplar rather than the number of similar exemplars. On the other hand, an atypical exemplar is NOT unlikely if the novel word refers only to the narrower category. Thus the suspicious coincidence can be eliminated if the learner assumes that the novel word refers to the more narrow category. In the experiment reported below, we investigate whether child and adult learners use the typicality of exemplars to infer the appropriate level of description for novel words.

It is well known that typicality plays a role in categorization tasks (Larochelle & Pineau, 1994; McCloskey & Glucksburg, 1978; Murphy, 2004; Murphy & Brownell, 1985; Rips, Shoben, & Smith, 1973; Van Overschelde, Rawson, & Dunlosky, 2004). For example, Meints, Plunkett, and Harris (1999) showed that one-year-olds restrict their understanding of many common categories to typical exemplars only, gradually including atypical exemplars over their second year of life. In a mouse-tracking study by Dale, Kehoe, and Spivey (2007), adults displayed more competition from a competing category (fish) when required to classify atypical exemplars of a category (e.g., whale as a mammal). A particularly relevant demonstration of typicality effects comes from Mervis and Pani (1980), who found that adults and five-year-old children generalized within novel categories better and more accurately when shown

typical exemplars of the categories compared to atypical exemplars (see also Rips, 1975). Overall, typicality is recognized to affect categorization with both known categories and in the context of category learning in adults and young children.

While we know that children are sensitive to typicality effects, whether exemplar typicality plays a role in the context of word learning has not been examined. The present study asks whether learners interpret labels of a single atypical exemplar more narrowly than they do if the label applies to a typical exemplar. Following the rationale from Murphy and Brownell (1985), the basic level is the most contextually appropriate level of description for typical exemplars in most contexts, but it is not the most appropriate level of description for atypical exemplars. Are word learners sensitive to what constitutes the contextually appropriate level of description vis-à-vis exemplar typicality? Do both adults and children expect speakers to adjust their labels when a different taxonomic level is more appropriate?

Our goal was to identify whether manipulating exemplar typicality affects word learning in four- to five-year-old children and adults. Specifically, we hypothesized that illustrating a novel word with an atypical exemplar would lead learners to narrow the interpretation of the novel word (e.g., as meaning ‘blowfish’ rather than ‘fish’). We also manipulated whether one or three exemplars of a category were witnessed in order to compare any effect of typicality with the expected effect of the number of exemplars on participants’ tendency to assign a more narrow interpretation to novel nouns (Xu & Tenenbaum, 2007).

Participants were shown one or three exemplars, which were referred to by a novel label (e.g., “This is a *fep*”). Exemplars were either typical or atypical exemplars. The pictures of exemplars were separately normed for typicality as members of dog, fish, flower, and bird categories as described in the ‘Methods’ section below. Participants were shown an array of eight entities and asked to “check the box(es) for any other *feps* that you find in the pictures below”. The eight pictures always included two subordinate-level matches (e.g., two additional golden retrievers, if the first exemplar was a golden retriever), two basic-level matches (e.g., a Labrador and a beagle), and four distractors (pictures of other categories). The order of the pictures included in each display was randomized across participants.

In a subsequent task, children were tested on all four categories in order to determine whether they were able to recognize that the atypical exemplars were members of the intended categories. Clearly children cannot be expected to interpret a label at a higher level of categorization if they do not recognize it is a member of the higher-level category. Children’s performance was therefore reanalyzed considering only those trials in which the child correctly recognized that the atypical exemplars WERE instances of the intended basic-level categories. If exemplar typicality does affect learners’ pattern of generalization, it will demonstrate that learners are selectively altering the level at which a novel word is interpreted, depending on the content of witnessed exemplar(s).

Methods

Participants

Children

Forty (40) monolingual English-speaking children aged four and five years (26 female, $M = 4;9$, range = 4;0–5;10; $SD = 6$ months) were recruited from the local area through a

variety of means: during visits to the Baby Lab, community children's events, or at a local preschool ($N = 17$). Monolingual exposure was defined as hearing 80% English or greater by parental report (language exposure for the included sample: $M = 96.68\%$, $SD = 5.12\%$, range: 80–100%). An additional 5 children were tested but not included because of technical errors or a failure to finish all trials in the experiment ($N = 2$) or because they did not meet our criteria for being monolingual ($N = 3$). No children were excluded on the basis of their selections during the test or categorization trials (discussed below). Those who took part in the experiment in the lab setting were given a T-shirt and children's book for their visit. Those who participated in the school setting were given a children's book. Caregivers provided consent for participation prior to the beginning the study.

Adults

Participants were 43 undergraduate students between the ages of 18 and 25 years (32 females, $M = 21$ years, $SD = 2.56$, 95.8% exposure to English), recruited at the student campus center and compensated with a cookie or cupcake. Participation was preceded by an explicit consent procedure, and participants were debriefed about the study afterwards.

Stimuli

Visual stimuli

We investigated the role of exemplar typicality and number of exemplars in both children and adults, using an interactive touch-screen tablet (iPad). Each trial contained either one or three exemplars of the same narrow (subordinate) category, and the exemplar(s) were either typical or atypical. Images of dogs, fish, birds, and flowers were used. The stimuli were extensively normed for typicality as reported in separate work by Emberson and Rubinstein (2016). Specifically, 62 participants rated how typical the pictures were on a scale of 1–5 across three experiments (i.e., "How typical is this picture? 1 for not typical, 5 for very typical"). In each experiment, participants reported significantly higher ratings for typical than the atypical exemplars employed (typical exemplars were given an average rating of 4 and atypical exemplars were given ratings averaging between 2 and 3). The examples of typical and atypical stimuli used in the present experiments are provided in Figure 1.

Figure 2 presents two representative trials. During the initial introduction to Mr Frog, participants saw a simple animation of a jumping frog. This same animation was used prior to the categorization phase at the end of the experiment.

Auditory stimuli

The current study used prosody intended to engage children, recorded by a female native English speaker. The four novel words that were used throughout the experiment were *fep*, *zak*, *lat*, and *galt*, which all obey English phonotactic constraints. The volume was set to 65% of the total iPad volume and was held at a constant level for every child in order to ensure that it was not too loud, but that they could hear the instructions clearly.

A categorization task followed the novel word interpretation task, as described below. During the categorization phase, children were asked to select examples of the English categories dog, flower, fish, and bird (e.g., "Can you show Mr Frog the dogs?").

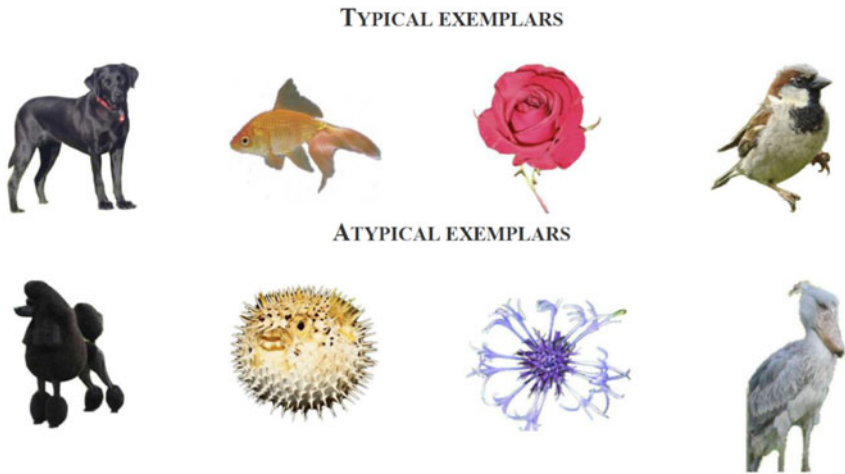


Figure 1. Representative stimuli.

Procedure

The experiment was administered using an iPad (Swift, 3.1). Children were generally seated beside the experimenter, who held the iPad screen at an angle that allowed the children to view and select the images. Children used a pair of child-sized headphones that allowed them to hear the instructions while attenuating any background noise.

The experiment had three phases: an orientation phase, a test phase, and a categorization phase. The orientation phase was included to familiarize the children with the on-screen testing method, as well as with the fact that they were able to select either a single image or multiple images in each trial. This phase consisted of two trials, each of which was repeated as many times as necessary until the child correctly completed them. The first of these trials consisted of an array of twelve images, all of which were primary shapes (e.g., squares, circles, etc.) of varying colors. The child was then asked by the experimenter to select all of the blue circles on the screen, of which there were three. If the child did not select all three blue circles, the experimenter would refresh the screen and ask them to try again. Once completed without error, children continued to the second orientation trial by selecting the arrow on the screen. In the second trial the experimenter asked the children to “find the queen” from a myriad of individuals depicting stereotypical occupational garb (e.g., a doctor, a king, a queen, a fireman, etc.). Once again, this trial was repeated until the child selected the single, correct image. Upon completing this trial, the children were prompted to move on to the test trials by pressing an orange arrow on the top right of the screen.

The test phase included four trials to test children’s generalization of novel words to typical and atypical exemplars. We also tested any effects of typicality in relation to the effect of exemplar number. As in Xu and Tenenbaum (2007, Experiment 2), exemplar number was manipulated between-subjects. Exemplar typicality was manipulated within-subjects with typical exemplars presented first and atypical exemplars presented second, for reasons described in the ‘Supplementary Materials’ (available

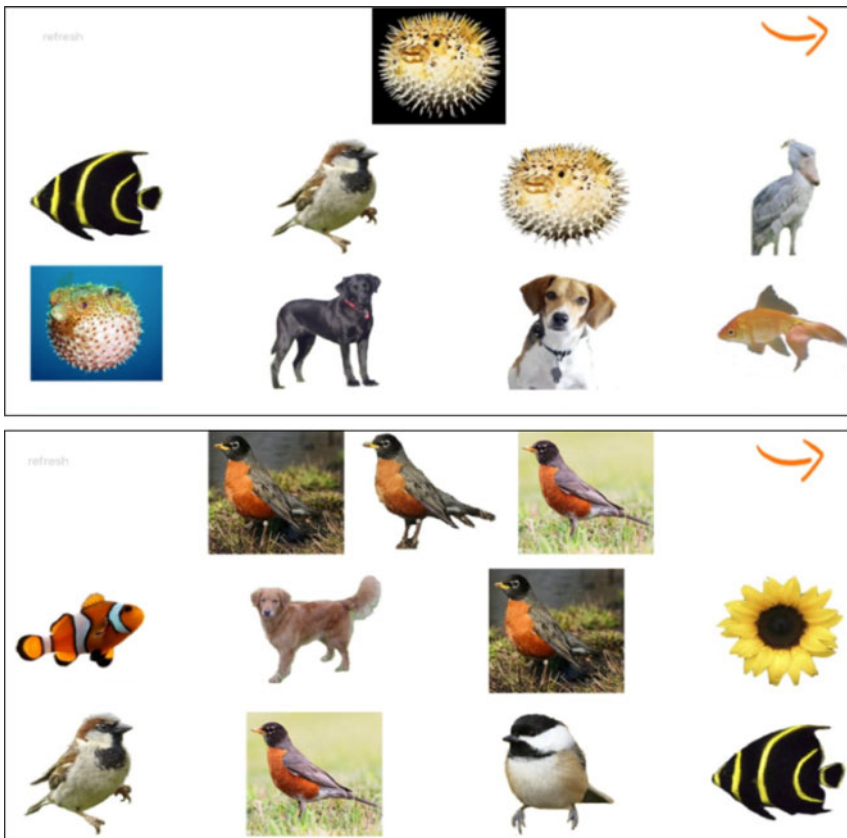


Figure 2. Sample selection screens for the single exemplar (top) and multiple exemplars (bottom) conditions. Children first witnessed only the exemplar(s), which would appear at the top of the screen while paired with a novel, verbal label. After pressing the orange arrow, they were shown an array of 12 images: 4 distractors as well as 2 subordinate-level matches and 2 basic-level matches. They could select as many pictures as they wanted. Note that this Figure only shows 8 images for visual clarity but an additional 4 distractors were presented.

at < <https://doi.10.1017/S0305000919000266>>). We have since reversed the order with a new group of children and found the same effects (see ‘Supplementary Materials’). Each child participated in four test trials: the first two consisted of typical target exemplars while the third and fourth trials consisted of atypical target exemplars.

The number of exemplars presented was manipulated between groups of participants, so that each person saw only 1-exemplar trials or only 3-exemplar trials. Within each trial, the child was first introduced to the exemplar(s) at the top of the screen, paired with a novel verbal label (i.e., “This is a fep” or “These are three feps”). The child then pressed the arrow at the top of the screen which added an array of twelve images below the exemplars: 8 distractors (i.e., unrelated to the category of the exemplar), 2 basic-level matches; 2 subordinate-level matches (Figure 2). The child was then asked, “Can you find the feps?” and was prompted to select as many of the images in the array as they wanted to, at their own pace, before proceeding to the next trial by pressing the arrow again. The order of the four categories of the

exemplars – fish, birds, dogs, and flowers – was counterbalanced across participants so each participant saw each category in one condition only.

Finally, the categorization phase was included to determine whether each child recognized the intended category in the case of atypical exemplars. In this task, children were shown the animation of Mr Frog again and asked to teach Mr Frog words from their language (English). This phase also consisted of four trials where a single, typical image was provided of each category, and children were asked to select “the dogs” (or fish, flowers, or birds) from a set of 12 pictures: 8 distractors, 2 typical, and 2 atypical images for that category. Thus in addition to the categorization task being used to determine whether children recognized atypical exemplars as members of the intended categories, the task was also useful as a check that any reduction in responses for later trials in the main task would not arise from fatigue or distraction. If any children grew weary or distracted by the end of the experiment, this would be evident in the final categorization trials as well.

With adult participants, the procedure was identical, except for an explanation that the experiment was intended for young children. Adult participants were allowed to hold the iPad themselves.

Results

Children’s and adults’ generalization to the basic level was modulated by both exemplar typicality and by the number of target exemplars provided. We employed a logistic regression to predict the number of basic-level responses (out of 2) for each participant based on two fixed effects, the number of exemplars (1 vs. 3) and exemplar typicality (atypical vs. typical), as well as their interaction. Given the within-subjects design, category was included as a random effect in order to control for any differences in generalization across categories. The binary variable of typicality had a reference level of typical and was contrast coded (–1 to 1, for typical to atypical). The binary variable of number of exemplars (1 vs. 3) used the reference level of 1-exemplar and it was also contrast coded (–1 to 1, for 1 to 3 exemplars). We constructed separate models for child (Figure 3) and adult (Figure 4) participants as well as a combined model to investigate systematic differences across age groups.

We find an effect of typicality in adults’ tendency to generalize ($\beta = -0.586$, $Z = -4.39$, $p < .001$). Importantly, we also find a similarly robust effect of typicality in children ($\beta = -0.460$, $Z = -2.82$, $p < .01$). We also replicate the effect of number of exemplars by finding a significant difference in generalization between single and multiple exemplars for children ($\beta = -0.579$, $Z = -3.013$, $p < .01$), as well as for adults ($\beta = -0.450$, $Z = -3.53$, $p < .01$). The effects of the number of exemplars and typicality are again of roughly the same size, and no interaction is evident in either children ($p = .37$) or adults ($p = .56$).

In addition, we sought to determine whether there were differences between the children and adults. We augmented the models that had been employed for children and adults separately to include age group (child vs. adult) as a fixed effect. We conducted this analysis separately for exemplar number and exemplar typicality as we find no interaction between these effects in either age group and these models are more straightforward to interpret. The model for exemplar typicality continues to find a robust effect ($\beta = -0.524$, $Z = -4.96$, $p < .001$). We find a small but significant main effect of age group with children generalizing to the basic level less than adults

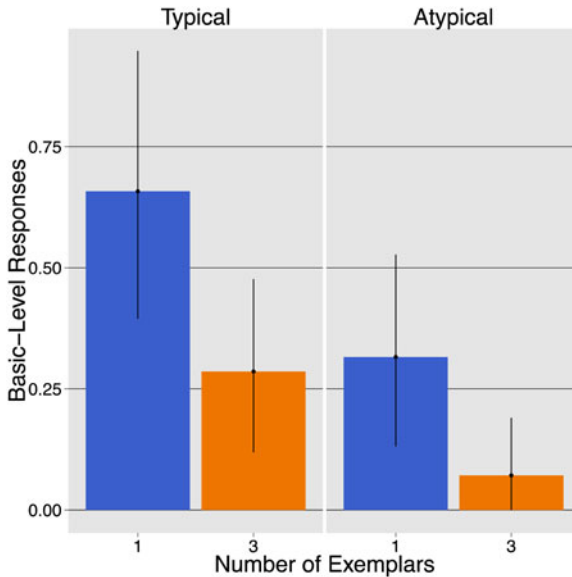


Figure 3. Children's mean # of basic-level selections (out of 2) when asked to find matches of a novel word (e.g., *galt*) when witnessing 1 or 3 typical or atypical exemplars. Error bars represent standard deviations.

($\beta = 0.226$, $Z = 2.14$, $p = .03$). Importantly, we find no interaction between age and exemplar typicality in the selection of basic-level pictures ($p = .51$), indicating that there is no modulation of this effect by age. We find the same results for exemplar number: In the pooled sample across children and adults, we continue to find a robust effect of exemplar number ($\beta = -0.476$, $Z = -4.59$, $p < .001$) and the same small main effect of age ($\beta = 0.253$, $Z = 2.44$, $p = .015$), but no interaction of exemplar number and age ($p = .81$). Thus, we confirm that exemplar typicality as well as number of exemplars modulates word learning similarly in children and adults; specifically, the presentation of an atypical exemplar during word learning results in a narrower interpretation of the novel word.

Child and adult performance were also consistent with task demands (Table 1). During test trials, children reliably selected the narrow (subordinate-level) matches ($M = 1.64$, $SD = 0.66$, out of two possible, averaged over all trial types), and rarely selected distractors ($M = 0.21$, out of 8 possible, $SD = 0.76$, averaged over all trial types). Adults' subordinate selections for test trials were near ceiling ($M = 1.98$, out of two possible, $SD = 0.19$), and they virtually never selected distractors ($M = 0.02$, out of 8 possible, $SD = 0.13$).

We also confirmed that the experimental manipulations (e.g., exemplar typicality) were not found in selections at the subordinate level in either group. Using the same models as above but applied to subordinate-level responses, we find all Z -values are less than the absolute value of 0.4 and all p -values are greater than .7 for children. We also apply these methods to the distractor responses. We find no effect of number of exemplars shown on the selection of distractors ($p = .7$). However, we find a marginal effect of the typicality of the exemplar on the numbers of distractors chosen ($\beta = -0.3688$, $Z = -1.93$, $p = .054$), with more distractors chosen when typical

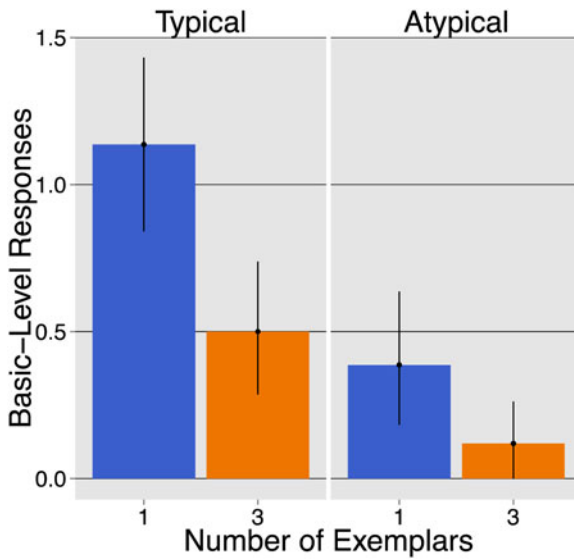


Figure 4. Adults' mean number of basic-level selections (out of 2) when asked to find matches of a novel word (e.g., *galt*) when witnessing 1 or 3 typical or atypical exemplars. Error bars represent standard deviations.

exemplars are presented than when atypical exemplars are chosen. Further examination of this data revealed that, on one trial, one subject selected 6 distractors (a clear outlier). We conducted this analysis again removing this subject and found no effect of typicality on distractor selection ($p = .28$), suggesting that this is not a group-level finding but one biased by this single trial. Adults also exhibited no differences in subordinate or distractor selections based on exemplar typicality or number of exemplars ($Z_s < |0.6|$, $p_s > .5$).

Recall that the subsequent categorization task asked participants to indicate all of the pictures that matched each familiar basic-level label (e.g., 'dog'). For each of the four categories, they were presented with 2 typical, 2 atypical exemplars, and 8 distractors. As expected, adults reliably included atypical selections when asked to indicate all of the pictures that matched each basic-level label ($M = 1.86$, out of 2 possible, $SD = 0.25$), virtually never included any distractors ($M = 0.01$, out of 8 possible, $SD = 0.11$), and selected atypical exemplars at a greater rate than distractors ($t(42) = 50.79$, $p < .001$). Children also reliably selected atypical exemplars (Figure 5). Specifically, children were much more likely to include atypical exemplars ($M = 1.49$, out of 2 possible, $SD = 0.77$) than distractors ($M = 0.25$, out of 8 possible, $SD = 1.03$; paired samples t -test: $t(39) = 10.17$, $p < .0001$). Thus, even though many more distractors were available for selection than atypical exemplars, children selected atypical exemplars that matched their familiar basic-level label far more reliably.

At the same time, there is evidence of a typicality effect for children in the categorization task insofar as they selected the typical exemplars more often than the atypical exemplars (typical exemplars: $M = 1.76$, out of 2 possible, $SD = 0.53$, $t(39) = 3.48$, $p = .001$). Since we cannot expect a child to interpret *feɹ* as 'fish' when shown a blowfish if the child failed to recognize that the blowfish was a fish, we re-ran the analysis for children, excluding test trials in which the child subsequently failed to include atypical exemplars as

Table 1. Selections of all types for both children (left) and adults (right) in all conditions. Two basic-level options were provided on each trial, so, e.g., 0.66 = 33%.

Children				Adults			
Exposure	Basic-level matches (out of 2)	Subordinate-level matches (out of 2)	Other (out of 8)	Exposure	Basic-level matches (out of 2)	Subordinate-level matches (out of 2)	Other (out of 8)
1-typical	M = 0.66	M = 1.63	M = 0.32	1-typical	M = 1.14	M = 2.0	M = 0.02
	SD = 0.91	SD = 0.67	SD = 1.10		SD = 0.95	SD = 0.0	SD = 0.15
3-typical	M = 0.28	M = 1.60	M = 0.26	3-typical	M = 0.50	M = 1.93	M = 0.02
	SD = 0.59	SD = 0.77	SD = 0.83		SD = 0.77	SD = 0.34	SD = 0.15
1-atypical	M = 0.32	M = 1.76	M = 0.07	1-atypical	M = 0.38	M = 1.98	M = 0.02
	SD = 0.66	SD = 0.43	M = 0.27		SD = 0.78	SD = 0.15	SD = 0.15
3-atypical	M = 0.07	M = 1.59	M = 0.19	3-atypical	M = 0.12	M = 2.0	M = 0.0
	SD = 0.34	SD = 0.70	SD = 0.64		SD = 0.45	SD = 0.0	SD = 0.0

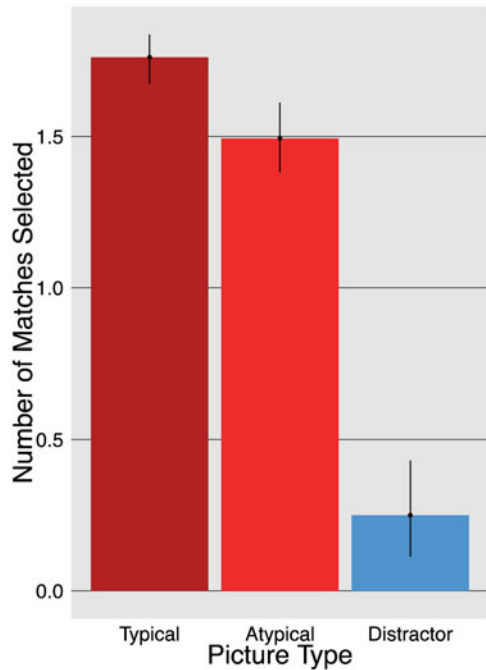


Figure 5. Average number of selections made by children of typical exemplars, atypical exemplars, and distractors during categorization trials. There were 2 possible typical and atypical exemplars and 8 possible distractors. Children were asked to select all of the members of each category (i.e., “Can you find all of the dogs?”). Children selected significantly more atypical matches than distractor images, but significantly fewer than typical images.

members of the basic-level category in the categorization task. Recall that 40 children each received four categorization trials for a total of 160 trials. Children failed to select at least one of these subordinate level pictures on 27 of these trials (17% of total trials). To determine the effectiveness of this approach, we confirmed that we do not find any difference in picture selection between typical and atypical exemplars at test ($t(37) = 0.74$, $p = .47$). Even so, the effect of atypicality on children’s generalization remained significant ($\beta = -0.480$, $Z = -2.64$, $p < .001$), as did the effect of the number of exemplars ($\beta = -0.750$, $Z = -3.65$, $p < .01$). Thus, the reduction in basic-level responses after seeing an atypical exemplar cannot be attributed to children not understanding which category the atypical exemplar refers to, as this effect is persistent, and even appears to be strengthened, when we only include trials where children demonstrate knowledge of the atypical exemplars for a given category.

In addition, we ran an exploratory analysis to determine how category knowledge related to the effects of exemplar number and typicality. Jenkins *et al.* (2015) found that increases in category knowledge resulted in a decrease of the effect of multiple exemplars on children’s generalization. This finding was contrary to the predictions of the Bayesian model by Xu and Tenenbaum (2007). We conducted an exploratory analysis to determine whether category knowledge as assessed during categorization trials had an effect on children’s use of exemplar typicality and number on their generalization to the basic level. To quantify category knowledge, we summed correct responses to the typical and atypical exemplars for each category trials (4 possible across 4 trials, for a total possible score of 16, $M = 13.03$, $SD = 3.11$, median = 13.5, range = 4–16). This category knowledge score had sufficient variance to separate

children into two groups with low category knowledge (M category knowledge = 10.8, range = 4–13, age = 4;8, $n = 20$, 12 female) and high category knowledge (M category knowledge = 15.25, range = 14–16, age = 4;9, $n = 20$, 14 female). Category knowledge is not correlated with age in this sample ($r(39) = 0.05$, $p = .76$).

We ran models separately for each to determine the presence or absence of the effects of exemplar typicality and number on their generalization to the basic level and then compared performance across groups based on category label. Note that this is an exploratory analysis and the dataset is divided in two; it is much less powered than our planned analyses. Overall, we find that the typicality effect that we report here is present in both groups (high category knowledge: $\beta = 0.490$, $Z = -1.95$, $p = .05$; low category knowledge: $\beta = -0.430$, $Z = -1.977$, $p = .05$), and there is no difference between groups (all children together, typicality: $\beta = -0.470$, $Z = -2.81$, $p < .01$; category label (low vs. high, contrast coded): $\beta = -0.17$, $Z = -1.01$, $p = .31$; interaction of typicality and category label: $\beta = -0.027$, $Z = -0.16$, $p = .87$). However, we find that category knowledge has an effect on children's suspicious coincidence effect (i.e., their generalization after seeing 1 or 3 exemplars). The high category knowledge group has the suspicious coincidence effect ($\beta = -1.15$, $Z = -3.04$, $p < .01$), but the low category knowledge group does not ($\beta = -0.23$, $Z = -1.16$, $p = .24$). When considering all children together, we find an effect of category knowledge on the number of basic-level pictures chosen regardless of trial type ($\beta = -0.43$, $Z = -2.0$, $p < .05$), and an interaction of the number of basic-level pictures selected and the number of exemplars children viewed ($\beta = -0.46$, $Z = -2.15$, $p < 0.5$; Figure 6).

Discussion

The present results find evidence of the widely assumed basic-level bias only when a novel label is illustrated by a typical exemplar of the basic-level category. When an atypical exemplar is provided, novel labels instead tend to be interpreted more narrowly. That is, when witnessing an unusual dog labeled as a *fep*, both adults and four- to five-year-old children are likely to interpret *fep* to apply narrowly to only the same type of unusual dog, not to dogs more generally. We find that the effect of witnessing an atypical exemplar affects both adults and four- to five-year-old children to roughly the same extent. Both groups generalize a single atypical exemplar to basic-level pictures less than 20% of the time. We also replicated the 'suspicious coincidence effect' (Xu & Tenenbaum, 2007), in which witnessing a novel label applied to multiple examples of the same narrow subcategory encouraged a narrow interpretation. Moreover, the two effects were roughly equally strong and independent of one another in both child and adult learners. While we have long known that children as well as adults distinguish typical and atypical exemplars of various categories (e.g., Rosch *et al.*, 1976), the present results demonstrate that children specifically make use of exemplar typicality in their interpretation of novel words, mitigating the tendency to interpret new words at the basic level. Thus, this work establishes a new source of information that learners use to make generalizations of novel words that are narrower than the basic level.

The present findings run counter to the idea that children necessarily assume that novel words should be interpreted at the basic level, at least if interpreted as a context-blind bias (Bloom, 2001). Instead, our findings suggest that young children as well as adults attribute more narrow meanings to novel labels when presented with unusual or atypical exemplars. We addressed the possibility that children simply

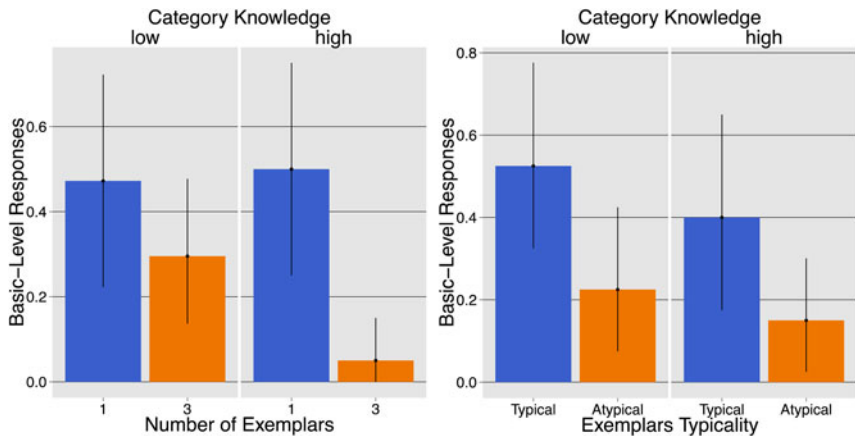


Figure 6. Left panel: suspicious coincidence effect by category knowledge. Right panel: blowfish effect by category knowledge.

did not recognize the atypical members as members of the intended categories with the categorization study results. The categorization task followed the word learning test phase and asked children to select “all of the dogs (fish, flowers and birds)”. Children demonstrated high, but not ceiling level accuracy. We therefore considered performance in the word learning task only on trials in which the same child accurately selected BOTH atypical exemplars as instances of the intended basic-level category during the following categorization task. Results demonstrated that, in fact, children who were shown a blowfish labeled as a *fep* interpreted *fep* to mean ‘blowfish’ rather than ‘fish’, even when they demonstrably recognized that the blowfish was in fact a fish in the categorization task. Thus, the blowfish effect remains in this conservative case and with substantially reduced statistical power (due to the rejection of 35% of total trials).

Recall that typicality was treated as a within-subjects variable and atypical trials always followed typical trials, used so that children would need to retreat from an anticipated basic-level bias upon witnessing atypical examples. We have seen that, in fact, children did select fewer basic-level items as members of the novel category for atypical items. Concerns about this order were addressed in a follow-up replication that we performed with a new group of children who all witnessed atypical exemplars first. Once again, children treated a novel term introduced with an atypical exemplar as only applying to the subordinate category, while they treated a novel term introduced by a typical exemplar as more likely to refer to the basic level (see ‘Supplementary Materials’).

Children tended to include all category members when asked to select all of the fish, dogs, flowers, or birds, and selected many more pictures than in the atypical test trials. And when we only considered trials in which children successfully selected BOTH atypical pictures as instances of the basic level category (e.g., both pictures of blowfish were selected as ‘fish’), children nonetheless did not treat the novel label as a basic-level description. That is, they did not interpret the novel label assigned to a blowfish as if it meant ‘fish’. Thus, the selection of fewer entities in the main task cannot be attributed to fatigue.

One might worry that the present results hinge on the fact that children at this age already have words for these particular basic-level terms and that THIS is why they tend to interpret the novel terms at the subordinate level; that is, it is possible that children interpret *fep* as something other than ‘fish’ because they already have the word *fish*. While we know that children do avoid multiple labels for a given concept, this possibility would not explain the difference between atypical vs. typical examples, nor would it address the fact that three instances of a subordinate category were interpreted differently than one (as established in other studies; e.g., Xu & Tenenbaum, 2007). That is, the idea that a *fep* should refer to something other than ‘fish’ may have reduced basic-level interpretations across the board, but it cannot predict the especially strong avoidance of basic-level interpretation for atypical exemplars reported here.

To explain the origin of the blowfish effect, at least three alternatives present themselves. The explanation we favor was alluded to in the ‘Introduction’: learners may use Bayesian reasoning to infer that an atypical exemplar was unlikely to be generated from a category that included many more typical exemplars. Alternatively, since typical exemplars tend to be labeled with basic-level terms and atypical exemplars tend to be labeled with more narrowly circumscribed terms (Murphy & Brownell, 1985), it could be that this specific correlation in our language input is learned through experience and implicitly affects learners’ future interpretations of novel word labels. This explanation would suggest that the blowfish effect is dependent on a certain amount of language experience; we might then expect that adults would exhibit a stronger atypicality effect than children, but we did not find evidence of this. It remains possible that the atypicality effect is learned as a correlation if little data is needed to observe the relationship between atypical exemplars and specific terms, possibly because the correlation is strongly present in the input that a child receives. Analyses of child language corpora or studies of naturalistic word learning scenarios would be useful to determine whether caregivers consistently provide more specific (or modified) basic-level terms when presenting atypical exemplars to their children (“Look at the greyhound’s long legs. I bet he runs faster than other dogs.”). Finally, it is possible that a lower-level, attentional explanation is involved, insofar as the more unusual features of atypical exemplars may attract more attention, which may in turn lead to a more specific, narrower interpretation. This explanation would be consistent with Spencer *et al.*’s (2011) interpretation of the suspicious coincidence effect; they argued that the reason three instances are more likely to lead to a subordinate interpretation of a novel word is that witnessing three instances of the same category leads to increased attention to the instances’ shared attributes. It is possible that both high-level reasoning, language experience, and low-level attentional factors are involved to different degrees and/or at different stages of language learning. Clearly more work is needed to draw firm conclusions as to why exemplar typicality has such a strong effect on word learners’ interpretation of the meanings of novel labels.

Overall, the present results demonstrate that the TYPICALITY of an exemplar plays a role in which taxonomic level is inferred from a novel label even in young children. Learners tend to restrict the interpretation of a novel label if it is illustrated by an atypical exemplar of a higher-level category, even when they recognize that the atypical exemplar is an instance of the higher-level category. This effect is independent of, and as strong as, witnessing multiple exemplars of the same subtype: the suspicious coincidence effect documented by Xu and Tenenbaum (2007). Thus,

we have replicated the finding that *galt* is likely to be interpreted as ‘fish’ if it labels a single salmon, while it tends to be interpreted as ‘salmon’ if it is illustrated by three salmon. The present work highlights a strong and independent effect: if *galt* labels even a single ATYPICAL fish – a blowfish – it is quite likely to be interpreted as ‘blowfish’ rather than ‘fish’.

Supplementary materials. For Supplementary materials for this paper, please visit <<https://doi.org/10.1017/S0305000919000266>>.

References

- Archambault, A., Gosselin, F., & Schyns, P. G. (2000). A natural bias for the basic level. In *Proceedings of the twenty-second annual conference of the Cognitive Science Society* (pp. 60–5). Mahwah, NJ: Erlbaum.
- Bloom, P. (2001). Roots of word learning. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 159–84). Cambridge University Press.
- Callanan, M. A., Repp, A. M., McCarthy, M. G., & Latzke, M. A. (1994). Children’s hypotheses about word meanings: Is there a basic level constraint? *Journal of Experimental Child Psychology*, 57(1), 108–38.
- Clark, E. V., Gelman, S. A., & Lane, N. M. (1985). Compound nouns and category structure in young children. *Child Development*, 1(56), 84–94.
- Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, 35(1), 15–28.
- Emberson, L. L., & Rubinstein, D. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition*, 153, 63–78.
- Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: the case for a developmental lexical principles framework. *Journal of Child Language*, 21(1), 125–55.
- Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, 107(20), 9066–71.
- Hall, D. G. (1993). Basic-level individuals. *Cognition*, 48(3), 199–221.
- Hall, D. G., & Waxman, S. R. (1993). Assumptions about word meaning: individuation and basic-level kinds. *Child Development*, 64(5), 1550–70.
- Jenkins, G. W., Samuelson, L. K., Smith, J. R., & Spencer, J. P. (2015). Non-Bayesian noun generalization in 3- to 5-year-old children: probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science*, 39(2), 268–306.
- Larochelle, S., & Pineau, H. (1994). Determinants of response time in the semantic verification task. *Journal of Memory and Language*, 33(6), 796–823.
- Lawson, C. A. (2014). Three-year-olds obey the sample size principle of induction: the influence of evidence presentation and sample size disparity on young children’s generalizations. *Journal of Experimental Child Psychology*, 123, 147–54.
- Lewis, M. L., & Frank, M. C. (2018). Still suspicious: the suspicious-coincidence effect revisited. *Psychological Science*, 29(12), 2039–47.
- Markman, E. M. (1989). *Categorization and naming in children*. Cambridge, MA: MIT Press.
- McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets? *Memory & Cognition*, 6(4), 462–72.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, 108(22), 9014–19.
- Meints, K., Plunkett, K., & Harris, P. L. (1999). When does an ostrich become a bird? The role of typicality in early word comprehension. *Developmental Psychology*, 35(4), 1072–8.
- Mervis, C. B., & Pani, J. R. (1980). Acquisition of basic object categories. *Cognitive Psychology*, 12(4), 496–522.
- Murphy, G. L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press
- Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 70–84.
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

- Rips, L. J.** (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 665–81.
- Rips, L. J., Shoben, E. J., & Smith, E. E.** (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12(1), 1–20.
- Rosch, E. H., Mervis, C. B., Gray, W. D., Boyes-Braem, P., & Johnson, D. N.** (1976). Basic objects in natural categories. *Cognitive Psychology*, 8(3), 382–439.
- Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K.** (2011). Learning words in space and time: probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science*, 22(8), 1049–57.
- Taylor, M., & Gelman, S. A.** (1989). Incorporating new words into the lexicon: preliminary evidence for language hierarchies in two-year-old children. *Child Development*, 60(3), 625–36.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J.** (2004). Category norms: an updated and expanded version of Battig and Montague (1969) norms. *Journal of Memory and Language*, 50(3), 289–335.
- Waxman, S. R.** (1990). Linguistic biases and the establishment of conceptual hierarchies: evidence from preschool children. *Cognitive Development*, 5, 123–50.
- Waxman, S. R., & Hatch, T.** (1992). Beyond the basics: preschool children label objects flexibly at multiple hierarchical levels. *Journal of Child Language*, 19(1), 153–66.
- Waxman, S. R., Shipley, E. F., & Shepperson, B.** (1991). Establishing new subcategories: the role of category labels and existing knowledge. *Child Development*, 62(1), 127–38.
- Xu, F., & Denison, S.** (2009). Statistical inference and sensitivity to sampling in 11-month-old infants. *Cognition*, 112(1), 97–104.
- Xu, F., & Tenenbaum, J. B.** (2007). Word learning as Bayesian inference. *Psychological Review*, 114(2), 245–72.

Cite this article: Emberson LL, Loncar N, Mazzei C, Treves I, Goldberg AE (2019). The blowfish effect: children and adults use atypical exemplars to infer more narrow categories during word learning. *Journal of Child Language* 1–17. <https://doi.org/10.1017/S0305000919000266>