Banana or fruit? Detection and recognition across categorical levels in RSVP

Banana or fruit? Detection and recognition across categorical levels in RSVP

Mary C. Potter and Carl Erick Hagmann

Word Count: 4101

Author Note

Mary C. Potter and Carl Erick Hagmann, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Acknowledgment This research was supported by a National Institutes of Health Grant MH47432.

Correspondence concerning this article should be addressed to Mary C. Potter, Department of Brain and Cognitive Sciences, 46-4125, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. E-mail: mpotter@mit.edu

Abstract Pictured objects and scenes can be understood in a brief glimpse, but there is a debate about whether they are first encoded at the basic level, as proposed by Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976) (e.g., *banana*), or at a superordinate level (e.g., *fruit*). The level at which we first categorize an object matters in everyday situations because it determines whether we approach, avoid, or ignore the object. In the present study, we limited stimulus duration in order to explore the earliest level of object understanding. Pictured objects were presented among five other pictures using RSVP at 80, 53, 27, or 13 ms/picture. On each trial, participants viewed or heard 1 of 28 superordinate names or a corresponding basic-level name of the target. The name appeared before or after the picture sequence. Detection (as d') improved as duration increased but was significantly above chance in all conditions and for all durations. When the name was given before the sequence, d' was higher for the basic than for the superordinate name, showing that specific advance information facilitated visual encoding. In the name-after group, performance on the two category levels did not differ significantly; this suggests that encoding had occurred at the basic level during presentation, allowing the superordinate category to be inferred. We interpret the results as being consistent with the claim that the basic level is usually the entry level for object perception.

Banana or fruit? Detection and recognition across categorical levels in RSVP

In an influential set of studies, Rosch and colleagues (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) asked whether there is a preferred level at which we categorize objects. When something comes into the room, do we first think that it is a thing? An animal? A dog? Or a German Shepherd? She hypothesized that the level that most immediately comes to mind will be the most useful level—the one with the greatest increase in the number of features, relative to the more general levels above or more specific levels below. Rosch et al. termed this level the *basic* level. In a series of experiments, they showed that the basic level is the one at which people normally name the object and is the preferred level by a variety of other criteria. In particular, they showed that, when participants are asked to verify what the object is, the fastest response is usually to the basic name, rather than to a superordinate name or a more specific name.

Jolicoeur, Gluck, and Kosslyn (1984) proposed the term *entry level* to describe the level at which a visual object is first categorized. They confirmed that Rosch's basic level is indeed the entry level for most objects, although a familiar but atypical exemplar such as a peacock may be categorized first at the more specific *subordinate* level of species, rather than at the animal class of bird. Moreover, experts in a given category are likely to categorize first at a subordinate level (e.g., Tanaka & Taylor, 1991). Despite these exceptions, the claim that the basic level is the visual entry level has been widely accepted. Indeed, Grill-Spector and Kanwisher (2005) found that a stimulus that was presented very briefly and then masked could be classified at the basic level (but not the subordinate level) with the same accuracy as that at which it could be distinguished as being a coherent object. (Mack, Gauthier, Sadr, & Palmeri, 2008, however, found evidence that detection of the presence of an object and basic-level categorization can be separate processes under some conditions.)

The claim that the basic level is the perceptual entry level has been questioned more recently by studies showing that a picture can be classified as including an animal as fast as or faster than it can be classified at the basic level as a dog or a bird, for example (Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009). For some categories, perhaps those such as *animal* that would have been important in evolution, the broader *superordinate* category may be the entry level. In a further challenge to the basic-level-as-entry-level hypothesis, Rogers and Patterson (2007) found that, although healthy controls show the usual basic-level advantage, patients in advanced stages of semantic dementia categorize pictures more accurately at the superordinate level than at the basic level.

Thus, previous research does not give a clear answer to the question of whether there is a single entry level in visual perception and, if so, whether it is the basic level defined by Rosch. One reason for the lack of agreement about what categorical level is perceived first may be that most previous studies have examined only a small number of categories, sometimes only the superordinate category *animals*. Another problem is the lack of consensus about the choice of task and the criteria for the categorical level that is perceived first. One measure that has been used is the response time to name the object (either at the basic level or at the superordinate level). This requires looking up and producing the name, which takes considerably longer than making a yes–no response to the object when the experimenter gives a name before the object is presented (Potter & Faulconer, 1975). Instead of making a yes–no response, one can use a go/no-go task and measure response time. Of particular interest is the fastest above-chance bin of response times (e.g., Macé et al., 2009; Poncet, Reddy, & Fabre-Thorpe, 2012), In the present study, instead of measuring response time, we constrain the time available for processing the target object by embedding it in a rapid serial visual presentation (RSVP) sequence. The task is to verify a match between the name (basic or superordinate) and a visual object. We propose the following criteria for the categorical level that is perceived first: The entry level (1) has the lowest threshold for above-chance detection, (2) is most accurately detected when named in advance, (3) is most accurately recollected when named after presentation, and (4) meets these criteria across most categories.

Experiment 1

To vary perceptual difficulty, in the present study, the target picture was presented in an RSVP sequence of six pictures, and the presentation duration was 13, 27, 53, or 80 ms per picture (see Potter, Wyble, Hagmann, & McCourt, 2014). The target was named just before or just after the sequence, between subjects, to evaluate the effect of advance information about the target. We used 28 superordinate categories with one to seven basic-level exemplars (M = 2.3) in each category.

The main questions we addressed in this study are whether viewers can more accurately categorize an object at the basic level or at a superordinate level and whether the difference is consistent across categories, presentation duration (a proxy for perceptual difficulty), and target name position (before vs. after the sequence).

Method

Participants The 32 participants (13 women, 19 men) were paid volunteers 18–59 years of age (M = 28); 16 participated in the *before* group, and 16 in the *after* group. All signed a consent form approved by the MIT Committee on the Use of Humans as Experimental Subjects. Participants were replaced if they made more than 50% false *yes* responses, overall, on nontarget trials, because such a high false alarm rate suggests that they did not follow the instructions. One participant was replaced in the *before* group.

Materials The stimuli on the 64 target-present trials were color photographs of objects in a natural setting, representing 28 superordinate categories, with one to seven exemplars in each category. On the 24 nontarget trials, the target names came from 21 of the same superordinate categories, but the basic-level names were never the same as those on target-present trials. Table 1 shows the list of superordinate categories and basic-level names for the target and nontarget trials. The other pictures in the sequences included diverse scenes as well as objects. No pictures were repeated, and all were new to the participants. A different set of pictures was used on the practice trials. The pictures were modified from those used by Potter, Wyble, Pandav, and Olejarczyk (2010). They were taken from the Web and from other collections of pictures available for research use. Pictures were resized to 300 × 200 pixels and were presented in the center of the monitor on a gray background. The horizontal visual angle was 10.3^{o} at the normal viewing distance of 50 cm.

Design and procedure Participants viewed an RSVP sequence of six pictures and tried to detect a target object specified by a written name. In the *before* group, each trial began with a fixation cross for 500 ms, followed by the name of the target for 700 ms, a 200-ms blank screen, and then the sequence of pictures. A 200-ms blank screen followed the sequence, and then the question "Yes or No?" appeared and remained in view until the participant pressed Y or N on the keyboard to report whether he or she had seen the target. In the *after* group, the trial began with a fixation cross for 500 ms, followed by a blank screen for 200 ms and the sequence. At the end of the sequence, there was a 200-ms blank screen, and then the name was presented simultaneously with the yes–no question until the participant responded.

There were two name conditions within subjects; in the basic name condition, the target was specified by the basic-level name (e.g., *tiger, car, pineapple, sofa*), and in the superordinate condition, by the superordinate name (e.g., *four-footed animal, vehicle, fruit, furniture*, respectively). Participants had one name condition in the first half and the other in the second half of the experiment; which condition came first was counterbalanced across subjects, within group.

In all conditions, a correct *yes* response was followed by a request to write something about the target: the specific name of the target (in the superordinate condition) or some additional description of the target (in the basic condition). When the participant finished typing and pressed the return key, the next trial began. After target-present trials on which the response was *no*, there was a prompt to press any key to begin the next trial. After responses on trials with no target, the words “no target” appeared in the center of the screen, followed by the prompt for the next trial.

Each half of the experiment began with a practice block presented for 133 ms per picture, followed by four experimental blocks for durations of 80, 53, 27, and 13 ms. Each experimental block included 8 target-present trials and 3 target-absent trials, for a total of 32 target-present and 12 target-absent trials in each half. Targets were never the first or last picture in the stream. Across every 8 participants in each group, the eight blocks of trials were rotated so that the pictures in each block of trials were seen equally often at each duration and in each half of the experiment.

Apparatus The experiment was programmed with MATLAB 7.14 and the Psychophysics Toolbox extension (Brainard, 1997), version 3, and was run on a Mac mini with 2.4-GHz, Intel Core 2 Duo processor. The Apple 17-in. CRT monitor was set to a 1,024 × 768 resolution, with a 75-Hz refresh rate. The room was normally illuminated. Timing precision and stimulus presentation were controlled with the Stream package for MATLAB. Trials containing an absolute timing error of 3 ms or greater that affected the target picture or the pictures immediately before and after the target were excluded from the analysis. This resulted in the rejection of, at most, two trials per participant.

Analyses In the yes–no detection task, d' measures were computed for each participant in each condition. Hits were counted independently of subsequent written identification or description. To account for durations with zero hits or misses, a log linear adjustment was applied to all conditions by adding 0.5 to the count of hits/false alarms and 1 to the number of trials before the d' calculation (Hautus, 1995). Repeated measures analyses of variance (ANOVAs) were carried out on d' as a function of before/after group, two categorical levels, and four presentation durations. Planned paired t-tests for each group, category condition, and duration compared d' with chance (0.0). An ANOVA of the proportion of correct written responses on target-present trials was also carried out.

Results and discussion

Yes–no accuracy (d′) As is shown in Fig. 2, *d′* was higher when the target name was basic level (M = 1.30) rather than superordinate (M = 1.08), $F(1, 30) = 18.14$, $p < .001$, $\eta_G^2 = .025$, and it was higher when the presentation duration was longer, $F(3, 90) = 76.57$, $p < .001$, $\eta_G^2 = .354$. Accuracy was also higher when the target name was provided before the sequence (M = 1.31) rather than after (M = 1.07), but this effect was not significant, $F(1, 30) = 1.54$, $p = .23$, $\eta_G^2 = .029$. The interaction of *before*/*after* group and categorical level approached significance, $F(1, 30) = 3.32$, $p = .078$, $\eta_G^2 = .005$; no other interactions were significant. Separate analyses of the *before* and *after* groups showed a sizable advantage of the basic level in the *before* group, $F(1, 15) = 13.78$, $p = .002$, $\eta_G^2 = .064$, and a marginal advantage in the *after* group, $F(1, 15) = 4.5$, $p = .051$, $\eta_G^2 = .006$. One-sample t-tests comparing *d′* with chance revealed that participants were significantly above chance at all durations in all conditions, *t*s(15) > 2.6, *p*s < .01.

In the *after* condition, the participant viewed the six pictures without any idea of which one (if any) was the target, so the response had to be based on what he or she remembered seeing. If he or she remembered only a superordinate representation (e.g., a four-footed animal), he or she would be able to say *yes* to that superordinate name but would have to guess or omit writing the basic name. If the participant remembered seeing a cow, he or she could correctly say *yes* whether the probe was “four-footed animal” or “cow.” That is, the superordinate name could be easily inferred from a basic-level understanding, but not vice versa.

Thus, in the superordinate *after* condition, if the superordinate category was perceived first, there should have been many trials on which the participant said *yes* correctly but was unable to write the basic-level name.

Differences among categories Although the experiment was not designed to evaluate differences among superordinate categories, we looked at the proportion of correct *yes* responses to items in the 28 superordinate categories separately for the superordinate name and the basic-level name. The *before* and *after* groups were combined. The results are shown in Fig. 3, ordered by accuracy in responding *yes* to the basic-level name. An advantage of the basic-level condition was observed for 19 of the 28 superordinate categories. A further analysis looked at basic versus superordinate accuracy for each of the 64 targets. Each target had been seen by 16 participants in the superordinate condition and 16 in the basic condition, collapsed across the four durations and *before* and *after* conditions. Mean detection accuracy (correct *yes* responses) was 47% in the superordinate condition and 55% in the basic condition. Of the 64 items, 41 were more accurate in the basic condition, 16 were more accurate in the superordinate condition, and 7 were equal.

Names and descriptions When participants had said *yes* correctly, they were asked to write the name of the target (in the superordinate condition) or to give some additional descriptive information about the target (in the basic name condition). We carried out an ANOVA on the proportion of target-present trials on which the written name (in the superordinate condition) or the written description (in the basic-level condition) was correct (regardless of whether the participant had said *yes*), with *before/after*, level, and duration as variables. There was a marked main effect of duration (M = .19, .24, .49, and .66 for the four durations), $F(3, 90) = 93.87$, $p < .001$, $\eta_G^2 = .409$, and a marginal effect of level (basic, correct description, M = .43; superordinate, correct basic name, M = .37), $F(1, 30) = 4.09$, $p = .052$, $\eta_G^2 = .016$. Neither the main effect of *before/after* (M = .42 and .37, respectively) nor any interaction was significant.

For the superordinate condition, inspection of the wrong names given on correct *yes* trials showed that most were guesses within the designated superordinate category but unrelated to the actual pictured item (e.g., given the name "four-footed animal," writing *sheep* or *zebra* for the target, a pig). In the superordinate condition, the proportion of such guesses at each duration, relative to the total number of target trials, was similar to the proportion of false *yeses* on nontarget trials, suggesting that the wrong names were given because the participant had guessed *yes* without having seen the target.

Before discussing the implications of the results for the question of whether the perceptual entry level is at the basic or superordinate level, we present the results of Experiment 2.

Experiment 2: Auditory target cues

A second objective of our study was to determine whether the sensory modality of the target name affected detection or description of target pictures. Short-term memory for recent spoken words is more accurate than that for written words (the modality effect; e.g., Crowder, 1986), so speaking the target name might increase retention and, thus, improve detection. We therefore replicated the *before* condition of Experiment 1 with spoken, instead of written, word cues.

Method

The method was the same as that of Experiment 1's *before* condition, except as indicated.

Participants The 16 participants (8 women, 8 men) were volunteers 18–35 years of age (M = 24.4). No participants had to be replaced for making more than 50% false *yes* responses on nontarget trials.

Materials Participants heard the target name through Sony MDR V-150 headphones (Sony, Japan), instead of seeing it. Audio was recorded in MATLAB with author Hagmann's voice spoken normally and was extracted into WAV files at a bit rate of 1,411 kbps, a sample size of 16 bits, and a sample rate of 44.1 kHz.

Design and procedure Auditory target names were presented beginning 2 s prior to the 200-ms blank screen preceding RSVP. No auditory files were longer than 2 s.

Results and discussion

The results were compared with those of Experiment 1's written name-before group, as shown in Fig. 4. In a modality × duration × level ANOVA of d', we observed no effect of modality and no interactions with modality (all Fs < 1.0). We found significant effects of level, $F(1, 30) = 53.74$, $p < .001$, $\eta_G^2 = .10$, with d' higher for basic-level names (M = 1.51) than for superordinate names (M = 1.15), and of duration, $F(3, 90) = 149.4$, $p < .001$, $\eta_G^2 = .58$. Detection at the basic level was significantly above that for the superordinate level at all durations, $p < .048$ (planned paired t-tests). No other effects or interactions had Fs > 1. The analyses of written names and descriptions gave results like those in Experiment 1's *before* condition; there were no main effects or interactions with modality.

In sum, spoken names in Experiment 2 gave results that were highly similar to those for the *before* group in Experiment 1. Because accuracy was no higher in Experiment 2, apparently it was no easier or harder to encode and retain target information from the spoken than from the written name.

General discussion

The main question we addressed was whether the entry level for object perception is the superordinate or the basic level. We proposed four criteria for the category level that is first perceived. The first criterion is that the entry level is the level with the lowest temporal threshold for above-chance performance. In the present experiments, the basic level was more accurate at 13 ms per picture than the superordinate level, although the effect was significant only when the target was named before the sequence. The second criterion is that the entry level is the level most accurately detected when the target is named before the sequence. For each of the durations in the *before* condition, the basic level was more accurate than the superordinate level, although the effect was not significant at 27 ms in Experiment 1 (in Experiment 2, the basic-level advantage was significant at all durations). The third criterion is that the entry level is the most accurately recollected when named after the sequence. Although the mean accuracy results favored the basic condition (p = .051), at none of the durations was the difference significant (we discuss this result in more detail below). The fourth criterion is that the entry level meets these criteria across most categories. In the present study, we sampled 28 superordinate categories with a total of 64 exemplars, and we found that the basic level had more correct *yeses* (combining before and after groups and all durations) for 19 of the 28 superordinate categories in Experiment 1 and 23 of the 28 superordinate categories in Experiment 2.

In the yes–no detection task, the basic-name advantage was substantial when the target was specified before viewing the sequence, indicating that having a specific target name in mind aided detection. When the target was named after the sequence, there was also a small but not significant basic-name advantage. If the perceptual entry level were at the superordinate level, we would expect to find trials in the superordinate-name-after condition in which a correct *yes* response was given but the target was perceived only at the superordinate level, causing the participant to write an incorrect basic-level name of the object. Indeed, we did find such trials, especially at the two shortest durations, 13 and 27 ms. However, the proportion of such trials matched fairly closely the proportion of false *yeses* at each duration, suggesting instead that many of these incorrect names were given because the participant had simply guessed *yes* with no actual perception of the target and was then forced to guess a name when prompted. Moreover, the mistaken names rarely corresponded with the visual properties of the target, as one might expect if the response were based on partial perception.

Is it possible (as suggested by a reviewer) that a superordinate representation is activated by the stimulus before the basic representation is activated but is overwritten as further evidence about the basic level is activated, so that only the basic level is consciously perceived? Our experiments cannot rule out this possibility, although the lack of an advantage for the superordinate name in the *before* condition seems inconsistent with this hypothesis.

Since detection was clearly more difficult the shorter the duration, the d' advantage of the basic-level name (the entry-level effect) might be expected to be more marked at shorter durations. The opposite result might be expected if perceptual information is degraded by presenting stimuli briefly, providing sufficient information only for superordinate classification (e.g., Rogers & Patterson, 2007). Neither of these predictions was correct: The effect of category level did not interact with duration, consistent with Grill-Spector and Kanwisher's (2005) results.

Theories of visual perception almost universally assume that visual features are analyzed in order to identify an object. One prominent theory, Treisman and Gelade's (1980) feature integration theory, proposes that features of an object are bound together by attention. However, individual salient features might activate a superordinate category before being bound. For example, features such as beaks, eyes, or feathers might preemptively activate the category *bird* or *animal* before binding. Evans and Treisman (2005) suggested that superordinate categorization of pictured animals presented as briefly as 100 ms or less might be based on such unbound features, whereas a much longer duration would be required to bind the features and identify the object at the basic level. However, Potter et al. (2010) found that viewers were able to detect and then identify at the basic level two successive exemplars of a superordinate target category presented in an RSVP sequence at 107 ms per picture, showing that objects could be fully identified at the basic level with a brief presentation. The present results also make it clear that sufficient features and associations with a name can converge to activate a basic-level category even with sequences of objects presented still more briefly, although only a fraction of the targets can be detected.

Unlike much of the previous research on category levels, participants in the present study could not anticipate the relevant target category until it was named less than a second before the sequence began, or only after the sequence. The large number of categories used, with no repetition of any basic-level categories, prevented participants from developing an attentional set for any particular objects or object categories.

A limitation of the present study is that in the basic-level condition, participants were not forced to discriminate between one basic-level item and another in the same superordinate category. That is, if the basic target was *dog*, we never showed them another four-footed animal, such as a fox or a horse, that would have forced them to discriminate between that animal and a dog. In consequence, the participant could have detected the target *dog* using a superordinate category (*foot-footed animal* or just *animal*), without having to determine whether the target matched the specified basic-level category. Had participants adopted that strategy in the present experiment, there should have been no advantage for the basic over the superordinate name, contrary to what we found.

A second limitation is that there may not be consensus on whether the names we used are all correctly designated as superordinate and basic level. The 28 superordinate names were not all terms in common use (e.g., *carrying item*, *personal hygiene item*). Separating animals into groups like birds and marine animals reduced the range of exemplars, but that should have made superordinate detection easier. Similarly, *bird*, *flower*, and *insect* are sometimes considered to be basic-level terms, with *duck*, *sunflower*, and *ladybug* as respective subordinate terms. Previous research (e.g., Grill-Spector & Kanwisher, 2005; Tanaka & Taylor, 1991) has shown that subordinate categories are perceived more slowly and less accurately than basic-level categories, so if some of our "superordinate" categories were actually at the basic level and their basic-level exemplars actually subordinate, that would have worked against the basic-level advantage that we found. In any case, the names we designated as basic level are all common names for the objects we presented, and the names designated as superordinate are all higher in the hierarchy than the basic-level names.

In conclusion, the preponderance of the evidence from the present experiments supports Rosch's original hypothesis that the basic-level concept is the level at which we first understand a visual object, even with extremely brief, masked exposure to the object.

References

Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *12*, 268–278. doi:10.1037/0278-7393.12.2.268

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1476–1492. doi:10.1037/0096-1523.31.6.1476

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: as soon as you know it is there, you know what it is. *Psychological Science*, *16*(2), 152–60. doi:10.1111/j.0956-7976.2005.00796.x

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*(1), 46–51. doi:10.3758/BF03203619

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275. doi:10.1016/0010-0285(84)90009-4

Macé, M. J.-M., Joubert, O. R., Nespoulous, J.-L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: you spot the animal faster than the bird. *PloS One*, *4*(6), e5927. doi:10.1371/journal.pone.0005927

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*(1), 28–35. doi:10.3758/PBR.15.1.28

Poncet M., L. Reddy, M. Fabre-Thorpe (2012), Presentation time does not affect superordinate-level advantage in ultra-rapid categorization, poster presented at the 11th VSS conference, Naples, Florida

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception & Psychophysics*, *76*(2), 270–279. doi:10.3758/s13414-013-0605-z

Potter, M. C., Wyble, B., Pandav, R., & Olejarczyk, J. (2010). Picture detection in rapid serial visual presentation: Features or identity? *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1486–1494. doi:10.1037/a0018730

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*, 451–469.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*. doi:10.1016/0010-0285(76)90013-X

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, *23*(3), 457–482. doi:10.1016/0010-0285(91)90016-H.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. doi:10.1016/0010-0285(80)90005-5

Table 1 Target and nontarget trial names

Superordinate	Basic target	Basic-non target
Amusement ride	ferris wheel	
Art supply	color pencils	paint brush
	crayons	
Baby Product	crib	diaper
Bathroom utility	sink	
	toilet	
Bird	chicken	stork
	duck	
	parrot	
	peacock	
Body part	hands	elbow
	tongue	
Carrying item	basket	purse
Cleaning product	broom	soap
Computer part	keyboard	monitor
Dessert	cookies	cake
	ice cream	
Dinner food	hamburger	
	pizza	
Flower	sunflower	
Footwear	running shoes	
Four-footed animal	bear	cat
	dog	horse
	lion	
	panda	
	pig	
	tiger	
	zebra	
Fruit	bananas	cantaloupe
	grapes	pineapple
	strawberries	
	watermelon	
Furniture	chair	stool
	sofa	
Insect	ladybug	
Marine animal	dolphin	shark
	killer whale	
	sea horse	
	seal	
Musical instrument	guitar	saxophone
	violin	
Personal hygiene item	toothbrush	hairbrush
Reptile	alligator	iguana
Sports equipment	basketball net	football helmet
	basketball	football
	hockey net	
	soccer ball	
Tableware	cup	spoon
	fork	
Tool	hammer	
	rake	
Toy	slinky	yo-yo
	teddy bears	
Vegetable	cabbage	artichoke
	onions	radish
	peas	
	peppers	
	tomatoes	
Vehicle	bus	jetski
	car	
	helicopter	
	motorcycle	
Weapon	knife	
	sword	

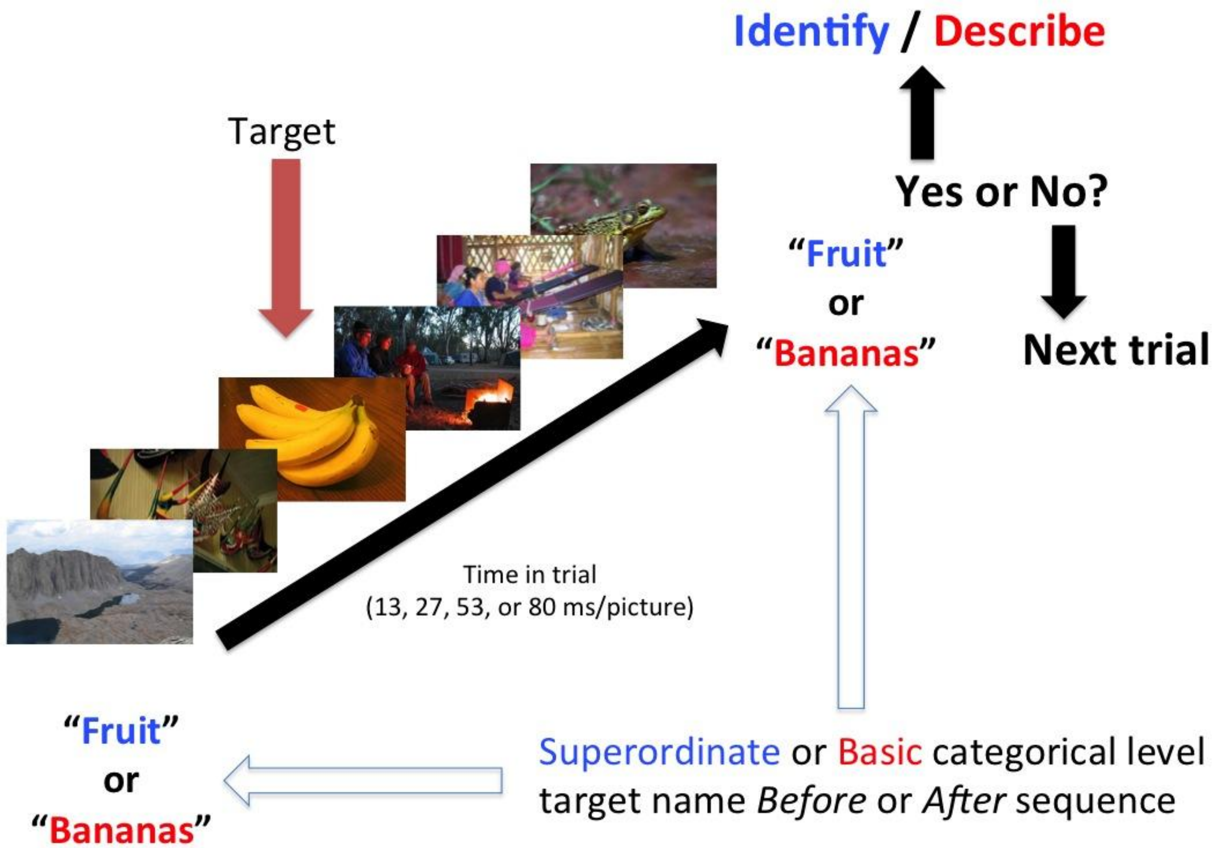
Figure Captions

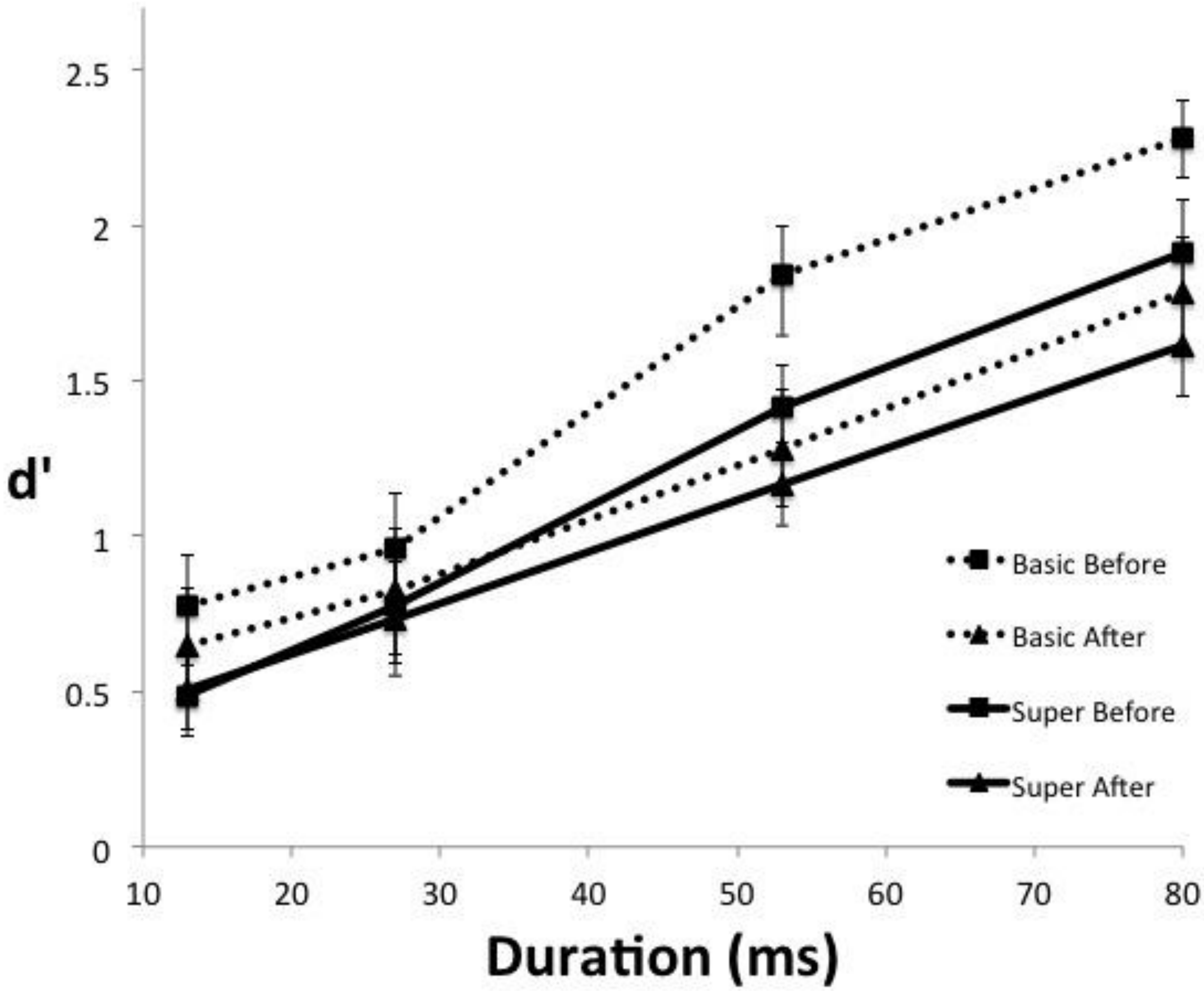
Fig. 1 Illustration of a trial with a superordinate (blue) and basic (red) target name given before or after a sequence of six color pictures. Correct detection resulted in a request for a written identification (superordinate) or description (basic) of the target object. Otherwise, the trial ended

Fig. 2 Yes–no accuracy (*d′*) at each presentation duration for each of the conditions (basic or superordinate names) in each of the groups (name before or after the sequence). Error bars represent standard errors of the means

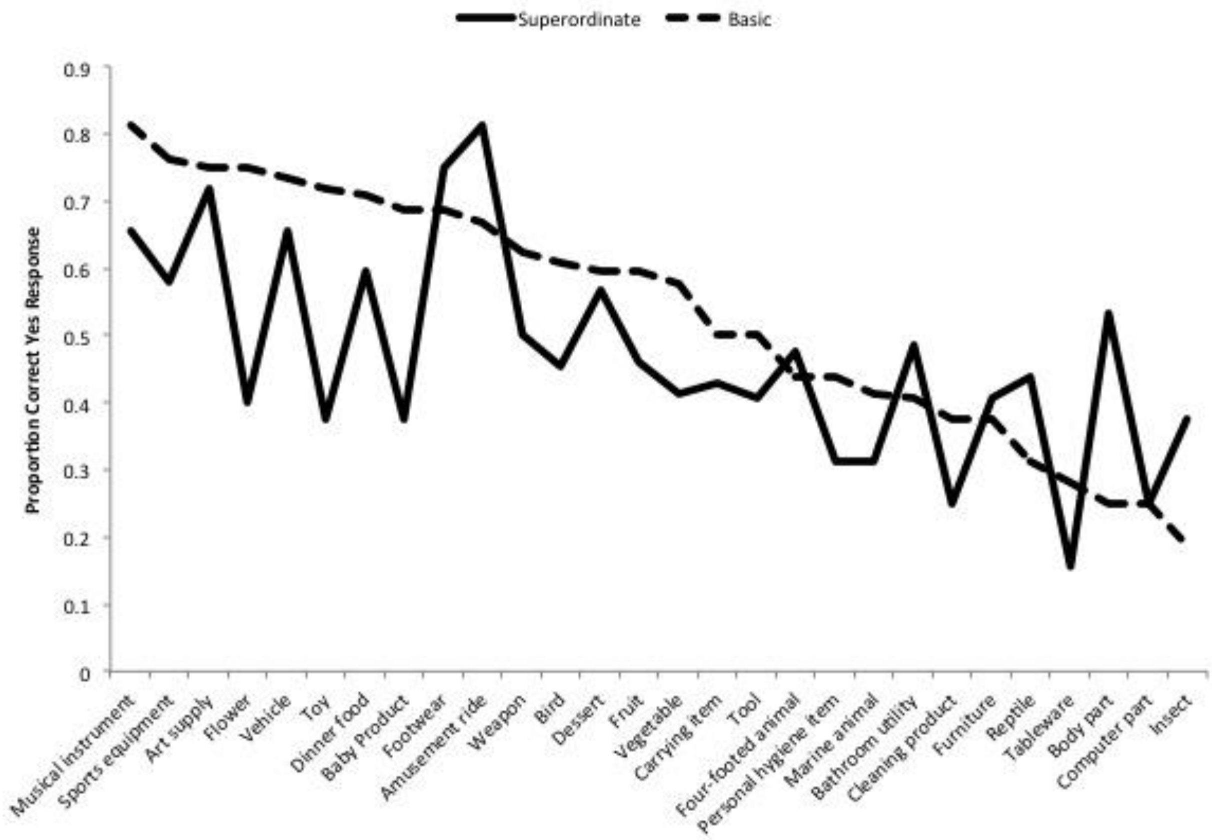
Fig. 3 Proportions of correct *yes* responses for items in each of 28 superordinate categories, separately for the basic-level name (dotted line) and the superordinate name (solid line), combining the before and after conditions and all four durations. The categories are ordered left to right with respect to performance in the basic-name condition. Each data point was based on between 16 and 112 trials, depending on the number of items in each category (one to seven: see Table 1)

Fig. 4 Yes–no accuracy (*d′*) at each presentation duration for each of the conditions (basic or superordinate names) in Experiment 2 with a spoken target name presented before each sequence, as compared with Experiment 1's group with a written name before the sequence. Error bars represent standard errors of the means


Identify / Describe
Target
Yes or No?
"Fruit"
or
"Bananas"
Next trial
Time in trial
(13, 27, 53, or 80 ms/picture)
"Fruit"
or
"Bananas"
Superordinate or Basic categorical level
target name *Before* or *After* sequence



d'
2.5
2
1.5
1
0.5
0
10
20
30
40
50
60
70
80
Duration (ms)
Basic Before
Basic After
Super Before
Super After


Superordinate
Basic
Proportion Correct Yes Response
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Musical instrument
Sports equipment
Art supply
Flower
Vehicle
Toy
Dinner food
Baby Product
Footwear
Amusement ride
Weapon
Bird
Dessert
Fruit
Vegetable
Carrying item
Tool
Four-footed animal
Personal hygiene item
Marine animal
Bathroom utility
Cleaning product
Furniture
Reptile
Tableware
Body part
Computer part
Insect

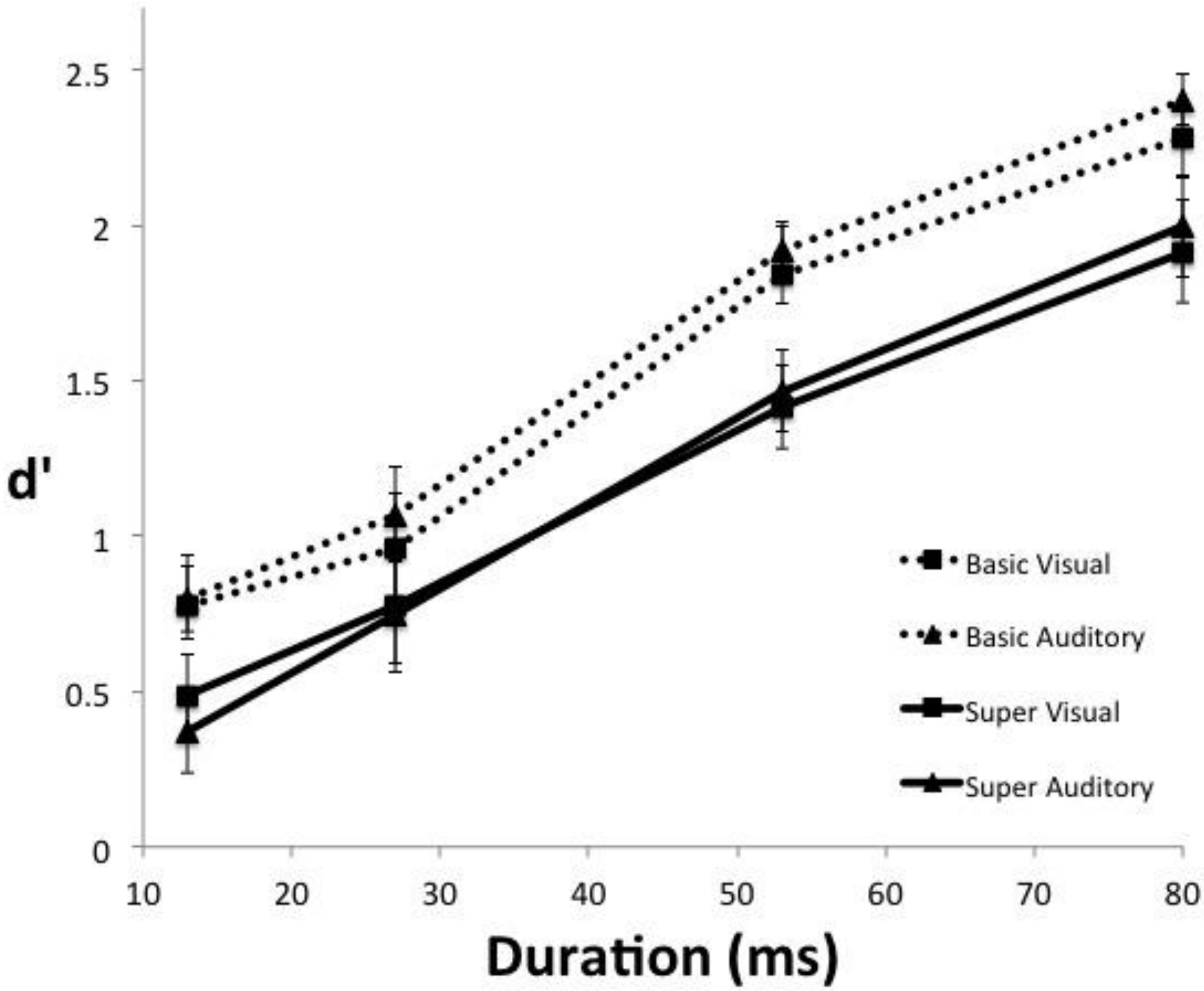

2.5
2
1.5
1
0.5
0
d'
10
20
30
40
50
60
70
80
Duration (ms)
Basic Visual
Basic Auditory
Super Visual
Super Auditory

Table 1 Target and nontarget trial names

Superordinate	Basic–Target	Basic–No Target
Amusement ride	Ferris wheel	
Art supply	color pencils	paint brush
	crayons	
Baby product	crib	diaper
Bathroom utility	sink	
	toilet	
Bird	chicken	stork
	duck	
	parrot	
	peacock	
Body part	hands	elbow
	tongue	
Carrying item	basket	purse
Cleaning product	broom	soap
Computer part	keyboard	monitor
Dessert	cookies	cake
	ice cream	
Dinner food	hamburger	
	pizza	
Flower	sunflower	
Footwear	running shoes	
Four-footed animal	bear	cat
	dog	horse
	lion	
	panda	
	pig	
	tiger	
	zebra	
Fruit	bananas	cantaloupe

	grapes	pineapple
	strawberries	
	watermelon	
Furniture	chair	stool
	sofa	
insect	ladybug	
Marine animal	dolphin	shark
	killer whale	
	sea horse	
	seal	
Musical instrument	guitar	saxophone
	violin	
Personal hygiene item	toothbrush	hairbrush
Reptile	alligator	iguana
Sport equipment	basketball net	football helmet
	basketball	football
	hockey net	
	soccer ball	
Tableware	cup	spoon
	fork	
Tool	hammer	
	rake	
Toy	slinky	yo-yo
	teddy bears	
Vegetable	cabbage	artichoke
	onions	radish
	peas	
	peppers	
	tomatoes	
Vehicle	bus	jetski
	car	

	helicopter	
	motorcycle	
Weapon	knife	
	sword	