How World Knowledge Shifts Adjective Interpretation

Sara Kessler (skessler@stanford.edu)
Department of Linguistics
Stanford University

Abstract

Dimensional adjective interpretation is dependent on the comparison class – the set of object representations – against which the object being modified by the adjective is judged. This paper explores the factors determining the composition of the comparison class, arguing that real world size information and prototypicality play crucial parts in its determination. Researchers often implicitly assume that only the objects in immediate visual context constitute the comparison class. However, Exp. 1 shows that this information from the visual context is integrated with knowledge of real world size and category properties to form the comparison class. Exp. 2 shows that prototype information is utilized when making size judgments of cartoon images, while size judgments of objects in photographs draw more heavily on a speaker's prior knowledge about the actual size of the objects in the world. Exp. 3 demonstrates that the effects observed in Exp. 1 and 2 were not caused by the adjectives used, but rather reflect differences between the size of the objects depicted in the images.

Keywords: semantics; pragmatics; adjectives; context; gradability; scale structure; prototype effects; size perception

Introduction

The interpretation of dimensional adjectives like *big*, *small*, *tall*, and *short* is dependent on context. The size an object needs to be to count as, e.g., *big* is different depending on the type of object and the other objects in the context. For instance, *big* for an elephant is an entirely different scale of *big* than *big* for a mouse. Similarly, the same elephant could count as *small* when standing with one group of elephants but no longer count as *small*, or even count as *big*, when amongst a different group. These observations indicate that the standard for what counts as *big* depends on the set that the object is being compared to, known as a comparison class. Speakers succeed in converging on similar enough comparison classes to successfully communicate when using adjectives, but the exact nature of a comparison class and the way in which it is contingent on the properties of the modified object have not yet been adequately explored. This paper argues that speakers utilize both the objects in the immediate visual context in which the adjective is uttered and prior knowledge about the relevant category of object in order to determine the comparison class, and therefore interpret a dimensional adjective.

A comparison class (Cresswell, 1976; Kennedy, 1999, 2007; Klein, 1980) can be thought of as as an underlying distribution of degrees of the property denoted by the adjective, within the category of object (Lassiter & Goodman, 2013, 2015). For example, in order to determine which elephants count as *big*, it is necessary to figure out what the distribution of the property (size) is among members of the class (relevant elephants). Then, some statistical function determines the standard of membership, i.e. the degree of size which an object must exceed in order to count as *big*. One proposal is that the standard is determined by the population mean, so that to be *a big elephant*, the elephant must exceed the mean size of all elephants (Bierwisch, 1989). Various experimental work has been done regarding the formula involved in determining the standard of membership (Barner & Snedeker, 2008; Hansen & Chemla, 2017; Solt & Gotzner, 2012a, 2012b).

However, even were a formula to be arrived at with certainty, the result of its application would still be dependent on the determination of the comparison class to which it is applied. As people experience the world around them, and interact with objects and depictions of objects, they build up mental representations of distributions for various properties of those objects, including their size. Thus, when encountering a new token of that object, its size can be compared to the distribution of sizes of relevant similar objects to determine whether its size exceeds the standard for counting as *big*.

Adjectives, including dimensional adjectives, demonstrate prototypicality effects, similar to nouns (Dirven, 1988, e.g.). Thus, elephants can be said, generally, to be *big*.[1] These prototypicality patterns are also part of a speaker's knowledge about objects which might influence the comparison class.

Although there has been an implicit assumption in much of the research on adjective judgments that only the objects in the immediate visual context determine the comparison class, there are reasons to suspect otherwise. Initial support for the idea that other information is also used when determining the comparison class is found in Tribushinina (2011). She argues for the integration of real world size knowledge with information from the immediate visual context. However, this evidence is only the beginning. Knowledge of real world size is conflated there with knowledge about items being classified as prototypical instances of *big* and *small* objects, and the stimuli used were clip-art images, which may strengthen this association. Clip-art images emphasize salient, prototypical features, and may vary widely from the actual appearance of the objects depicted, whereas a photograph depicts an object that exists in the real world and had properties with specific values attached to it. Therefore, clip-art may trigger access to prototypical, generalized information about object categories, rather than information about the actual distribution of properties found among these objects in the world. This hypothesis will be further probed to further determine the influence of world knowledge of various types on the comparison class.

[1] This is most noticeable in child directed speech and children's books. Children are in the process of building up these networks of association and forming category representations (Tribushinina, 2011).

Experiment 1: Replication and validation

The initial evidence for integration of size knowledge and information from the immediate visual context comes from the first experiment reported in Tribushinina (2011). This experiment was conducted in Dutch and in a lab, so I conducted a replication to validate the methodological changes I implemented (see the discussion section for more details). Each trial displays a series of images, e.g. elephants, and the participant is asked " Which elephants are big/small?" There are two main predictions. Since all of the images are smaller than the objects they depict, they might all be considered small if only world knowledge were used to determine the comparison class. If only information from the visual context were used, symmetry between *big* and *small* would predict an equal number of items considered *big* as considered *small*.[2] If people integrate both types of information, I predict that there will be an asymmetry between the adjectives such that the mean number of images selected as *big* (the *big* zone) will be **smaller** than the mean number of images selected as *small* (the *small* zone). Additionally, the depicted objects differed in whether they were considered prototypically big, prototypically small or neither prototypically big nor prototypically small (designated here as neutral), but the images themselves were all of the same set of sizes, regardless of the object they depicted. Therefore, any differences in the number of items counted as *big* or as *small* between objects can be attributed to the influence of knowledge about their size, rather then to different size information from the visual context. Thus, prototypically big objects are predicted to have a smaller *big* zone and a bigger *small* zone than prototypically small objects. Neutral objects are predicted to fall in the middle of the other two categories.

Methods

Participants I recruited 57 people on-line using Amazon's Mechanical Turk (MTurk).

Materials and Design Each trial of the experiment consisted of a slide containing seven cartoon images of the same object, each scaled to a different size. The images were arranged by size, in either ascending or descending order. They ranged from 38 to 265 pixels in width, increasing by 38 pixels from image to image (Figure 1). Since the experiment was conducted using MTurk, the actual size of the slides that participants saw varied between participants. The images scaled to fit their displays, but were consistent relative to each other across participants. The images used were identical to those used in Tribushinina (2011) and acquired from that author. There were twelve different images, four of prototypically big objects (elephant, hippo, house, plane), four of prototypically small objects (baby, mouse, chick, gnome), and four of neutral objects (monkey, umbrella, cake, balloon)[3]. As the slide was displayed the participant heard the question in (1). The audio stimuli were created using an online text-to-speech generator[4] in order to maintain consistent prosody across all of the trials.

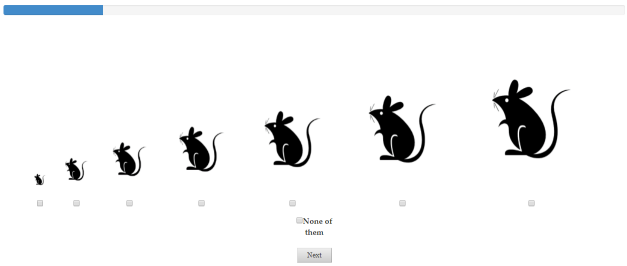

Figure 1: Sample trial slide, showing mice in ascending order, accompanied by the question "Which mice do you find big?"

(1) Which [object]s are [big/small]?

Each image could be selected either by clicking on the image or on the checkbox underneath it.

Procedure The experiment consisted of two introductory slides, followed by 48 trials, followed by a brief demographic questionnaire. The two introductory slides used objects (flower, car) and adjectives (*pretty*, *ugly*) not in the trials. The purpose of these questions was to introduce participants to the task and make them aware that they could select multiple answers. In both the introductory phase and in the trials, as each slide was displayed, the participant heard the audio prompt, and could not select any images until it was done playing. The participant was required to make a selection (either at least one of the images or the "None of the above" box), before clicking on the "Next" button to advance to the next trial. At the end of the experiment, participants were asked about their native language and the size of the screen they used for the experiment. Each participant saw each object, in each order (ascending or descending), for each adjective (*big* or *small*) – 48 trials in total. The trials appeared in a randomized order with the caveat that the same object not appear twice in a row.

Exclusions A number of criteria were used to ensure that the participants were cooperating with the task. All participants who reported using a screen smaller than 12 inches were excluded from analysis, as were those who reported a native language other than English. Due to the scale structure of *big* and *small*, when asked to either select the big objects or the small objects, participants were expected to include the appropriate extreme image (either the biggest or

[2] It is not necessary for my argument to assume symmetry between *big* and *small*, though Tribushinina (2011) does take this stance. If world knowledge is not taken into account, the *big* zone should be the same across all objects, as they are the same size. This is shown not to be the case, both in Exps. 1 and 2, and even more so if these are compared to Exp. 3.

[3] For details on how the prototypicality status was determined, see Tribushinina (2011). I am assuming that Dutch and English speakers are sociologically similar enough that the protoypicality status of the items in Dutch is retained in English. The results of this experiment provide support for this assumption.

[4] https://text-to-speech-demo.mybluemix.net/

the smallest). If they failed to do so more than 10% of the time were excluded. Additionally, participants who provided non-consecutive responses[5] more than 10% of the time were excluded. Also excluded were all individual trials in which no appropriate endpoint was selected or in which non-consecutive images were selected.

Results

In Exp. 1, two participants were excluded for screen size, two for language and three for not marking an endpoint on more than ten percent of trials. Additionally, 26 individual trials were excluded,[6], leaving a data set of 2373 trials from 50 participants. Despite the methodological differences, I successfully replicated the results from (Tribushinina, 2011). The critical measure compares the mean number of images counted as *big* (the *big* zone) in each trial to the mean number of images counted as *small* (the *small* zone). As predicted, the *big* zone (mean = 2.11, sd = 0.81) is indeed smaller than the *small* zone (mean = 2.87, sd = 1.09). A Wilcoxon signed-rank test (SRT) shows that this difference is significant, $Z = -21.44, p < 0.001$. Additionally, Friedman tests showed that there were differences within the *big* zone (Friedman $\chi^2 = 28.15, df = 2, p < 0.001$) and within the *small* zone (Friedman $\chi^2 = 32.56, df = 2, p < 0.001$) between objects of different prototypicality status (Figure 2).[7] As predicted, the *big* zone is significantly smaller for big objects than for neutral objects (Wilcoxon SRT, $Z = -2.23, p = 0.02$), for neutral objects than small objects (Wilcoxon SRT, $z = -3.57, p < 0.001$), and for big objects than small objects (Wilcoxon SRT, $Z = -4.37, p < 0.001$). For a prototypically big object, fewer images were selected as *big*, than for a prototypically small object, meaning that in order to count as *big*, prototypically big objects needed to be bigger than prototypically small ones did. Thus, it is harder to count as a *big* object if you are prototypically big, and easier if you are prototypically small. World knowledge is influencing adjectival judgments. The inverse pattern is displayed in the *small* zone, as predicted. The *small* zone is significantly bigger for big objects than for neutral objects (Wilcoxon SRT, $Z = 3.23, p < 0.001$), for neutral objects than small objects (Wilcoxon SRT, $Z = 3.82, p < 0.001$), and for big objects than small objects (Wilcoxon SRT, $Z = 5.02, p < 0.001$). Similar to the result above, it is harder to count as a *small* object if you are prototypically small, and easier if you are prototypically big.

Discussion

These results replicate the findings in the first experiment in Tribushinina (2011), despite the methodological differences between the experiments. I conducted this experiment in English rather than in Dutch, and on MTurk rather than in a lab. The shift to MTurk required participants to perform the task on their own devices, resulting in lack of consistency of screen size, and thus stimulus size, across participants. Since they were being asked to judge size, the variation in the size of the stimuli across participants was a concern. Additionally, in the lab, there was an experimenter asking the questions, and the participants responded by pointing to the images they wanted to select. This interactional format might have led to more cooperativeness than on MTurk where participants may be multi-tasking and paying less attention to the task. On MTurk there was also the possibility of misclicking, or thinking you clicked without successfully selecting an image. These factors necessitated the exclusions described above, which were not required in the original experiment in Tribushinina (2011). Despite these changes, the results support an account where world knowledge is integrated with the information from the immediate visual context in determining the comparison class used for the adjective judgment. If the comparison class had been solely determined by the size of the images in the immediate visual context then, across objects, each zone should be the same size, since each set of images were the same range of sizes. Assuming that the formula for determining the cut-off point is fixed regardless of the distribution of items in the comparison class, then different sized zones indicate different distributions of sizes in the comparison class. Therefore, the observed difference between the size of the *big* zone for prototypically small objects and for prototypically big objects indicates that world knowledge is being used to determine the comparison class.

Exp 1 suggests that world knowledge is a significant factor in determining a comparison class, but raises the question of what type of world knowledge is being used. Tribushinina (2011) argues that it is knowledge of the size of objects in the world that is being taken into account, and probes this using prototype status. However, it is not clear that these are

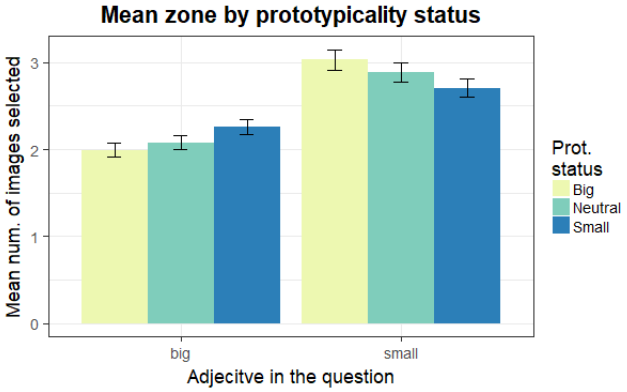

Figure 2: Mean number of objects selected as *big* and *small* by prototypicality status of the object when stimuli images were cartoons. Error bars represent 95% confidence intervals.

[5] A response was non-consecutive, for example, if for *big* the largest and third largest images were selected but the second largest was not.

[6] Twelve for having no endpoint, four for having an inappropriate endpoint and eleven for having non-consecutive images selected.

[7] The Friedman test is a non-parametric test for one-way repeated measures analysis of variance by ranks, similar to a Wilcoxon test, but can be used when there are more than two groups.

equivalent. Prototype status and real size often, but not always, overlap. For example, a baby, which is taken to be prototypically small, is actually approximately the same size as a balloon, taken to be neutral, if not bigger. The stimuli used were cartoon images. Cartoons evoke salient, stereotypical features of the items depicted. Therefore prototype status information may be more salient for cartoon images than for real pictures. On the other hand, when judging photographs of real objects, which more closely resemble the objects themselves than cartoon renderings do, prototypical size information may be down-weighted as compared to real-world size information. Exp. 2 examines these differences using real pictures rather than cartoons in a setup otherwise identical to that of Exp. 1.

Experiment 2: Photographs

Given the differences between cartoons and phototgraphs, if real-world size information is given more weight when photographs are judged than when cartoons are judged, I predict that the difference in the mean number of objects selected as *big* or as *small* between prototypically small and neutral objects will disappear. Prototypically *big* objects are still predicted to be distinguishable from the other objects since the real-world size of the prototypically big objects in this study is many times the size of the other objects. Since all images are still smaller then the objects they depict, the overall difference in size between the *big* zone and the *small* zone is not predicted to change from Exp. 1. Namely, the *big* zone is still predicted to be smaller than the *small* zone.

Methods

Participants Using MTurk, I recruited 50 people who had not participated in Exp. 1.

Materials and Design Exp. 2 used photographs rather than cartoon images. The photographs showed the same objects that appeared in Exp. 1, with two exceptions. There was a duckling instead of a chick, and a bunny instead of a mouse. The images were substituted out of concern that the largest image in the relevant set was not actually smaller than a live mouse or chick. The assumption that the images are smaller than the objects depicted in them is central to the prediction that the *big* zone will be smaller than the *small* zone. Therefore, two objects that were still prototypically small, and were similar to the original objects, but were slightly bigger than them in real life were used instead.[8]

Procedure The procedure is identical to that of Exp. 1. Each subject saw two introductory slides (photographs, unlike in Exp. 1), followed by 48 trials, in a random order, except that the same object could not appear twice in a row.

Results

Nine subjects were excluded from the analysis based on the exclusion criteria described under Exp. 1: five for using screens too small, and four for exceeding the threshold for inappropriate endpoints. Additionally, 35 trials were excluded,[9] leaving a total dataset of 1933 trials over 41 subjects. As predicted, since the images are all smaller than the objects they depict, the *big* zone (mean = 2.33, sd = 0.8) is smaller than the *small* zone (mean = 3.14, sd = 0.8). A Wilcoxon SRT shows that the difference between the zones is statistically significant, $Z = -16.16, p < 0.001$)). However, when broken down by prototypicality status the results for real pictures differ from those found for cartoon images in Exp. 1. There are still significant differences within the *big* zone (Friedman $\chi^2 = 7.38, df = 2, p = 0.025$) and marginally significant differences within the small zone (Friedman $\chi^2 = 5.05, df = 2, p = 0.08$) (Fig. 3). These differences are not the same ones seen in Exp. 1, though. The *big* zone is significantly smaller for prototypically big objects than for neutral objects (Wilcoxon SRT, $Z = -2.58, p = 0.008$), and is marginally smaller for big objects than for small objects (Wilcoxon SRT, $Z = -1.34, p$ (one-tailed)$= 0.09$).[10] However, the difference between the *big* zone for prototypically neutral objects and for prototypically small objects is not statistically significant (Wilcoxon SRT, $Z = 0.94, p = 0.35$). Within the *small* zone, a similar pattern is seen. The *small* zone is significantly bigger for prototypically big objects than for neutral objects (Wilcoxon SRT, $Z = 2.65, p = 0.007$), and marginally bigger for prototypically big than for prototypically small objects (Wilcoxon SRT, $Z = 1.78, p = 0.07$). However, the difference between the *small* zone for prototypically neutral and small objects is not significant (Wilcoxon SRT, $Z = -0.007, p = 0.99$).

Discussion

The pattern of significant differences found in (Tribushinina, 2011) and in Exp. 1, is not replicated with photographs of images. Exp. 2 demonstrates that when speakers judge the size of objects in photographs, prototypicality status is not taken into account, as opposed to when they judge cartoons. When judging photographs, the difference between prototypically small and neutral objects is no longer significant. This difference may be driven by the fact that the real world size

[8] The substitutions might raise some concern that they are in the direction supporting my hypothesis, since making the size of the prototypically small objects bigger minimizes the real-world size distinction between the prototypically small and neutral objects. Initial norming data on the stimuli indicate that chicks and ducklings are the same size, but did find bunnies to be thought of as bigger than mice. However, the phrasing of the question there might have been probing real world size more than people's conceptions of prototypicality status. Further investigation is required to distinguish the two.

[9] 22 trials were excluded for having no endpoint selected, 3 for the wrong endpoint and 10 for non-consecutive images.

[10] In a variation on this experiment, in which the order of the images for each trial was randomized, rather than ascending or descending (in order to discourage participants from forming strategies to answer the question without evaluating each display), the *big* zone was found to be significantly smaller for prototypically big objects than for small object, as well as for neutral objects, but was not significantly different for prototypically neutral and small objects.

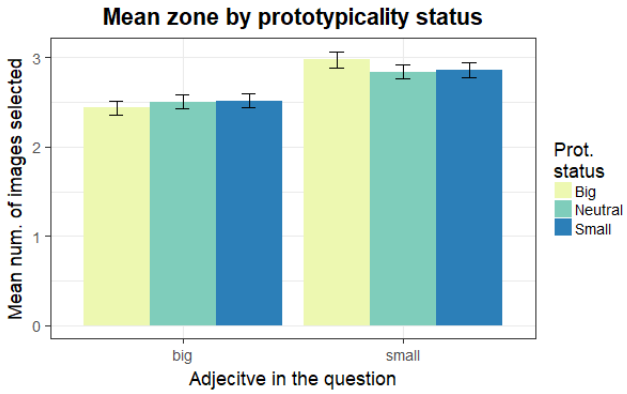

Figure 3: Mean number of objects selected as *big* and *small* by prototypicality status of the object when stimuli images were photographs. Error bars represent 95% CIs.

of the objects in these two categories is not very different, whereas it is still different from the size of the prototypically big objects. While it seems these two experiments support the theory that real world size information is integrated with information from the immediate visual context when making adjectival judgments, an alternate explanation is possible.

Experiment 3: Tiny Objects

An alternative explanation for the observed asymmetry between the *big* zone and the *small* zone relates to the difference in meaning between *big* and *small*. There is an asymmetry between *big* and *small* in that while an object can get bigger and bigger indefinitely, there is a limit on how small it can get. While the scale of size is unbounded on the upper end, it is effectively bounded on the lower end. At some point an object gets so small that the distinctions are no longer relevant. This asymmetry could mean that it is easier in general to count as small, since the lower limit of smallness is known. On the other hand, the upper limit of bigness is not, so speakers may be more reluctant to determine that an object in front of them counts as *big*. In order to rule out this possible interpretation, I conducted Exp. 3 using a paradigm similar to the one used in Exps. 1 and 2, but with stimulus images that are larger than the items they depict. I predict that if speakers are integrating real world size knowledge with the information from the immediate visual context when making adjectival judgments, then the *big* zone should be **bigger** than the *small* zone. However, if the difference between the zones is simply due to the different scale structures of *big* and *small* then the effect should be the same as in Exps. 1 and 2, and the *big* zone should remain smaller than the *small* zone.

Methods

Participants Using MTurk I recruited 49 participants who had not participated in either of the other experiments.

Materials and Design The materials for Exp. 3 are similar in set up to Exps. 1 and 2, but instead of using images which are smaller than the objects depicted in them, the images are photographs of very small objects. Each trial still consisted of seven images of an object, sized as before, arranged in order, either ascending or descending from left to right. The objects in the images were ant, bean, blueberry, candy,[11] earring, fly, hazelnut, ladybug, pill, pin, seed, tack. Since all of the objects are very small there are no prototypicality status differences between them. The audio for each display was created using the same text-to-speech generator as previously.

Procedure As in the previous experiments, there were two introductory trials (identical to Exp. 2), and then 48 trials. The trials were in a randomized order for each participant, with a caveat that objects could not appear twice in a row.

Results

Seven subjects were excluded from the analysis based on the exclusion criteria described for Experiment 1, all for exceeding the threshold for inappropriate endpoints. Additionally, 29 trials were excluded,[12] leaving a total dataset of 1987 trials over 42 subjects. As predicted if the mean zone size differences observed in Exps. 1 and 2 indicate integration of real world size information with the information from the immediate visual context, the pattern flips when the images judged are larger than the actual size of the objects depicted in them. The *big* zone (mean = 3.81, sd = 1.49) was found to be **bigger** than the *small* zone (mean = 1.54, sd = 0.63). The mean *big* and *small* zone findings are summarized in Table 1. A Wilcoxon SRT shows that the difference between the zones is statistically significant, $Z = 21.76, p < 0.001$.

Discussion

Exp. 3 demonstrates that when the stimulus images are bigger than the objects they depict, more images are selected as *big* than *small*, i.e. the *big* zone is bigger than the *small* zone. This result strongly suggests that the zone size effects demonstrated in Tribushinina (2011) and throughout this paper (Table 1), show that speakers integrate real world size information along with the immediate visual context, when making adjectival judgments. If these effects had been caused by the different scale structure of *big* and *little* then the relative size of the zones should not have reversed in Exp. 3.

General Discussion

The comparison class of an adjective determines its interpretation – which objects will have a sufficient degree of the denoted property in order to appropriately be described by the adjective. Previous research has established that the comparison class is context dependent, and this paper further specifies the nature of this context dependency. A null hypothesis is often implicitly assumed to be that only the objects in the immediate visual environment of the object being modified

[11] An image of a single candy corn piece was used, but the audio used the label "candy".

[12] 11 trials were excluded for no endpoint, 2 for the wrong endpoint and 16 for non-consecutive images.

Table 1: Comparison of the mean *big* and *small* zones in Exps. 1–3.

Experiment	Image type	Relative size	*Big* zone	*Small* zone	*Big* zone > *small* zone?
1	cartoon	object bigger than image	2.11	2.87	N
2	photograph	object bigger than image	2.33	3.14	N
3	photograph	object **smaller** than image	3.81	1.54	**Y**

by the adjective matter for determining the comparison class. However, Exp. 1 demonstrates that this hypothesis is an oversimplification. The comparison class integrates information both from the immediate visual context and from prior world knowledge. The design of this experiment equated prototype knowledge with knowledge of real size. However, these two do not always overlap.

Exp. 2 further elaborates this picture, showing that the type of images being judged influences the composition of the comparison class. When the images being judged are cartoons, then the prototypical features of the objects depicted are easily accessed. The comparison class may then be a distribution made up of generalized members of the category, accessing features such as how this category is conceived of and spoken of. These images do not depict actual individuals with real dimensions. Therefore the distribution of dimensions they access may be closer to a category average, or prototype. Those distributions are fairly narrow, with low variance. In contrast, when the images being judged are photographs of objects, which are closer in appearance to the objects depicted in them, they may render more accessible a speaker's experience with the objects themselves. This experience includes information about the real size of these objects. The distribution of dimensions in this class may be wider, with more variance. This would explain the contrast between the behavior observed in Exp. 1 and that observed in Exp. 2. Neutral objects are treated differently depending on whether they are depicted in cartoons or in photographs. When they are depicted in cartoons, the *big* and *small* zones for prototypically neutral objects diverge from those for prototypically small objects. However, when they are depicted in photographs, the *big* and *small* zones of prototypically neutral and small objects overlap, corresponding to the similarity of the real size of these objects. The distributions for prototypically small and neutral objects are distinct for cartoons since each is narrower, whereas they overlap for photographs where the distributions are wider. Therefore, prototypicality becomes more relevant to adjectival judgments for cartoons. In both these cases it seems that world knowledge of some type is being integrated with information from the immediate visual context to determine the comparison class.

Exp. 3 supports my account of the results of Exps. 1 and 2. It might have been argued that the observed asymmetry between the *big* zone and the *small* zone is due to an asymmetry in the scale structure of the adjectives themselves. If this were the case, the asymmetry should have persisted in Exp. 3. However, the flipped result in Exp. 3 argues strongly that the results are due to integration of knowledge of real world size with information from the immediate visual context.

Acknowledgments

I would like to thank Meghan Sumner, Dan Lassiter and Beth Levin for discussions of this project, Michael Frank, Cayce Hook, and Eric Smith for and Stanford Psychology for providing participant funds and help on Exp. 1, and the reviewers for their feedback.

References

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret *tall* and *short* based on the size distributions of novel noun referents. *Child Development*, *79*(3), 594-608.

Bierwisch, M. (1989). The semantics of gradation. In M. Bierwisch & E. Lang (Eds.), *Dimensional Adjectives* (pp. 71–261). Springer.

Cresswell, M. J. (1976). The semantics of degree. In B. Partee (Ed.), *Montague Grammar* (p. 261-292). Academic Press.

Dirven, J. R., Rene Taylor. (1988). The conceptualisation of vertical space in english: The case of tall. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics* (pp. 379–402). Amsterdam & Philadelphia: John Benjamins.

Hansen, N., & Chemla, E. (2017). Color adjectives, standards, and thresholds: An experimental investigation. *Linguistics and Philosophy*, 1-40. doi: 10.1007/s10988-016-9202-7

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge.

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*(1), 1-45.

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*(1), 1–45.

Lassiter, D., & Goodman, N. D. (2013). Context, scale structure, and statistics in the interpretation of positive-form adjectives. In *Semantics and Linguistic Theory* (Vol. 23, pp. 587–610).

Lassiter, D., & Goodman, N. D. (2015). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, 1-36.

Solt, S., & Gotzner, N. (2012a). Experimenting with degree. In *Proceedings of SALT* (Vol. 22, p. 353-364).

Solt, S., & Gotzner, N. (2012b). Who here is tall? Comparison classes, standards and scales. In *Pre-proceedings of the international conference linguistic evidence* (pp. 79–83). Tubingen: Eberhard Karls Universitat.

Tribushinina, E. (2011). Once again on norms and comparison classes. *Linguistics*, *49*(3), 525-553.