LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Learning Maximum Absolute Meaning Through Reasoning About Speaker Intentions

Crystal Lee [a] and Chigusa Kurumada [b]

[a]Princeton University and [b]University of Rochester

Abstract: Three experiments investigated adult learners' acquisition of a novel adjective. In English and other languages, meanings of some gradable adjectives are said to include an absolute standard of comparison (e.g., *full* means completely filled with content). However, actual usage is often imprecise, where a maximum absolute standard of comparison, strictly speaking, does not apply (e.g., a 90% full cup can be full depending on how it will be used). This creates problems for learners who acquire the absolute meaning from variable mappings between word forms and observations. We demonstrated that adult learners infer a maximum standard of comparison when they receive information about agents' intended goals that lie behind imprecise word usage. Results suggested these inferences are conditioned on the amount of visually ambiguous observations made during learning. We conclude that access to agents' intended goals allows learners to explain contextual sources of imprecision and helps the learning of a maximum absolute meaning from primarily nonabsolute observations.

Keywords absolute gradable adjectives; artificial language; linguistic relativity; word learning; intention reasoning

---

This research was supported in part by a National Science Foundation Graduate Research Fellowship awarded to the first author.

We would like to thank Florian Jaeger for assistance in conducting data analyses, Scott Grimm for thoughtful comments on the manuscript, and the members of the University of Rochester Kinder Lab for their constructive comments when we were conducting this project. We would also like to acknowledge the audience of the Boston University Conference on Language Development for helpful discussion points on an earlier version of this manuscript.

Correspondence concerning this article should be addressed to Crystal Lee, Department of Psychology, Princeton University, South Drive, Princeton, NJ 08544, USA. Email: cl40@princeton.edu

The handling editor for this article was Emma Marsden.

DOI: 10.1111/lang.12439

Introduction

Language is said to provide a basis for carving the world into meaningful events and categories (Lucy, 1997; Whorf, 1956). Learning a second language (L2), therefore, often requires a learner to "take a different perspective on the same event" (Flecken, von Stutterheim, & Carroll, 2014, p. 42). For instance, some languages, such as English, preferentially encode the manner of motion in verbs (e.g., *walk*, *skip*, *stagger*) while others, such as Spanish, focus more on its path or direction (e.g., verb equivalents of *walk towards*, *through*, *into*, such as *entrar* for *go into*; Papafragou, Hulbert, & Trueswell, 2008; Slobin, 1996; Talmy, 1983). Competent L2 users and bilinguals differentially classify concepts (Park & Ziegler, 2014; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009) and events (Athanasopoulos et al., 2015) according to the constraints of the language in which the concepts and events operate. A growing body of research has demonstrated that these typological differences influence language users' attention and interpretation of scenes and events even in some nonlinguistic tasks (Athanasopoulos & Bylund, 2012; Bylund, Athanasopoulos, & Oostendorp, 2013; Flecken et al., 2014; Kersten et al., 2010; Montero-Melis & Bylund, 2017; von Stutterheim, Andermann, Carroll, Flecken, & Schmiedtová, 2012).

This linguistic relativity raises the question: How do word learners know which aspect of an event is encoded in a novel word meaning? In particular, how do L2 learners discover a property of an object or an event that is not linguistically encoded in their native language (L1)? The current study approached this question by examining the acquisition of so-called absolute gradable adjectives such as *full*, *empty*, or *straight* (Kennedy, 2007; Kennedy & McNally, 2005; McNally, 2009). The semantic meaning of these adjectives is considered to include an absolute standard of comparison according to which a gradable property is assessed. For instance, an absolute standard of comparison for *full* is that a given container is full when it holds the maximum amount of content without spilling over.

We investigated how word learners identify an absolute standard of comparison during cross-situational learning of a new adjective—a maximum absolute gradable adjective, in particular. One major challenge associated with this process concerns variability across observable instances of the input. Objects that are described as full, for example, can be visually or conceptually highly distinct from one another (e.g., a glass, a theater, a train, or a stomach can all be full), and each of them often exhibits varying degrees of fullness (e.g., a 90%-filled glass can be felicitously labeled as full). This variability makes it difficult for learners, by simply observing objects that are described

as full, to conclude that the meaning includes a maximum standard of comparison. We hypothesized that one way in which learners acquire a maximum absolute standard of comparison is for them to explain the likely sources of the visual variability based on their comprehension of the surrounding context. In particular, learners benefit from understanding latent variables, such as an agent's intended goals and purposes regarding an observed scene or event.

Background Literature

Learning Relative and Absolute Adjectives

Although relatively limited in comparison to research on nouns and verbs, studies on the acquisition of adjectives have shown intricate relationships between lexical meanings and pragmatic contexts of word use (Groba, De Houwer, Mehnert, Rossi, & Obrig, 2018; Hall, Williams, & Bélanger, 2010; Kobayashi, 1997; Mintz, 2005; O'Neill, Topolovec, & Stern-Cavalcante, 2002). We focused on gradable adjectives, whose meanings pertain to a varying degree of a given property such as size, volume, and cost (Kennedy, 2007; Kennedy & McNally, 1999, 2005; Syrett, Kennedy, & Lidz, 2010). According to widely accepted accounts, gradable adjectives can be further categorized into two subtypes: absolute versus relative (Kennedy, 2007; Kennedy & McNally, 1999, 2005; McNally & Kennedy, 2013; Pinkal, 1995; Rotstein & Winter, 2004; Rusiecki, 1985; Syrett et al., 2010; Unger, 1975). The meaning of absolute gradable adjectives is considered to include a fixed, context-independent standard of comparison that can be either at a maximum or minimum (Kennedy, 2007; Kennedy & McNally, 2005). For instance, the truth value of a sentence such as *The glass is full* is evaluated against a maximum standard of comparison (i.e., the maximum level of fullness), and no other comparison to other glasses is necessary. Likewise, *empty* is another maximum absolute gradable adjective: A train is empty when there is no one on it. Maximum absolute gradable adjectives are thus often conceptualized as a mapping of objects onto endpoints of a closed scale (e.g., a degree of fullness; see Figure 1, Panel a).

In contrast, the positive form of a relative gradable adjective, such as *tall*, *heavy*, or *expensive*, is evaluated against a contextually determined comparison class (Bartsch & Vennemann, 1972; Kennedy, 2007; Kennedy & McNally, 1999; Klein, 1980; Ludlow, 1989; McNally, 2009; Stechow, 1984; Syrett et al., 2010; Syrett, Bradley, Kennedy, & Lidz, 2006; Toledo & Sassoon, 2011): A tall tree is taller relative to an average tree or to other trees in a contextually defined contrast set. Relative gradable adjectives are, therefore, considered to map objects onto an open-ended scale (see Figure 1, Panel b). Learning a relative

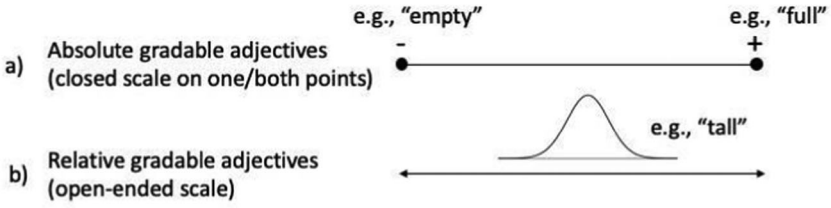

Figure 1 The hypothesized scale structure of absolute and relative gradable adjectives (modified from Syrett et al., 2010).

gradable adjective has been hypothesized to include estimating distributions of relevant values (e.g., height of instances of the same object kind; Barner & Snedeker, 2008) and evaluating a given referent in relation to an appropriate standard (Gelman & Ebeling, 1989; e.g., a mitten that is small for a boy can still be large for a miniature doll).

How absolute gradable adjectives are acquired presents a puzzle due to a phenomenon termed imprecision (Burnett, 2014; Kennedy, 2007; Pinkal, 1995; Qing & Franke, 2014; Solt, 2015; van Rooij, 2011), "informative uses of expressions in contexts in which they, strictly speaking, do not apply" (Syrett et al., 2010, p. 1). For instance, one can ask for *the full glass* when a glass is almost, but not completely, full. Similarly, the sentence *The train is empty today* can routinely refer to a situation in which a few people are on the train (Kennedy, 2007). These cases of imprecision are ubiquitous and have been analyzed in relation to language users' tolerance for language use that is, strictly speaking, false but close enough to true (Lasersohn, 1999). Put simply, a glass can be described as full when it is recognized as sufficiently full, given relevant expectations, intended goals, and constraints for the purposes at hand.

For language learners, the frequent and routinized cases of imprecision in the use of absolute gradable adjectives create two classes of problems. First, the maximum standard of comparison or the end point of a scale that learners need to identify is rarely visually observable. In reality, *full* is quite rarely used to refer to a glass that is full to the brim; it is far more often used for a glass that is only 80% or 90% full. Learners nonetheless somehow come to understand that expressions such as *The glass is full* have truth conditions that refer to a fixed (maximal) standard of comparison (e.g., as seen in Syrett et al., 2010, summarized below). Second, the mapping between a given instance of an object and the word form is often ambiguous. The same physical glass that is 90% full can be described as full or not full depending on the context (e.g., it would

have to be as near to 100% full as possible in order to adhere to a recipe). Consequently, and paradoxically, learners must acquire a maximum absolute standard of comparison from these heterogeneous visual observations, most of which, strictly speaking, do not represent an end point of the gradable adjective's scale structure.

Indeed, it has been empirically demonstrated that young learners of English differ from proficient adults in how they interpret absolute gradable adjectives in context (Syrett et al., 2010). Most relevant to the current discussion, preschoolers and adult participants in the Syrett et al. (2010) experiment heard a request such as "Give me the full jar" in the presence of two jars, neither of which was completely full. They were given three options: choosing one of the two jars or rejecting the request by choosing the response *Neither*. Although adults reliably chose the neither option, a large proportion of preschoolers selected the "more full" of the two jars. However, the study also found that preschoolers' responses became similar to adults' in the presence of more contextual information (e.g., children were more likely to reject the fuller of the two jars when they had seen a maximally full jar in a previous trial). Syrett et al. concluded that the two groups of participants were more or less equal in their knowledge about the semantic meaning of the adjectives but drew on different sources of contextual information as they provided their responses. In other words, the two groups might have critically differed in their "judgment about which of them [the utterances] counts as tolerable deviations from the actual, precise, meaning of the expression" (p. 30).

Although influential, the above conclusion left open the question about how learners might come to acquire the actual, precise meaning of the maximum absolute gradable adjective in the first place. One possibility proposed so far is that learners bootstrap the meaning from co-occurring adverbials, where "proportional modifiers (e.g., completely) tend to modify absolute maximum standard GAs (e.g., full), while intensifiers (e.g., very) tend to modify relative GAs (e.g., big)" (Syrett & Lidz, 2010, p. 258). The current study presented an additional possibility in which participants reasoned about the meaning of a new word through inferring the intention and the goal of the agent in a scene that they observed.

The Current Study

In the current study, we tested the hypothesis that the meaning of a maximum absolute gradable adjective is, in fact, learnable from variable visual observations if learners reason about a new word meaning in context. As an illustration, one can imagine a learner who observed 100 instances of *full* being used where

only a small fraction of instances accompanied observations of the maximum degree of fullness. In most cases, an object that is not maximally full (e.g., a 90% full glass) is referred to as full or not full, depending on context. One can assume that learners can identify a relevant scale (e.g., fullness) from visual observations.[1] The key questions for learners, therefore, are (a) whether it is a relative or absolute gradable adjective; and, if absolute, (b) what constitutes the relevant end point of a scale. Given a large number of observations depicting not-maximally full objects, visual inspection of an object alone may not allow learners to conclude that a meaning is absolute. Rather, it might be more reasonable to assume that *full* has a relative meaning which requires a contextual comparison in the same way as a given tree can be tall or not tall, depending on a contextually relevant comparison class.

The same set of observations, however, could support learning an absolute meaning if one makes the following assumptions. First, one can assume that word learning leverages learners' knowledge of context, including an intended goal of an agent's action (Tomasello, 1995). For instance, by observing an agent serving water, learners would recognize a part of the agent's goal is to pour a glass without spilling. Second, one can assume that learners expect speakers to be rational (in the Gricean sense; Grice, 1975) and their word use to be informative in context. This supports the inference that, when a speaker describes the same object as sometimes full and other times as not full, it is unlikely due to random noise; there must be a reason for the speaker to do so. There is now a rich literature that has targeted how word learners leverage this basic rationality assumption to reason about word meaning (Barner, Brooks, & Bale, 2011; Clark, 2020; Hochstein, Bale, Fox, & Barner, 2014; Horowtiz & Frank, 2012; Nordmeyer & Frank, 2014; Papafragou, Fairchild, Cohen, & Friedberg, 2017; Ramscar, Dye, & Klein, 2013; Scofield & Behrend, 2008; Smith, Goodman, & Frank, 2013).

By combining these two assumptions, learners can explain why the same physical object may be described as full or not full across different scenarios. For instance, when a glass is described as full in the context of serving water, it is interpreted against the expectation that it could not be 100% full. In contrast, the same glass might be described as not full in the context of precisely measuring its content, where the glass would have to be 100% full. Assuming that the speaker is conveying a coherent meaning across these scenarios, over time, learners can infer that *full* concerns the maximum level of fullness (i.e., the absolute standard of comparison) for an agent's goal and purpose at hand. We called this *intention-based* reasoning and tested whether participating listeners would, indeed, be more likely to assign an absolute meaning rather than a

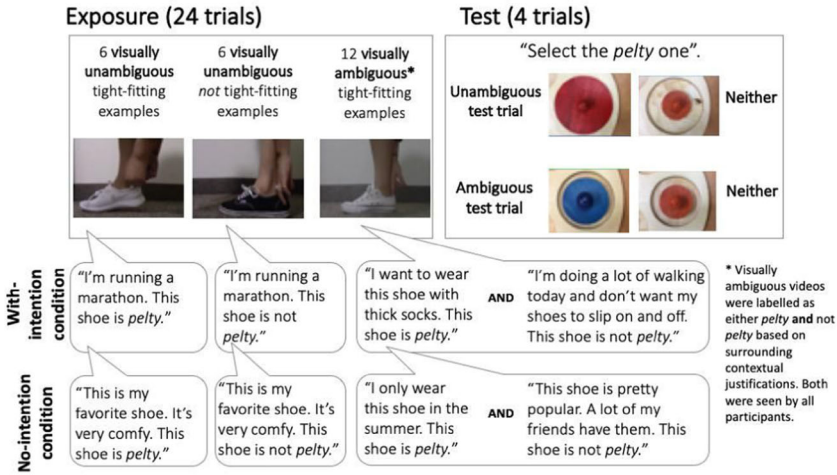

Figure 2 The experimental procedure, including the exposure and test phase. Participants were shown varying visual observations during exposure (e.g., illustrated here, different shoes). During testing, participants were asked to select the *pelty* one from three different choices. [Color figure can be viewed at wileyonlinelibrary.com]

relative meaning to a novel adjective when they had access to the agent's goals and intentions.

We constructed three word-learning experiments in which adult speakers of English heard a novel adjective *pelty* while observing scenes that depicted varying degrees of tightness of fit between two objects. We designed the visual stimuli (still images and videos) to draw participants' attention to this gradable property (e.g., a snugly fitting shoe; see Figure 2). Seminal work on language and cognition has suggested that the semantic distinction of tight fit versus a loose fit is lexically encoded in some languages (e.g., Korean), accessible even to young children who speak these languages, but this distinction is not salient in English (Choi, McDonough, Bowerman, & Mandler, 1999; Norbury, Waxman, & Song, 2008). We expected that the English-speaking participants in our study would need to learn this distinction as part of a novel word meaning.

We constructed the visual stimuli provided in the exposure phase so that they could support either the relative (i.e., more tight-fitting compared to a contextually determined standard) or the absolute (i.e., maximally tight-fitting) meaning to examine the effects of intention-based reasoning. We later tested participants on their interpretation of the novel adjective by extending the presupposition accommodation task in Syrett et al. (2010), as described above.

This phase enabled testing of whether participants would respond to a request, for example, "Select the *pelty* one," in a way that was consistent with the relative or the absolute meaning. We found that knowing the goals and intentions of the agent made it more likely for participants to learn the absolute meaning from highly (visually) variable scenes. The effects of intention-based reasoning, however, were significantly attenuated when observed scenes never visually represented the absolute standard.

Experiment 1

We taught adult English L1 speakers a novel gradable adjective, *pelty*, by exposing them to objects that varied along the dimension of tightness of fit. Half of the exposure stimuli depicted objects that were either unambiguously tight-fitting or unambiguously loose-fitting, thus providing evidence for a potential maximum absolute gradable meaning. The other half of the stimuli were ambiguous, that is, mostly, but not completely, tight-fitting, reflecting visual variability in word usage that could potentially lead to nonmaximal meaning inferences. We provided no explicit instructions about the meaning, as would be the case in naturalistic (as opposed to formal) language-learning scenarios.

Method

Participants

We recruited a total of 79 participants from the Internet-based crowd-sourcing platform Amazon Mechanical Turk (https://www.mturk.com/mturk/). These participants self-identified as monolingual English-speaking adults. To ensure participants' attention and English proficiency, we included two catch trials at the beginning of the experiment and at the beginning of the test phase. In these catch trials, we asked the participants to listen to instructions spoken at a relatively fast rate and to type a keyword (i.e., *pineapple*, *umbrella*) into a text-box. We excluded two participants due to poor catch trial performance and 14 participants due to past participation in a similar experiment.[2] We randomly assigned the remaining 63 participants either to the with-intention or the no-intention condition (30 in the with-intention and 33 in the no-intention conditions). The mean duration of the experiment was approximately 10 minutes, and each participant received $1. Visual and audio materials used in all the experiments reported in this article can be found at https://osf.io/b9mah/. The University of Rochester Institutional Review Board approved all procedures. Before starting the study, all participants gave written consent.

Table 1 Stimuli presentation across the three experiments

	Visually unambiguous		Visually unambiguous		
Experiment	Labeled as *pelty*	Labeled as *not pelty*	Labeled as *pelty*	Labeled as *not pelty*	Total
Experiment 1	6	6	6	6	24
Experiment 2	3	3	9	9	24
Experiment 3	0	0	12	12	24

Stimuli

We selected six different objects (T-shirt, bracelet, bookshelf, shoe, laptop case, and card) to create 18 videos that illustrated the uses of *pelty*. Each object was repeated three times for an unambiguously tight-fitting scene, an unambiguously loose-fitting scene, and an ambiguously tight-fitting scene (see Figure 2). We normed all of the videos with an independent group of 31 participants to ensure that each scene type could be reliably distinguished with respect to the degree of tightness of fit (for the results of this norming study, see Appendix S2 in the online Supporting Information).

To manipulate the effect of intention-based reasoning, we constructed two sets of 24 video stimuli by pairing the scenes with two types of verbal descriptions. In both sets, we paired the six unambiguously tight-fitting scenes and the six ambiguously loose-fitting scenes with descriptions that labeled them as *pelty* or not *pelty*, respectively. We constructed the 12 ambiguously tight-fitting scenes from six videos, each repeated twice while paired with a different verbal description (see Table 1).

All of the verbal descriptions of the scenes in the two between-participant (with-intention vs. no-intention) conditions are provided in Appendix S3 in the online Supporting Information. For the with-intention condition, the verbal descriptions provided information that clued the participant into the intentions and goals of the agent. For instance, a visually ambiguous scene with a shoe was introduced with the description "I want to wear this shoe with a thick sock. This shoe is *pelty*." In this case, the situational constraint relevant to the agent's intended goal (i.e., wearing shoes with thick socks) can explain why the shoe is not maximally tight-fitting when worn on a bare foot. For 50% of the visually ambiguous scenes, verbal descriptions labeled them as *not pelty* with relevant intention information (e.g., "I'm doing a lot of walking today, and don't want my shoes to slip on and off. This shoe is *not pelty*."). In contrast, in the no-intention condition, verbal descriptions were of similar details and

content while not including any reference to the agent's goal (e.g., "I only wear this shoe in summer because I don't want it to get muddy from spring rain. This shoe is *pelty*.").

Our test phase stimuli included two types of novel objects (cylinders and rings, see Appendix S3 in the online Supporting Information) that were devoid of any apparent functional goal or purpose. With each item type, we created two sets of test trials to be used in a three-alternative forced-choice task. In the unambiguous trials, these three choices were an unambiguously tight-fitting object (the tighter option), an unambiguously loose-fitting object (the not tighter option), and *Neither* printed next to the illustrations of the two objects (see top row of Figure 2). In the ambiguous trial, the choices were a mostly (ambiguously) tight-fitting object (i.e., the tighter option), a largely loose-fitting object (i.e., more room around the edges, the not tighter option), and *Neither* (see bottom row of Figure 2).

Procedure

We instructed the participants that a native speaker of a regional variant of English (Singaporean English) would be teaching them a new word, *pelty*, which is not used in American English. During the exposure phase, participants watched 24 videos as described above in a randomized order and answered the question "Is this *pelty*?" at the end of each video by selecting a radio button (labeled as *Yes* or *No*). This was to ensure that the participants listened to all the verbal descriptions presented via audio. In the test phase, the participants went through four test trials in which they saw two still images and the printed word *Neither* and received the prompt to "Select the *pelty* one" (see Figure 2).

Analysis Procedure

We planned to analyze the data in the following two steps. First, following the logic of Syrett et al. (2010), we predicted that those participants who inferred *pelty* to be an absolute gradable adjective would choose (a) the tighter option in an unambiguous test trial and (b) *Neither* in an ambiguous test trial. In contrast, those participants who inferred *pelty* to be a relative gradable adjective would select the tighter option in either type of test trial. Responses other than these two would signal that the participants had not learned *pelty* as a gradable adjective. We expected that the participants in the with-intention condition should be more likely than those in the no-intention condition to arrive at the absolute meaning. Responses in the no-intention condition would allow us to assess the extent to which the participants had been able to learn the absolute meaning based on the visual representations observed in the exposure phase.

To test this prediction, we used the glmer function of the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2016). We employed a mixed-effects logit regression (Breslow & Clayton, 1993; Jaeger, 2008) with a full two by two factorial design to analyze the participants' responses. We created two separate models to analyze the participants' choices of (a) the tighter option versus the other options and (b) the neither option versus the other options. We ANOVA (i.e., sum) coded all of the fixed effect predictors (ambiguous test trials $= 1$, unambiguous test trials $= -1$; with-intention condition $= 1$, no-intention condition $= -1$). We implemented a maximal random effects structure, including test object types (cylinder vs. rings) and participants. However, we removed higher-order terms one-by-one in the event of model convergence failure (Barr, Levy, Scheepers, & Tily, 2013). We set our alpha level at .05. We calculated odds ratios as a measure of the size of the effect.

Second, to gain additional insight, we computed joint distributions of responses between the ambiguous and unambiguous trials within each participant. Because there were three possible responses in each trial type, there were nine possible response patterns that a given participant could provide. We classified the participants into three categories:

1. those who chose the tighter option in the unambiguous trial and *Neither* in the ambiguous trial, thus exhibiting an absolute gradable adjective interpretation;
2. those who chose the tighter option in both trial types, thus exhibiting a relative gradable adjective interpretation;
3. those who showed other response patterns, indicating that they did not interpret *pelty* to be a gradable adjective (of any type).

We expected that, if intention-based reasoning plays a role in learning, the participants in the with-intention condition should be more likely than those in the no-intention condition to fall under Category 1.

Results

Proportions of the Tighter Option and Neither Responses

Figure 3 summarizes the response patterns in Experiment 1. (All the data and analysis codes can be found at https://osf.io/b9mah/. See Appendix S1 in the online Supporting Information for full model specifications and results.) First, we constructed a binomial mixed-effects logistic regression model predicting the choice of the tighter option over the other two options. Our final model included by-participant and by-item random intercepts. We found a

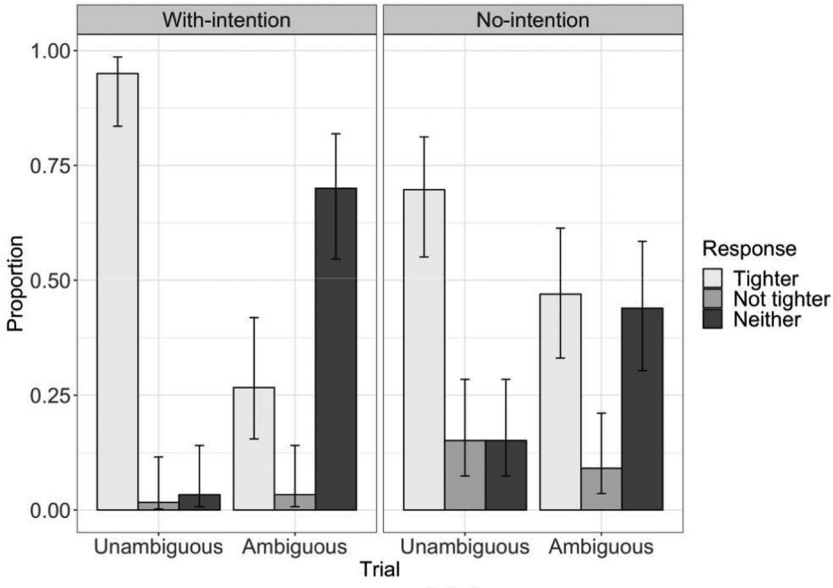

Figure 3 Proportion of responses collapsed across participants and test trials in Experiment 1. Error bars represent the 95% confidence interval on the proportion based on Goodman (1965).

significant main effect of test trial type, $b = -1.23$, 95% CI [−1.65, −1.17], $z = -6.51$, $p < .001$, $OR = 0.29$, 95% CI [0.19, 0.41], such that the likelihood of choosing the tighter option was greater in the unambiguous trial than in an ambiguous trial (as demonstrated by an OR of less than 1). This suggested that participants, overall, regarded the maximally tight-fitting object, rather than the mostly tight-fitting object, as a better representation of the meaning of *pelty*. Participants' choices of the tighter option did not differ significantly across the exposure conditions ($p = .11$). Crucially, there was a significant interaction of condition with test trial type such that participants were more likely to choose the tighter option in the unambiguous test trial, compared to the ambiguous test trial, if they had had access to intention information in the exposure phase (95% in the with-intention condition vs. 69.7% in the no-intention condition), $b = -0.75$, 95% CI [−1.17, −0.41], $z = -3.99$, $p < .001$, $OR = 0.47$, 95% CI [0.31, 0.67].

Second, we constructed a similar binomial model predicting the participants' choice of the neither option over the other two options. We found a significant main effect of test trial type, $b = 1.66$, 95% CI [1.19, 2.24],

$z = 6.30$, $p < .001$, $OR = 5.24$, 95% CI [3.29, 9.40], such that participants were about five times more likely to choose the neither option in an ambiguous trial than in an unambiguous trial. Participants' choices of the neither option did not differ significantly across the exposure conditions ($p = .75$). Most importantly, the likelihood of choosing the neither option was about twice as likely in the ambiguous trial, compared to the unambiguous trial, if they had access to intention information (70% in the with-intention condition vs. 43.9% in the no-intention condition), $b = 0.76$, 95% CI [0.34, 1.29], $z = 3.23$, $p = .001$, $OR = 2.14$, 95% CI [1.40, 3.63]. Together these two models supported the prediction that the participants in the with-intention condition would be more likely to show the response patterns compatible with the absolute interpretation (i.e., *pelty* means maximally tight-fitting) compared to the participants in the no-intention condition.

Joint Distributions of Responses

We classified the participants according to their responses in the unambiguous and ambiguous trials by the two object test types (i.e., rings/cylinders; see Figure 4). Overall, the participants provided more consistent responses in the with-intention condition: 63% (19/30) of them showed identical responses across the two types of objects. In contrast, only 42% (14/33) of the participants did so in the no-intention condition.

In the with-intention condition, 70% of the participants chose the tighter option in the unambiguous trials and the neither option in the ambiguous trials (i.e., the top-left corner of each three-by-three panel in Figure 4). This choice pattern corresponded to the absolute interpretation. Approximately 20% (20% in the cylinder trial, 23% in the ring trial) chose the tighter option in both ambiguous and unambiguous trial types (i.e., the bottom-left corner of the with-intention panels in Figure 4), a response pattern which corresponded to the relative interpretation. In contrast, the responses were more heterogeneous in the no-intention condition.[3] This corroborated what we had found in our binomial mixed-effect models; participants were more likely to assign an absolute gradable adjective interpretation in the with-intention condition than in the no-intention condition.

Discussion

The results strongly support the hypothesis that the availability of intention information facilitates the acquisition of an absolute meaning. The participants in the with-intention condition, compared to those in the no-intention condition, were more likely to provide responses that corresponded to the interpretation

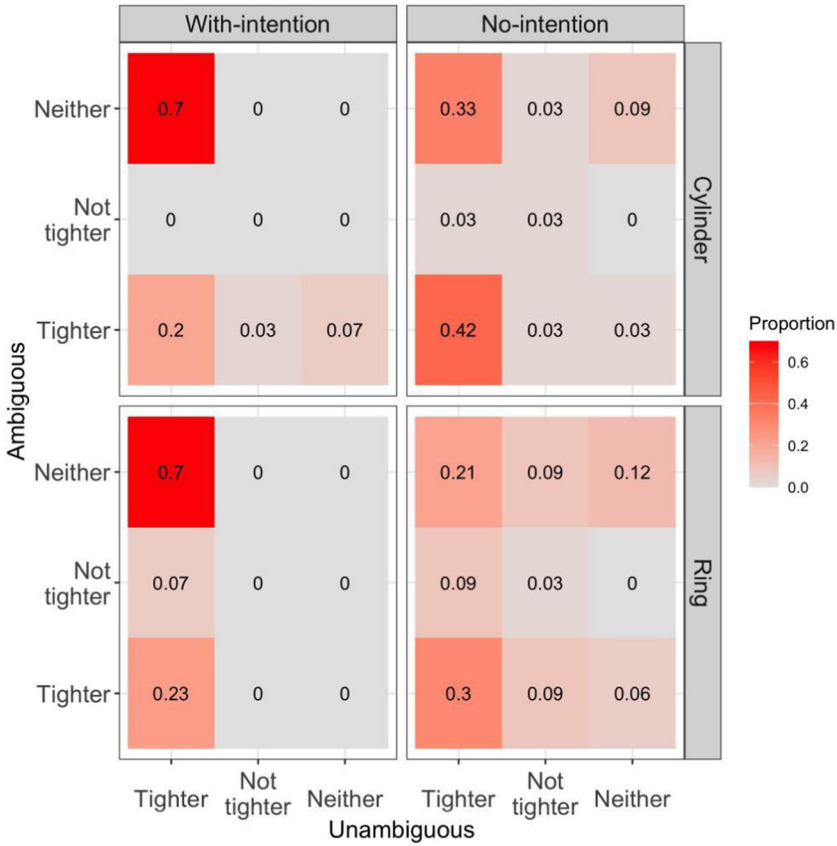

Figure 4 Proportions of participants falling into the nine possible response patterns across the unambiguous and ambiguous trial types (separated by the object types) in Experiment 1. The color gradation of the cells represents the percentages of the participants: Darker colors represent higher percentages. [Color figure can be viewed at wileyonlinelibrary.com]

of *pelty* as maximally tight-fitting. This is notable because only 25% (six out of 24) of the exposure stimuli visually represented the maximum degree of tightness of fit. Nonetheless, the participants in the with-intention condition, just like the adults in Syrett et al. (2010), reliably rejected the not-maximally tight-fitting objects and selected the neither option. Admittedly, varying degrees of tightness were made more topical in the descriptions of the with-intention condition compared to the no-intention condition. However, the added focus on the

dimension of tightness alone did not guarantee learning the absolute meaning because the exposure items included the same number of unambiguous scenes (maximally tight-fitting) and ambiguous scenes (mostly tight-fitting) labeled as *pelty* (see Table 1). To reject the tighter option in the ambiguous trials, the participants had to infer from the verbal descriptions that the maximum degree of tightness was intended even when it was not shown visually.

In Experiment 2, we put this hypothesis to a stronger test and examined whether participants could learn the absolute meaning with an even smaller number of absolute learning instances.

Experiment 2

Experiment 2 reused the paradigm of Experiment 1 with an increased number of visually ambiguous scenes in the exposure phase.

Method

Participants

We recruited a total of 83 participants from Amazon Mechanical Turk. We used the same exclusion criteria as in Experiment 1, and then analyzed the remaining data from 63 participants. Of these, 30 participants were in the with-intention condition and 33 were in the no-intention condition. The mean duration of the experiment was approximately 10 minutes, and each participant received $1.

Stimuli

We used stimuli identical to those that we had used in Experiment 1, except that we halved the number of visually unambiguous items (i.e., 12 in Experiment 1, six in Experiment 2) and increased the number of visually ambiguous items by 50% (i.e., 12 in Experiment 1, 18 in Experiment 2; see Table 1). We created two lists to counterbalance the object types presented as ambiguous and unambiguous scenes. For instance, some participants saw visually unambiguous scenes that were labeled appropriate usages of *pelty* involving the items book, shoe, and card while others saw the items laptop, shirt, and bracelet in unambiguous scenes; these were then counterbalanced in the other list. We created six additional ambiguous video stimuli to complete the lists.

Procedure

The procedure was identical to that of Experiment 1.

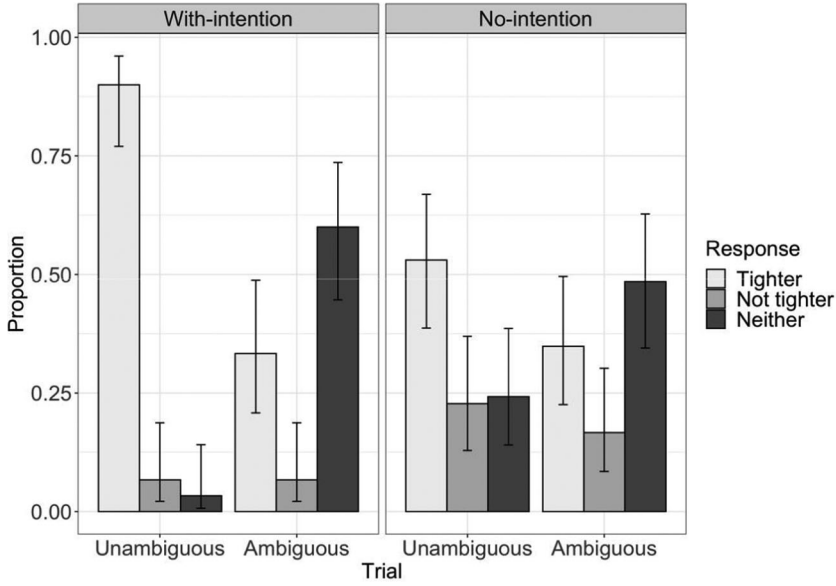

Figure 5 Proportion of responses collapsed across participants and test trials in Experiment 2. Error bars represent the 95% confidence interval on the proportion based on Goodman (1965).

Analysis Procedure

The analysis plan was identical to that of Experiment 1. To directly compare the results from Experiments 1 and 2, we also conducted analyses using a combined data set.

Results

Proportions of the Tighter Option and Neither Responses

Figure 5 summarizes the response patterns in Experiment 2. As in Experiment 1, we first constructed a binomial mixed-effects logistic regression model predicting participants' responses of tighter over the other two. We found a main effect of test trial type, $b = -1.26$, 95% CI [-1.75, -0.85], $z = -5.60$, $p < .001$, $OR = 0.29$, 95% CI [0.18, 0.43], where the participants were about a third as likely to choose the tighter option in the ambiguous test trials compared to the unambiguous test trials. Condition was also was significant, $b = 0.68$, 95% CI [0.17, 1.27], $z = 2.53$, $p = .01$, $OR = 1.97$, 95% CI [1.19, 3.55], suggesting that the participants were about twice as likely to choose the tighter option when they had had access to intention information in the

exposure phase. Notably, we found an interaction of condition with test trial type, where the likelihood of participants choosing the tighter option was about half as likely in the ambiguous trials compared to the unambiguous trials, when they had had access to the intention information in the exposure phase (90% in the with-intention condition vs. 53% in the no-intention condition), $b = -0.73$, 95% CI [−1.17, −0.36], $z = 3.61$, $p < .001$, $OR = 0.48$, 95% CI [0.31, 0.70].

Second, we constructed a similar binomial model that predicted participants' choice of the neither option over the other two options. We found a main effect of test trial type, $b = 1.42$, 95% CI [0.98, 1.97], $z = 5.73$, $p < .001$, $OR = 4.12$, 95% CI [2.67, 7.19], similar to what we had found in Experiment 1. Participants were about four times as likely to choose the neither option in the ambiguous trial than the unambiguous trial. Further, there was an interaction of condition with test trial type when participants chose the neither option (60% in the with-intention condition in the ambiguous trial vs. 49% in the no-intention condition), $b = 0.75$, 95% CI [0.34, 1.27], $z = 3.28$, $p = .001$, $OR = 2.12$, 95% CI [1.41, 3.57]. Together, these two models supported the prediction that the participants in the with-intention condition were more likely to assign an absolute meaning to the novel adjective, even from predominantly ambiguous observations.

Comparisons Between Experiments 1 and 2

To directly compare the results of Experiments 1 and 2, we constructed two additional binomial mixed-effect logistic regression models for a combined data set. The first model predicted tighter option responses using data from both experiments: exposure condition (with- vs. no-intention), test trial type (unambiguous vs. ambiguous trials), experiment (Experiment 1 = 1; Experiment 2 = −1), and we ANOVA coded their interactions and entered them as fixed effects. We entered test object types and participants as random intercepts. We again found main effects of condition, $b = 0.44$, 95% CI [0.15, 0.75], $z = 2.93$, $p = .003$, $OR = 1.55$, 95% CI [1.16, 2.11], and test trial type, $b = -1.21$, 95% CI [−1.51, −0.95], $z = -8.46$, $p < .001$, $OR = 0.30$ [0.22, 0.39]. Importantly, the data supported the critical, predicted interaction of condition with test trial type, $b = -0.71$, 95% CI [−0.99, −0.47], $z = -5.36$, $p < .001$, $OR = 0.49$, 95% CI [0.37, 0.63]. Participants in the with-intention condition were more likely than those in the no-intention condition to choose the tighter option in the unambiguous test trial. The main effect of experiment ($p = 0.13$), and its interactions with other fixed effects ($p = .41$), were not significant. This indicated that participants' choice of the tighter option was largely intact, even when the number of unambiguous scenes was halved in Experiment 2.

Next, we constructed a similar model to analyze participants' choice of the neither option. As we expected, the odds of participants choosing the neither option was more than four times as likely in the ambiguous trials compared to the unambiguous trials, $b = 1.54$, 95% CI [1.21, 1.92], $z = 8.52$, $p < .001$, $OR = 4.65$, 95% CI [3.35, 6.84]. Critically, we found a significant interaction of exposure condition with test trial type, $b = 0.76$, 95% CI [0.45, 1.11], $z = 4.60$, $p < .001$, $OR = 2.13$, 95% CI [1.58, 3.03]. The three-way interaction including experiments was not a significant predictor for neither option responses ($p = .98$).

Joint Distributions of Responses

As in Experiment 1, participants' responses were more consistent across the object types (i.e., rings/cylinders) in the with-intention condition than in the no-intention condition (see Figure 6); 80% (24/30) of the participants in the with-intention condition showed consistent response patterns across the two object types. In contrast, only 45% (15/33) of the participants in the no-intention condition provided consistent responses across the object types. This suggested that the increased number of visually ambiguous scenes affected the responses of those who were in the no-intention condition more strongly, making it less likely for them to arrive at a coherent and reliable interpretation of the novel adjective.

Overall, the joint distributions of response patterns between the ambiguous and the unambiguous trials were strikingly similar to those in Experiment 1. Of the participants in the with-intention condition, 60% (both in the ring and the cylinder trials) provided a response consistent with an absolute gradable adjective interpretation, but only 39% (24% in the ring trial, 39% in the cylinder trial) of those in the no-intention condition did so. This yielded additional support for the results of the mixed-effect analyses. Despite the reduced exposure to unambiguous scenes, the participants in the with-intention condition were more likely to consider *pelty* as an absolute gradable adjective compared to the participants in the no-intention condition.

Discussion

The participants in the with-intention condition were able to learn an absolute meaning at approximately the same rate (i.e., over the same number of stimuli seen) as in Experiment 1. This is remarkable given that the tightness of fit in the vast majority of the scenes in the exposure (18/24 tokens, 75%) was visually ambiguous. Only three out of 24 scenes depicted the maximum degree of tightness of fit. Based on the same number and types of learning observations,

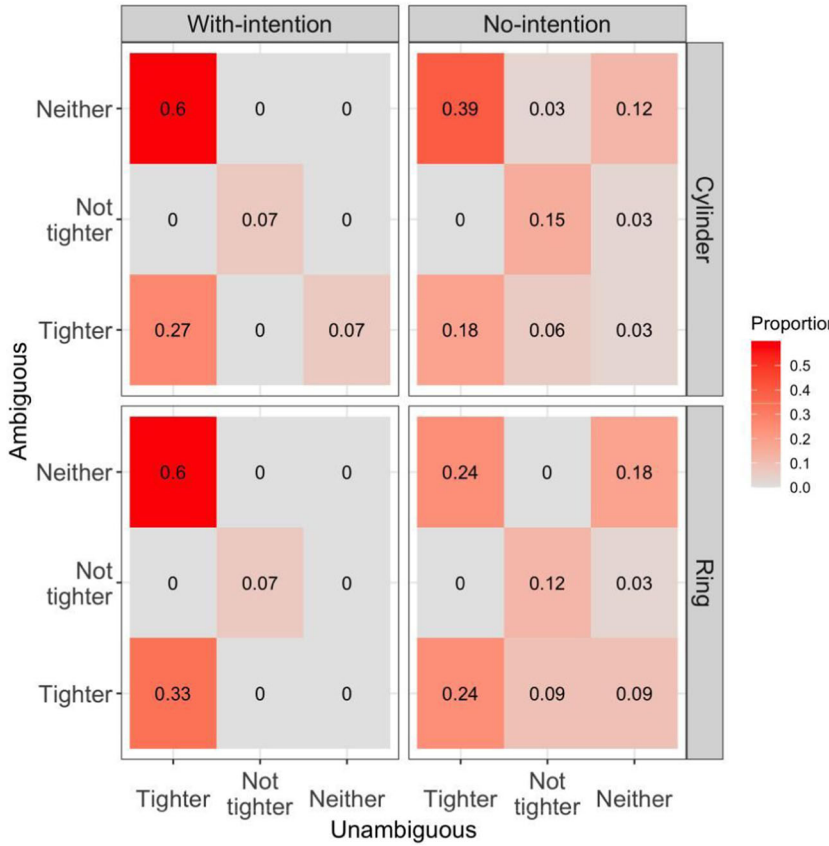

Figure 6 Proportions of participants who fell into the nine possible response patterns across the unambiguous and ambiguous trial types (separated by the object types) in Experiment 2. The color gradation of the cells represents the percentages of the participants: Darker colors represent higher percentages. [Color figure can be viewed at wileyonlinelibrary.com]

the participants in the no-intention condition arrived at a much wider variety of meanings for the novel gradable adjective.

This led us to ask about the limits of intention-based reasoning. It is possible that the reasoning process is sufficiently powerful to support learners' acquisition of an absolute meaning with no direct observations of scenes that represent a maximum standard of comparison. Alternatively, a minimal amount of exposure to unambiguous learning observations could be necessary. We

conducted Experiment 3 to test whether participants might be able to infer the absolute meaning of *pelty* exclusively from visually ambiguous observations.

Experiment 3

We replicated Experiment 2 but replaced all the visually unambiguous (i.e., maximally tight-fitting) exposure scenes with visually ambiguous ones.

Method

Participants

We recruited a total of 99 participants to participate from Amazon Mechanical Turk. We used the same exclusion criteria used in Experiments 1 and 2. Due to poor catch trial performance ($n = 1$) and past participation in a similar experiment ($n = 25$), we excluded 26 participants. We included the remaining 73 participants (34 in with-intention and 39 in no-intention conditions) in our analyses. The mean duration of the experiment was approximately 11 minutes, and each participant received $1.

Stimuli

We used stimuli identical to those used in Experiment 1 except that the participants saw only ambiguous scenes (24 items). We labeled 12 of the items as *pelty* and the other 12 as *not pelty* (see Table 1).

Procedure

The procedure and analysis were identical to those of Experiments 1 and 2.

Analysis Procedure

The analysis plan was identical to that of Experiments 1 and 2.

Results

Proportions of the Tighter Option and Neither Responses

Figure 7 summarizes the response patterns in Experiment 3. As in Experiments 1 and 2, we first constructed a binomial mixed-effects logistic regression model predicting the participants' responses for the tighter option over the other two options. We found a main effect of test trial type, $b = -0.34$, 95% CI [-0.61, -0.08], $z = -2.50$, $p = .01$, $OR = 0.71$, 95% CI [0.54, 0.93], but the exposure condition (with- vs. no-intention) was not significant ($p = .55$). That is, the likelihood of choosing the tighter option was smaller for ambiguous test trials compared to the unambiguous ones, and this likelihood was similar between the two conditions (with- vs. no-intention). Unlike in Experiments 1 and

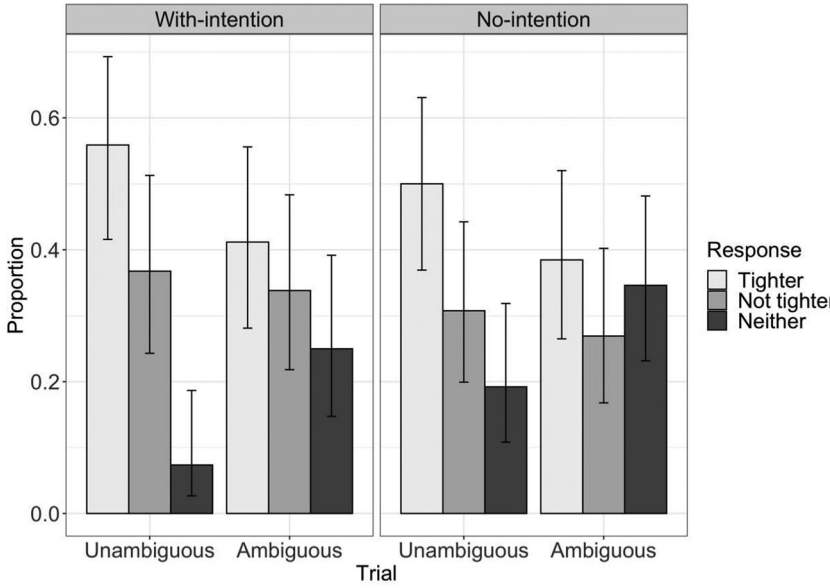

Figure 7 Proportion of responses collapsed across participants and test trials in Experiment 3. Error bars represent the 95% confidence interval on the proportion based on Goodman (1965).

2, there was no interaction of the test trial type with the exposure condition ($p = .76$). This suggested that in Experiment 3 the availability of the intention information did not significantly alter the participants' responses across the unambiguous and ambiguous trials.

Second, we constructed a similar binomial model predicting participants' choice of the neither option over the other two options. We found significant main effects for the test trial type and the exposure condition: Participants demonstrated greater odds of choosing the neither option when (a) neither of the choice options represented the maximum degree of tightness (i.e., in ambiguous trials), $b = 0.64$, 95% CI [0.30, 1.01], $z = 3.56$, $p < .001$, $OR = 1.89$, 95% CI [1.35, 2.74], and (b) in the no-intention condition, $b = -0.44$, 95% CI [−0.89, −0.02], $z = -2.03$, $p = .04$, $OR = 0.65$, 95% CI [0.41, 0.98]. These two effects did not interact with each other ($p = .36$), suggesting that the participants in the two between-subject exposure conditions were equally likely to choose the neither option in the ambiguous trial. The overall higher number of neither responses in the no-intention condition relative to the with-intention condition was unexpected and not readily explainable. We speculated that the

level of uncertainty about the adjective meaning was particularly high in the no-intention condition, which made it more likely for the participants to assign a *nongradable* adjective interpretation to *pelty*. We return to this question in the General Discussion section.

Joint Distributions of Responses

Compared to Experiments 1 and 2, the participants' responses were overall more variable in Experiment 3 (see Figure 8); 44% of the participants in the with-intention condition and 33% of participants in the no-intention condition provided consistent responses across the object types. It is interesting to note that, in Experiment 3, a large proportion of the participants chose the not tighter option in an unambiguous trial at least once (53% in the with-intention and 56% in no-intention conditions, respectively) despite the presence of the maximally tight-fitting option in unambiguous trials. This might have indicated that they had inferred room between two objects to be permitted and perhaps expected (e.g., *pelty* means loose-fitting). If so, their interpretation likely deviated from one that is canonically associated with a maximum absolute gradable adjective. The results thus suggested that the participants did not readily learn the absolute meaning when their exposure consisted exclusively of ambiguous scenes.

Discussion

Based solely on visually ambiguous observations, the participants were significantly less likely to exhibit response patterns associated with an absolute interpretation, even with access to information about an agent's intention. The result (combined with those of Experiments 1 and 2) suggests that the intention-based reasoning facilitates but does not replace, a more commonly assumed cross-situational word-learning process (Smith & Yu, 2008; Vouloumanos, 2008; Vouloumanos & Werker, 2009; Yu & Ballard, 2007). Language learners are known to build their hypotheses about a novel word meaning based on accumulating experiences of mappings between linguistic forms and visual observations (Akhtar & Montague, 1999; Fisher, Hall, Rakowitz, & Gleitman, 1994; Gillette, Gleitman, Gleitman, & Lederer, 1999; Gleitman, 1990; Pinker, 1984; Trueswell, Medina, Hafri, & Gleitman, 2013). In this process, a new observation provides evidence that can either upweight or downweight current hypotheses about possible word meanings. Without direct observations of visual scenes that clearly represent a maximum standard of comparison, multiple candidate meanings will remain feasible. Those meanings can include gradable interpretations that we have considered here (such as relative and absolute with

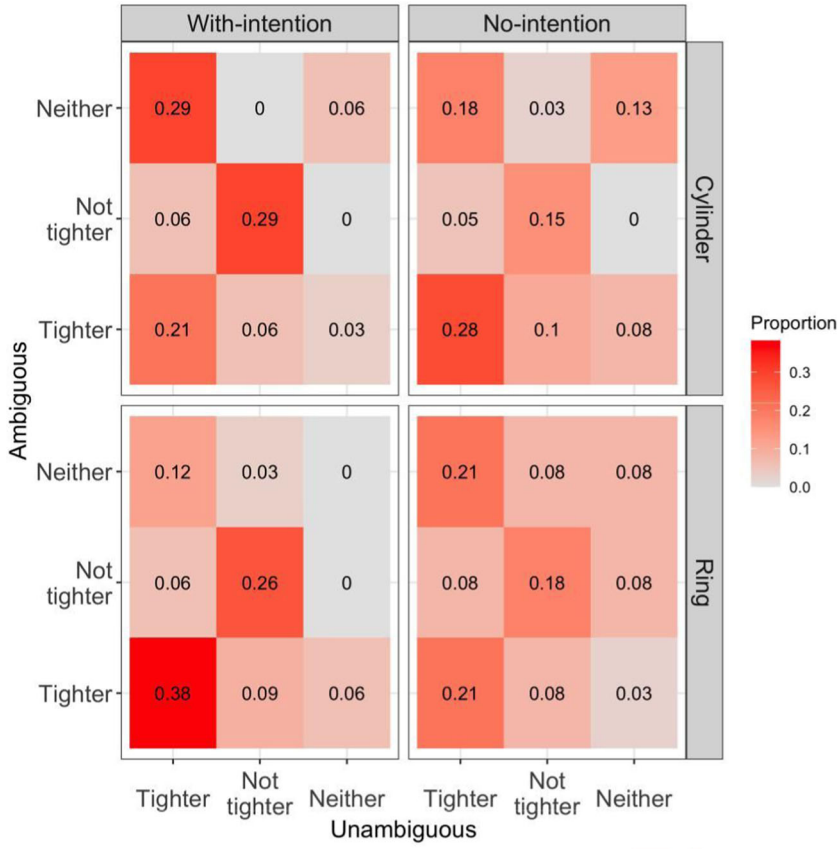

Figure 8 Proportions of participants who fell into the nine possible response patterns across the unambiguous and ambiguous trial types (separated by the object types) in Experiment 3. The color gradation of the cells represents the percentages of the participants: Darker colors represent higher percentages. [Color figure can be viewed at wileyonlinelibrary.com]

higher and lower precision thresholds) as well as more ad hoc ones that learners have inferred in the context. In such cases, contextually provided information about the agent's intended goals was not sufficient to support acquisition of a maximum absolute gradable adjective. We conclude that intention-based reasoning comes into effect when the absolute meaning is sufficiently salient by virtue of unambiguous visual observations.

General Discussion

We set out to investigate the effects of linguistic relativity on L2 learning. When a target language guides attention to a gradient scale that is not linguistically encoded in L1, how do learners determine whether the meaning is absolute or relative? We focused on a case of maximum absolute gradable adjectives because ubiquitous imprecision in their use patterns presents a significant puzzle to word learners. To learn the scalar structure that involves a maximum or minimum standard of comparison, learners must go beyond what is visually represented in an observed scene or object. Moreover, the physical properties of an object do not reliably predict the use of an absolute gradable adjective (e.g., a glass can be full or not full, depending on its purpose of use; a line with a slight unevenness is considered straight when it is hand-drawn but not straight when it is drawn with graphic software). Recognizing why a given context is afforded the use of an absolute gradable adjective is, therefore, a critical step in the process of learning an adjective's meaning.

Experiments 1 and 2 demonstrated that the adult participants who had access to information about the agent's intended goals, compared to those who did not, were more likely to associate a novel word form with an absolute meaning. The contrasting results from Experiments 2 and 3, however, demonstrated that the effects of intention-based reasoning critically interact with direct observations of unambiguous scenes. In the absence of maximally tight-fitting objects in exposure, the participants considered a wider range of possible meanings even in the with-intention condition. This supports the idea that intention-based reasoning works in tandem with a more commonly assumed hypothesis-testing process where learners accumulate positive and negative evidence for hypothesized meanings. To further examine the role of positive and negative evidence, we re-ran Experiment 1 with only the trials in which a referent was labeled as *pelty* (i.e., positive evidence only). In this Experiment 4, the participants in the with-intention condition did not choose the neither option over the tighter option as reliably as in Experiment 1 (Experiment 1: tighter = 26.7%, neither = 70% vs. Experiment 4: tighter = 55.8%, neither = 44.2%). This suggests that the absolute meaning was made accessible in part by the evidence determining loose-fitting objects as *not pelty*. (More details of this follow-up experiment are provided in Appendix S5 in the online Supporting Information.)

Inferred intention of an agent was thus factored into a possible word meaning mainly when the participants already had sufficient evidence for an endpoint oriented interpretation for *pelty*. The maximum standard (i.e., maximum tightness of fit) seems to arise as learners evaluate possible meanings for a

novel gradable adjective not only in a given, current context, but also across multiple contexts. This is broadly consistent with the proposal that imprecision associated with maximum standard absolute gradable adjective is pragmatic in nature (Aparicio, Xiang, & Kennedy, 2015; Kennedy, 2007; Leffel, Xiang, & Kennedy, 2016; Qing & Franke, 2014; Syrett et al., 2010). Unlike other types of adjectives that are semantically vague (e.g., A long rope can be arbitrarily long), maximum absolute gradable adjectives like *full* or *straight* are said to have a definitive, unique value specified in a context which is conveyed loosely according to pragmatic purposes at hand. A core insight from the current results is that a clear grasp of these pragmatic purposes can, over time, help learners arrive at a definitive, unique value as part of the semantic meaning of a maximum absolute gradable adjective.

Inferring Precision Threshold

One crucial assumption that we made throughout Experiments 1, 2, and 3 was that the choice patterns exhibited in the test phase corresponded to the participants' internal semantic representations for the novel adjective. In particular, based on Syrett et al. (2010), we associated the choice of the tighter option in both the ambiguous and the unambiguous trials with the relative meaning (i.e., tighter-fitting of the two). However, interpretations that gave rise to this response pattern may have been more heterogeneous. Most relevant to the current discussion, this pattern is, in principle, compatible with a maximum standard absolute interpretation with a low precision threshold (e.g., close to an end-point of the underlying scale). In other words, the choice of the tighter option across the ambiguous and the unambiguous trials can stem from the same pragmatic weakening that would be observed with a 90% full glass being construed as full enough. If so, perhaps a greater number of the participants than we had expected might be responding with a maximum standard interpretation in mind, not a relative one.

To gain additional information about the exact nature of adjective meaning learned in different conditions, we conducted post hoc replication experiments to elicit explicit responses about the inferred interpretation of *pelty* at the end of the experiments. We had these new groups of participants go through the exposure and the test phases of Experiments 1, 2, and 3, ($n = 94$, 97, 95, respectively), which we have labeled Experiments 1′, 2′, and 3′. We subsequently asked these participants a multiple-choice probe question: "What do you think pelty means?" They responded to the question by selecting one option from the following five alternatives:

1. Maximally tight-fitting with no space around an inner object (i.e., absolute interpretation with a high precision threshold);
2. More or less tight-fitting with some space around an inner object (i.e., absolute interpretation with a low precision threshold);
3. The tighter-fitting of the two (i.e., relative interpretation);
4. Loose-fitting;
5. Something else (free text answer).

It should be noted that this probe question, although being a direct approach to obtaining information aboutthe acquired meaning, might not have been an optimal means to assess the participants' intuitive understanding of the novel word. When asked to provide an explicit judgment on a word, learners tend to vary in their ways of verbalizing internal representations of word meaning and may revise or meta-reason about their interpretation as they answer the question. Responses collected in this replication experiment should therefore be interpreted with a grain of salt and be considered supplemental to the main results that we have reported.

With that being said, participants' answers to the probe question provided intriguing insight into the nature of the meanings inferred in the with- versus no-intention conditions across the three (replicated) experiments. A summary table of the answers can be found in Appendix S6 of the online Supporting Information. Notably, the results corroborated the main findings of Experiments 1, 2, and 3: The participants were more likely to derive an absolute interpretation (i.e., the sum of the Options 1 and 2 listed above) than a relative interpretation both in Experiment 1′ (with-intention = 79.1% vs. no-intention = 74.6%) and in Experiment 2′ (with-intention = 87.5% vs. no-intention = 55.1%). In Experiment 3′, a much smaller proportion of participants chose Option 1 or 2 (with-intention = 53% vs. no-intention = 48.2%). In fact, the responses were almost evenly distributed across the five choice options in Experiment 3′, which is compatible with the variable response patterns that we had observed in the test phase in the original Experiment 3 (see Figure 8). A comparatively large proportion of participants in Experiment 3′ selected Option 4 (i.e., loose-fitting: with-intention = 26.5% vs. no-intention = 16.7%), which suggested that the participants who had observed nothing but ambiguous scenes often recognized the non-maximum fitness represented in the scenes to be part of the novel word meaning.

The responses also highlighted possible differences between Experiments 1 and 2 for a precision threshold applied in the test phase. The selection of Option 1 (i.e., an absolute interpretation with a high precision threshold) was

most common in the with-intention condition in Experiment 1′ (44.2%) compared to any other condition across all the replication experiments. In the corresponding condition in Experiment 2′, only 25% of the participants chose Option 1 whereas a much larger proportion of participants (62.5%) selected Option 2 (i.e., an absolute interpretation with a low precision threshold). This may suggest that participants in Experiment 1′might have derived a stronger expectation for an end-point oriented interpretation of the new absolute gradable adjective (i.e., *pelty* means maximally tight-fitting). That is, the reduced number of maximally tight-fitting objects observed during exposure in Experiment 2 might have prompted participants to adopt a lower precision threshold when providing judgments on the meaning of *pelty*. Albeit rudimentary, the converging evidence from our main and post hoc investigations points to an interactive relationship between intention-based reasoning and visual observations. Learners draw on multiple sources of information to infer a novel adjective meaning.

Using Contextual Understanding to Learn Semantic Meaning

The importance of leveraging speaker intentions has been addressed in the long tradition of word-learning research. Even before learners have mastered object names and the meanings of basic nouns, social cues such as eye-gaze and pointing can effectively signal to what a speaker means to refer (see Baldwin, 1993; Bloom, 2002; Estigarribia & Clark, 2007; Hollich, Hirsh-Pasek, & Golinkoff, 2000, for L1 acquisition; see Kuhl, Tsao, & Liu, 2003; Sueyoshi & Hardison, 2005, for L2 acquisition). More broadly, grasping an agent's intended goal of an action allows learners to bootstrap meanings beyond those of basic nouns (e.g., *chasing* vs. *fleeing*, or accidental vs. intentional actions) (Bloom, 1997; Brooks, Audet, & Barner, 2013; Carpenter, Akhtar, & Tomasello, 1998; Golinkoff & Hirsh-Pasek, 2008; Poulin-Dubois & Forbes, 2002). For a novice learner, understanding a word meaning, in large part, begins with understanding the intention that the speaker attempted to communicate by choosing the word form in context (Bloom, 2002; Clark, 1990; Frank, 2014; Frank, Goodman, & Tenenbaum, 2009).

The current study presented a proof-of-concept for extending this logic to the acquisition of maximum absolute gradable adjectives. We used scenarios in which a gradable property itself (e.g., tightness of fit) could be identified through visual observations. Ambiguity remains, however, because of imprecision. That is, speakers expect comprehenders to map the adjective form (e.g., *The glass is full*) to an intended meaning through the aid of contextual knowledge, including an agent's intended goal of action. In a standard, linguistic

analysis of adjective meanings, it is commonplace to assume that comprehenders must first call to mind the semantic meaning of an adjective. They then interpret (or "relax") it in consideration of relevant aspects of a given context (Lasersohn, 1999). For word *learners*, who, by definition, lack the semantic meaning, the process does not straightforwardly apply. What we demonstrated here was that word learners could instead leverage their understanding of context to inductively break into the semantics of absolute gradable adjectives. Such reasoning would allow an absolute standard of comparison to emerge as a hypothesis that correctly predicts a speaker's use of an adjective for observed objects and scenes once relevant contextual variables have been considered.[4]

One key process that likely plays a role in this inference is a well-known inductive bias called "explaining away" (Pearl, 1988). For example, in a scenario in which the word *pelty* is used for a more or less tight-fitting shoe, there would be at least two plausible explanations: (a) *Pelty* means relatively (but not maximally) tight-fitting and (b) The agent had a reason to keep the shoe not maximally tight-fitting. Logically, the observation could be compatible with either of these explanations (or both; i.e., *pelty* could mean relatively tight-fitting and the agent also might have intended to keep it not maximally tight-fitting). In other words, confirmation of Explanation (a), for example, does not make Explanation (b) objectively any more or less likely. Under typical probabilistic reasoning, however, confirmation of Explanation (a) leads to reduction of the probability of Explanation (b). As a result, learners come to conclude that the nonabsolute nature of the observation is due to the incidental (contextual) cause, for example, Explanation (a), rather than the word's semantic meaning, for example, Explanation (b). Similar reasoning across contexts and contextual causes would support the inference that the word meaning should involve an absolute (maximum) standard of comparison.

The fundamental assumption behind intention-based reasoning is, thus, that word learners a priori expect speakers to be imprecise in a way that makes their word choice contextually informative. We consider this to be a reasonable assumption for adult (L2) learners with pragmatic competence and ample experiences of word use. In fact, an increasing number of L2 word-learning studies have demonstrated learners' active inferencing about word meaning through linguistic and extra-linguistic context (e.g., de la Garza & Harris, 2017; Nagy, Anderson, & Herman, 1987; Nagy, Herman, & Anderson, 1985; Nassaji, 2006). Previous work has so far focused mainly on acquisition of nouns and other nominals (cf. recent research on event categorization by Vanek, 2019). The current results extend the idea to adjectives and add to the

body of the work by suggesting that the contextual information involves speakers' intentions as relevant to the goal of an action.

The extent to which child (L1) learners assume that speakers' language use is contextually informative is much less clear. Traditionally, children were thought to have weak expectations regarding informative language use, making their pragmatic language use non-adult-like (Davies & Katsos, 2010; Deutsch & Pechmann, 1982; Eskritt, Whalen, & Lee, 2008; Matthews, Lieven, & Tomasello, 2007; Morisseau, Davies, & Matthews, 2013; Sonnenschein, 1982). More recent studies, however, have found that young children have a sophisticated cognitive framework for reasoning about others' actions, including language use, according to their cost and effectiveness of information transmission (see Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016, for a review). For instance, 4-year-olds and adults would learn from "This is a special *glorp* [a nonce word]. This is a red *glorp*," that *glorps* are not usually red (Horowitz & Frank, 2016; Horowitz & Frank, 2012). This indicates that 4-year-olds, like adults, interpret the adjective *red* not merely as a descriptive modifier but according to the speaker's likely intention to highlight a meaningful contrast. With this type of reasoning, children can learn word meanings and concepts (e.g., What is a *glorp*?) beyond what they can directly observe in the input. We take this and related findings as suggestive evidence that children can use intention-based reasoning in learning gradable adjectives, perhaps with a substantially larger number of observations than what would be necessary for adults. It would be a fruitful avenue of research to examine children's as well as adults' assumptions about informative language use and other inductive biases in word-learning.

Limitations and Future Directions

The current study leaves unaddressed some important questions regarding the process of adjective learning in L1 and L2 acquisition. First, it has been shown that L1 vocabulary knowledge strongly impacts the ease and trajectory of L2 lexical learning (e.g., Athanasopoulos & Albright, 2016; Ellis, 1995; Jiang, 2002; Krashen, 1981; Ringbom, 1983). Although the participants of our study did not exactly know what *pelty* meant, they likely employed their knowledge about similar lexical items in their L1—relative and absolute gradable adjectives, in particular. For instance, the participants might have called to mind an English lexical item whose meaning could serve as a stand-in for the new semantic concept (e.g., *snug*). If so, the current findings are a result of semantic or conceptual transfer (Giacobbe, 1992; Ringbom, 1983), rather than acquisition of a brand-new semantic meaning. Future studies should investigate

the extent to which existing lexical knowledge provides necessary scaffolding for the intention-reasoning process involving adjectives. Importantly, such research should verify that this process works for acquiring a meaning or concept that was previously unavailable to the learner.

One methodological question raised by the current research was how to examine a scalar structure inferred for a novel adjective. The presupposition accommodation test provides a useful diagnosis for whether observed responses better correspond to an absolute or a relative interpretation. This, however, is arguably an indirect assessment, leaving open the question of whether the meaning was in fact evaluated against a particular standard of comparison. With real, existing adjectives, researchers have devised a number of diagnostics to answer this question, for example, Does an adjective have a comparative or superlative form? Can it be modified by proportional modifiers (e.g., *completely*, to indicate an absolute meaning) or intensifiers (e.g., *very*, to indicate a relative meaning)? (Syrett & Lidz, 2010; Kennedy, 2007; Kennedy & McNally, 1999, 2005; Paradis, 2001; Rotstein & Winter, 2004). More research is needed to determine whether such semantic judgments can be effectively extended to nonce adjectives. If applicable, these judgments will provide a window into how underlying scalar structure can be learned through accumulating exposure to the input.

Also of importance is to test intention-based reasoning with a wider variety of gradable adjectives. Syrett et al. (2010) reported on an intriguing asymmetry in children's and adults' responses to absolute gradable adjectives, with a maximum versus a minimum standard of comparison. Perhaps, learners begin to learn absolute gradable adjectives separately, as individual items, and later extrapolate a common feature (e.g., having a minimum or maximum standard of comparison) among them. It is important to test whether, and, if so, how, learners apply the intention-based reasoning for different absolute gradable adjectives. Such attempts will elucidate how the effects of intention-based learning might interact with other variables, such as semantic meaning of adjectives, frequencies of use, and the typical pragmatic functions associated with them. In addition, future studies could examine how intention-based reasoning can interact with other sources of linguistic information. For instance, adverbials, such as *maximally* or *completely*, used together with a gradable adjective can support syntactic bootstrapping of the adjective meaning (Syrett & Lidz, 2010). Future examinations must include multiple syntactic constructions and co-occurring adverbs to examine their effects.

Finally, the logic of intention-based reasoning can be extended beyond adjectives. For example, some motion and change of state verbs have an

event- and situation-dependent start/end point as part of their semantics (Comrie, 1976; Levin & Rappaport Hovav, 2005; Vendler, 1967). When one finishes dinner, it usually means that she has achieved a state where no food remains on her plate. But an exact visual representation of the plate can vary according to multiple variables (e.g., food items, amount of food served, formality of the dining event). Sometimes, a plate with a few pieces of vegetables left can be considered finished. In addition, *finishing dinner* can have a different set of situational constraints from *finishing work* or *finishing a race*. To extract an abstract semantic meaning that pertains to an end point of an event, learners must understand how the meaning is realized under various intentionally and situationally defined conditioning variables. Inferring an agent's goals and intentions is thus a critical step in constructing abstract semantic representations from the contextually situated evidence of word use.

Conclusion

We sought to resolve a puzzle: How can adult learners overcome challenges stemming from prevalent cases of imprecision in adjective use? Our three experiments and two follow-up experiments presented a novel gradable adjective with or without information that explains contextual sources of imprecision. Results demonstrated that contextual information (e.g., inferred goals of the agent), in particular, can help learners reason about an intended maximum standard of comparison from variable visual observations. We also found that effects of intention-based reasoning must be examined within the broader scope of hypothesis-testing processes that are critical in word learning, in which observed evidence is evaluated against a priori expectations about possible and likely meanings. Our findings thus highlight the importance of pragmatic reasoning in adjective learning, providing a new path forward to understanding word learning in natural conversational contexts.

Final revised version accepted 11 June 2020

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/b9mah. All proprietary materials have been precisely identified in the manuscript.

Notes

1 We verified this assumption in our experiments.
2 Here and in the subsequent experiments, we excluded participants who had previously participated in one of the related experiments (including pilot experiments and stimuli norming surveys). We took two measures. First, in a Mechanical Turk HIT page, we described the task with visual material used in the experiment and requested Turkers not to accept the current HIT if they recognized the study content. Second, after collecting data and compensating participants, we removed repeating participants based on their IP addresses.
3 Here and in the other experiments, we analyzed the joint distribution of the three-way outcomes with a Bayesian multilevel (mixed) multinomial regression using the brm function of R package brms (Bürkner, 2017). This is one of the few available implementations suitable for the analysis of repeated-measures of unordered multinomial outcomes (i.e., tighter, not tighter, and neither options) across conditions. The results of these analyses, supporting the observations made here, are provided in Appendix S7 in the online Supporting Information. For a comprehensive introduction to basic principles and approaches of Bayesian analyses in language research, see Norouzian, de Miranda, and Plonsky (2018).
4 This intention-based reasoning could also be applicable to minimum standard absolute gradable adjectives (e.g., *open*, *asleep*). In general, minimum absolute gradable adjectives are not considered to be subject to imprecision or pragmatic weakening in the same way as maximum absolute gradable adjectives are. For instance, the expression *The door is open* can describe various degrees of openness from slightly ajar to completely open (Kennedy, 2007). Unlike the case of maximum absolute gradable adjectives, the non–end point interpretations (e.g., a door that is widely open) are not semantically false; in fact, they involve strengthening of the semantic truth-conditions. The challenge and ambiguity that learners would face, however, can resemble those that we discussed with maximum absolute gradable adjectives. Learners routinely observe the same physical scene that can be described as open or not open depending on contextual expectations and the agent's intended goals. We therefore expect that the basic logic laid out here about intention-based reasoning would in principle go beyond cases of imprecision and be applicable to both minimum and maximum absolute gradable adjectives.

References

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, *19*, 347–358. https://doi.org/10.1177/014272379901905703

Aparicio, H., Xiang, M., & Kennedy, C. (2015). Processing gradable adjectives in context: A visual world study. In S. D'Antonio, M. Moroney, & C. R. Little (Eds.),

Proceedings of the 25th Semantics and Linguistic Theory Conference (pp. 413–432). https://doi.org/10.3765/salt.v25i0.3128

Athanasopoulos, P., & Albright, D. (2016). A perceptual learning approach to the Whorfian hypothesis: Supervised classification of motion. *Language Learning*, *66*, 666–689. https://doi.org/10.1111/lang.12180

Athanasopoulos, P., & Bylund, E. (2012). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*, 286–309. https://doi.org/10.1111/cogs.12006

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., … Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, *26*, 518–526. https://doi.org/10.1177/0956797614567509

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*, 395–418. https://doi.org/10.1017/s0305000900008345

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84–93. https://doi.org/10.1016/j.cognition.2010.10.010

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*, 594–608. https://doi.org/10.1111/j.1467-8624.2008.01145.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartsch, R., & Vennemann, T. (1972). Relative adjectives and comparison. *UCLA Papers in Syntax*, *2*, 80–107.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Bloom, P. (1997). Intentionality and word learning. *Trends in Cognitive Sciences*, *1*, 9–12. https://doi.org/10.1016/s1364-6613(97)01006-1

Bloom, P. (2002). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9–25. https://doi.org/10.1080/01621459.1993.10594284

Brooks, N., Audet, J., & Barner, D. (2013). Pragmatic inference, not semantic competence, guides 3-year-olds interpretation of unknown number words. *Developmental Psychology*, *49*, 1066–1075. https://doi.org/10.1037/a0029384

Bürkner, P.-C. (2017). brms: An r package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Burnett, H. (2014). A delineation solution to the puzzles of absolute adjectives. *Linguistics and Philosophy*, *37*, 1–39. https://doi.org/10.1007/s10988-014-9145-9

Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics*, *51*, 929–955. https://doi.org/10.1515/ling-2013-0033

Carpenter, M., Akhtar, N., & Tomasello, M. (1998). Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant behavior and development*, *2*, 315–330. https://doi.org/10.1016/S0163-6383(98)90009-1

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language specific spatial categories in English and Korean. *Cognitive Development*, *14*, 241–268. https://doi.org/10.1016/S0885-2014(99)00004-0

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*, 417–431. https://doi.org/10.1017/S0305000900013842

Clark, E. V. (2020). Perspective-taking and pretend-play: Precursors to figurative language use in young children. *Journal of Pragmatics*, *156*, 100–109. https://doi.org/10.1016/j.pragma.2018.12.012

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge, UK: Cambridge University Press.

Davies, C., & Katsos, N. (2010). Over-informative children: Production/comprehension asymmetry or tolerance to pragmatic violations? *Lingua*, *120*, 1956–1972. https://doi.org/10.1016/j.lingua.2010.02.005

de la Garza, B., & Harris, R. J. (2017). Acquiring foreign language vocabulary through meaningful linguistic context: Where is the limit to vocabulary learning? *Journal of Psycholinguistic Research*, *46*, 395–413. https://doi.org/10.1007/s10936-016-9444-0

Deutsch, W., & Pechmann, T. (1982). Social interaction and the development of definite descriptions. *Cognition*, *11*, 159–184. https://doi.org/10.1016/0010-0277(82)90024-5

Ellis, N. C. (1995). The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning*, *8*, 103–128. https://doi.org/10.1080/0958822940080202

Eskritt, M., Whalen, J., & Lee, K. (2008). Preschoolers can recognize violations of the gricean maxims. *British Journal of Developmental Psychology*, *26*, 435–443. https://doi.org/10.1348/026151007X253260

Estigarribia, B., & Clark, E. V. (2007). Getting and maintaining attention in talk to young children. *Journal of Child Language*, *34*, 799–814. https://doi.org/10.1017/S0305000907008161

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333–375. https://doi.org/10.1016/0024-3841(94)90346-8

Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal

recognition task. *Language and Cognition*, *6*, 45–78. https://doi.org/10.1016/0024-3841(94)90346-8

Frank, M. C. (2014). Learning words through probabilistic inferences about speakers' communicative intentions. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in interaction: Studies in honor of Eve V. Clark* (pp. 207–230). Amsterdam, Netherlands: John Benjamins.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 578–585. https://doi.org/10.1111/2Fj.1467-9280.2009.02335.x

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in judgments of functional size. *Child Development*, *60*, 920–932. https://doi.org/10.2307/1131033

Giacobbe, J. (1992). A cognitive view of the role of L1 in the L2 acquisition process. *Interlanguage Studies Bulletin*, *8*, 232–250. https://doi.org/10.1177/026765839200800304

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135–176. https://doi.org/10.1016/S0010-0277(99)00036-0

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55. https://doi.org/10.1207/s15327817la0101_2

Golinkoff, R. M., & Hirsh-Pasek, K. (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, *12*, 397–403. https://doi.org/10.1016/j.tics.2008.07.003

Goodman, L. A. (1965). On simultaneous confidence intervals for multinomial proportions. *Technometrics*, *7*, 247–254. https://doi.org/10.1080/00401706.1965.10490252

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics* (Vol. 3, pp. 41–58). New York, NY: Academic Press.

Groba, A., De Houwer, A., Mehnert, J., Rossi, S., & Obrig, H. (2018). Bilingual and monolingual children process pragmatic cues differently when learning novel adjectives. *Bilingualism: Language and Cognition*, *21*, 384–402. https://doi.org/10.1017/S1366728917000232

Hall, D. G., Williams, S. G., & Bélanger, J. (2010). Learning count nouns and adjectives: Understanding the contributions of lexical form class and social-pragmatic cues. *Journal of Cognition and Development*, *11*, 86–120. https://doi.org/10.1080/15248370903453592

Hochstein, L., Bale, A., Fox, D., & Barner, D. (2014). Ignorance and inference: Do problems with Gricean epistemic reasoning explain children's difficulty with scalar implicature? *Journal of Semantics*, *33*, 107–135. https://doi.org/10.1093/jos/ffu015

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). What does it take to learn a word? *Monographs of the Society for Research in Child Development*, *65*, 1–16. https://doi.org/10.1111/1540-5834.00091

Horowitz, A. C., & Frank, M. C. (2016). Children's pragmatic inferences as a route for learning about the world. *Child Development*, *87*, 807–819. https://doi.org/10.1111/cdev.12527

Horowitz, A., & Frank, M. (2012). Learning from speaker word choice by assuming adjectives are informative. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 473–478). Austin, TX: Cognitive Science Society.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences*, *20*, 589–604. https://doi.org/10.1016/j.tics.2016.05.011

Jiang, N. (2002). Form–meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition*, *24*, 617–637. https://doi.org/10.1017/S0272263102004047

Kennedy, C. (2007). Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, *30*, 1–45. https://doi.org/10.1007/s10988-006-9008-0

Kennedy, C., & McNally, L. (1999). From event structure to scale structure: Degree modification in deverbal adjectives. In T. Matthews & D. Strolovitch (Eds.), *Proceedings of the 9th Semantics and Linguistic Theory Conference* (pp. 163–180). https://doi.org/10.3765/salt.v9i0.2820

Kennedy, C., & McNally, L. (2005). Scale structure, degree modification, and the semantics of gradable predicates. *Language*, *81*, 345–381. https://doi.org/10.1353/lan.2005.0071

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General*, *139*, 638–653. https://doi.org/10.1037/a0020507

Klein, E. (1980). A semantics for positive and comparative adjectives. *Linguistics and Philosophy*, *4*, 1–45. https://doi.org/10.1007/bf00351812

Kobayashi, H. (1997). The role of actions in making inferences about the shape and material of solid objects among Japanese 2 year-old children. *Cognition*, *63*, 251–269. https://doi.org/10.1016/s0010-0277(97)00007-3

Krashen, S. (1981). *Second language acquisition and second language learning*. Oxford, UK: Pergamon.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning.

Proceedings of the National Academy of Sciences, *100*, 9096–9101. https://doi.org/10.1073/pnas.1532872100

Lasersohn, P. (1999). Pragmatic halos. *Language*, *75*, 522–551. https://doi.org/10.2307/417059

Leffel, T., Xiang, M., & Kennedy, C. (2016). Imprecision is pragmatic: Evidence from referential processing. In M. Moroney, C. R. Little, J. Collard, & D. Burgdorf (Eds.), *Proceedings of the 26th* Semantics and Linguistic Theory Conference (pp. 836–854). https://doi.org/10.3765/salt.v26i0.3937

Levin, B., & Hovav, R. (2005). M*Argument realization*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/cbo9780511610479

Lucy, J. A. (1997). Linguistic relativity. *Annual review of Anthropology*, *26*, 291–312. https://doi.org/10.1146/annurev.anthro.26.1.291

Ludlow, P. (1989). Implicit comparison classes. *Linguistics and Philosophy*, *12*, 519–533. https://doi.org/10.1007/bf00632474

Matthews, D., Lieven, E., & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*, *78*, 1744–1759. https://doi.org/10.1111/j.1467-8624.2007.01098.x

McNally, L. (2009). The relative role of property type and scale structure in explaining the behavior of gradable adjectives. In R. Nouwen, R. van Rooij, U. Sauerland, & H.-C. Schmitz (Eds.), *Vagueness in communication: International workshop, ViC 2009* (pp. 151–168). Heidelberg, Germany: Springer. https://doi.org/10.1007/978-3-642-18446-8_9

McNally, L., & Kennedy, C. (2013). Degree vs. manner well: A case study in selective binding. In J. Pustejovsky, P. Bouillon, H. Isahara, K. Kanzaki, & C. Lee (Eds.), *Advances in generative lexicon theory* (pp. 247–262). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-007-5189-7_11

Mintz, T. H. (2005). Linguistic and conceptual influences on adjective acquisition in 24-and 36-month-olds. *Developmental Psychology*, *41*, 17–29. https://doi.org/10.1037/0012-1649.41.1.1

Montero-Melis, G., & Bylund, E. (2017). Getting the ball rolling: The cross-linguistic conceptualization of caused motion. *Language and Cognition*, *9*, 446–472. https://doi.org/10.1017/langcog.2016.22

Morisseau, T., Davies, C., & Matthews, D. (2013). How do 3-and 5-year-olds respond to under-and over-informative utterances? *Journal of Pragmatics*, *59*, 26–39. https://doi.org/10.1016/j.pragma.2013.03.007

Nagy, W. E., Anderson, R. C., & Herman, P. A. (1987). Learning word meanings from context during normal reading. *American Educational Research Journal*, *24*, 237–270. https://doi.org/10.3102/00028312024002237

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, *20*, 233–253. https://doi.org/10.2307/747758

Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, *90*, 387–401. https://doi.org/10.1111/j.1540-4781.2006.00431.x

Norbury, H. M., Waxman, S. R., & Song, H.-J. (2008). Tight and loose are not created equal: An asymmetry underlying the representation of fit in English- and Korean-speakers. *Cognition*, *109*, 316–325. https://doi.org/10.1016/j.cognition.2008.07.019

Nordmeyer, A. E., & Frank, M. C. (2014). The role of context in young children's comprehension of negation. *Journal of Memory and Language*, *77*, 25–39. https://doi.org/10.1016/j.jml.2014.08.002

Norouzian, R., de Miranda, M., & Plonsky, L. (2018). The Bayesian revolution in second language research: An applied approach. *Language Learning*, *68*, 1032–1075. https://doi.org/10.1111/lang.12310

O'Neill, D. K., Topolovec, J., & Stern-Cavalcante, W. (2002). Feeling sponginess: The importance of descriptive gestures in 2-and 3-year-old children's acquisition of adjectives. *Journal of Cognition and Development*, *3*, 243–277. https://doi.org/10.1207/s15327647jcd0303_1

Papafragou, A., Fairchild, S., Cohen, M. L., & Friedberg, C. (2017). Learning words from speakers with false beliefs. *Journal of Child Language*, *44*, 905–923. https://doi.org/10.1017/s0305000916000301

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, *108*, 155–184. https://doi.org/10.1016/j.cognition.2008.02.007

Paradis, C. (2001). Adjectives and boundedness. *Cognitive Linguistics*, *12*, 47–65. https://doi.org/10.1515/cogl.12.1.47

Park, H. I., & Ziegler, N. (2014). Cognitive shift in the bilingual mind: Spatial concepts in Korean-English bilinguals. *Bilingualism: Language and Cognition*, *17*, 410–430. https://doi.org/10.1017/s1366728913000400

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Francisco, CA: Morgan Kaufman. https://doi.org/10.1016/C2009-0-27609-4

Pinkal, M. (1995). Vagueness and imprecision. In M. Pinkal (Ed.), *Logic and lexicon: Studies in linguistics and philosophy* (Vol. 56, pp. 257–289). Dordrecht, The Netherlands: Springer. https://doi.org/10.1007/978-94-015-8445-6_8

Pinker, S. (1984). Visual cognition: An introduction. *Cognition*, *18*, 1–63. https://doi.org/10.1016/0010-0277(84)90021-0

Poulin-Dubois, D., & Forbes, J. N. (2002). Toddlers' attention to intentions-in-action in learning novel action words. *Developmental Psychology*, *38*, 104–114.

Qing, C., & Franke, M. (2014). Gradable adjectives, vagueness, and optimal language use: A speaker-oriented model. In T. Snider, S. D'Antonio, & M. Weigand (Eds.), *Proceedings of the 24th Semantics and Linguistic Theory Conference* (pp. 23–41). https://doi.org/10.3765/salt.v24i0.2412

R Core Team. (2016). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ramscar, M., Dye, M., & Klein, J. (2013). Children value informativity over logic in word learning. *Psychological Science*, *24*, 1017–1023. https://doi.org/10.1177/0956797612460691

Ringbom, H. (1983). Borrowing and lexical transfer. *Applied Linguistics*, *4*, 207–212. https://doi.org/10.1093/applin/4.3.207

Rotstein, C., & Winter, Y. (2004). Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics*, *12*, 259–288. https://doi.org/10.1023/b:nAls.0000034517.56898.9a

Rusiecki, J. (1985). *Adjectives and comparison in English: A semantic study*. London, UK: Longman.

Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, *23*, 278–290. https://doi.org/10.1016/j.cogdev.2008.01.003

Slobin, D. I. (1996). Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 195–217). Oxford, UK: Clarendon.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Smith, N. J., Goodman, N. D., & Frank, M. C. (2013). Learning and using language via recursive pragmatic reasoning about other agents. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 3039–3047). Cambridge, MA: MIT Press.

Solt, S. (2015). Vagueness and imprecision: Empirical foundations. *Annual Review of Linguistics*, *1*, 107–127. https://doi.org/10.1146/annurev-linguist-030514-125150

Sonnenschein, S. (1982). The effects of redundant communications on listeners: When more is less. *Child Development*, *53*, 717–729. https://doi.org/10.2307/1129385

Stechow, A. v. (1984). Comparing semantic theories of comparison. *Journal of Semantics*, *3*, 1–77. https://doi.org/10.1093/jos/3.1-2.1

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning*, *55*, 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Syrett, K., Bradley, E., Kennedy, C., & Lidz, J. (2006). Shifting standards: Children's understanding of gradable adjectives. In *The Proceedings of the Inaugural Conference on Generative Approaches to Language Acquisition–North America* (Vol. 2, 345–352). https://doi.org/10.7282/T39W0HHS

Syrett, K., Kennedy, C., & Lidz, J. (2010). Meaning and context in children's understanding of gradable adjectives. *Journal of Semantics*, *27*, 1–35. https://doi.org/10.1093/jos/ffp011

Syrett, K., & Lidz, J. (2010). 30-month-olds use the distribution and meaning of adverbs to interpret novel adjectives. *Language Learning and Development*, *6*, 258–282. https://doi.org/10.1080/15475440903507905

Talmy, L. (1983). How language structures space. In H. L. Pick, Jr & L. P. Acredolo (Eds.), *Spatial orientation: Theory, research, and application* (pp. 225–282). New York, NY: Springer. https://doi.org/10.1007/978-1-4615-9325-6_11

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences*, *106*, 4567–4570. https://doi.org/10.1073/pnas.0811155106

Toledo, A., & Sassoon, G. W. (2011). Absolute vs. relative adjectives—variance within vs. between individuals. In N. Ashton, A. Chereches, & D. Lutz (Eds.), *Proceedings of the 21st Semantics and Linguistic Theory Conference* (pp. 135–154). https://doi.org/10.3765/salt.v21i0.2587

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). New York, NY: Erlbaum.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*, 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Unger, P. (1975). *Ignorance: A case for scepticism*. Oxford, UK: Clarendon. https://doi.org/10.2307/2218871

Vanek, N. (2019). Changing event categorization in second language users through perceptual learning. *Language Learning*, *70*, 309–348. https://doi.org/10.1111/2Flang.12377

van Rooij, R. (2011). Vagueness and linguistics. In G. Ronzitti (Ed.), *Vagueness: A guide* (pp. 123–170). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/2F978-94-007-0375-9_6

Vendler, Z. (1967). *Linguistics in philosophy*. Cornell University Press.

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics*, *50*, 833–867. https://doi.org/10.1515/2Fling-2012-0026

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*, 729–742. https://doi.org/10.1016/j.cognition.2007.08.007

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*, 1611–1617. https://doi.org/10.1037/a0016134

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Tables of Analysis Output.
Appendix S2. Norming Study for Visual (Video) Stimuli.
Appendix S3. Exposure Verbal Narrations.
Appendix S4. Test Stimuli.
Appendix S5. Experiment 4, Positive Evidence Only.
Appendix S6. Explicit *Pelty* Interpretations.
Appendix S7. Bayesian Analyses Examining Joint Distribution of Responses.

Appendix: Accessible Summary (also publicly available at https://oasis-database.org)

Understanding Others' Intentions Supports the Learning of Adjective Meanings

What This Research Was About and Why It Is Important

Everyday language use is heavily dependent on context and often considered "imprecise". For example, one can label a cup that is only 80% full as "full" although it is false based on the literal sense of the word. Moreover, the same 80% full cup is not considered "full" when used for following a recipe. This imprecision and context-dependency present a major problem for word learners. How do adult learners of English come to know the precise meaning of words like "full" if objects described by these words can vary across contexts? We addressed this by constructing and teaching a novel adjective (*pelty*, meaning "tight-fitting") in different contextual conditions. We showed a series of videos in which objects exhibited different degrees of tightness of fit, and were described by a speaker as *pelty* or not *pelty*. Just as in the cases of *full* or not *full*, this novel adjective applied to a given scene only sometimes. We predicted that adult learners would interpret distinct meanings depending on the availability of information about intention behind the word use (e.g., a mostly tight-fitting shoe is labeled as *pelty* when the speaker *intends to* wear it with a thick sock). Results showed that when learners received information relevant to a speaker's intentions, they were more likely to interpret *pelty* to mean

"completely tight-fitting," despite varying degrees of fitness seen during exposure to the word. On the other hand, when learners did not receive information about the speaker's intention, they interpreted *pelty* to mean "relatively tight-fitting," even though the visual context was exactly the same (i.e., only the presence/absence of speaker's intention was different). Our findings provide novel evidence that learners actively incorporate their understanding of the context, such as the intentions of others, when learning the meaning of a new adjective. This suggests that nonlinguistic, contextual, information can help shape word learning in naturalistic contexts.

What the Researchers Did

- In three studies, we taught adult participants a novel adjective intended to refer to tightness of fit between two objects.
- Participants were shown 24 examples of the novel adjective, depicting varying degrees of tight-ness with information about whether the scene can be described by the adjective or not.
- Participants were assigned to one of two conditions that either provided information about the speaker's intentions or not. The participants provided their interpretations of the meaning of *pelty*.
- Types and numbers of scenes that participants were exposed to were systematically manipulated across three experiments.

What the Researchers Found

- We found that participants who had access to a speaker's intentions inferred a "completely tight-fitting" meaning for the adjective compared to those who did not (Experiment 1).
- This was also the case even when the visual variety of examples shown increased (Experiment 2).
- However, participants did not infer a "completely tight-fitting" interpretation (relative to Experiments 1 and 2) when given *only* imprecise visual examples (Experiment 3).

Things to Consider

- We concluded that learners use information from the context in which the language is used, including inferred intention of the speaker, to acquire a novel word meaning.
- The consideration of contextual information may be different depending on the individual (e.g., adults vs. children).

- Future research can extend this line of thinking to examine other types of linguistic constructions, such as verbs.

Materials and data: Materials and data are publicly available at https://osf.io/b9mah.

How to cite this summary: Lee, C., & Kurumada, C. (2020). Understanding others' intentions supports the learning of adjective meanings. *OASIS Summary* of Lee & Kurumada (2021) in *Language Learning*. https://oasis-database.org