The structure of the lexicon reflects principles of communication

Molly Lewis
mll@stanford.edu
Department of Psychology
Stanford University

Elise Sugarman
elisesug@stanford.edu
Department of Psychology
Stanford University

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

Abstract

We explore a previously undescribed regularity in language: a bias for longer words to map to relatively more complex meanings. While theories of communication make the more general prediction that longer utterances should be associated with more complex meanings, this prediction has not yet been explored at the level of words. Through norming and experimental studies, we find evidence in support of a complexity bias in word meanings. We conclude by discussing hypotheses about the nature of this bias.

Keywords: communication; pragmatics

Introduction

Is the structure of language shaped by its use in communication? Hockett (1960) argued that several major features of language—from its productivity to its arbitrary mappings from signifier to signified—emerge from communicative pressures. This kind of functional explanation has since been proposed for a wide range of phenomena, from the presence of ambiguity in language (Piantadosi, Tily, & Gibson, 2011a) to the way languages organize the semantic space of kinship meaning (Kemp & Regier, 2012). In our current study, we identify a regularity in the lexicon—that more complex ideas tend to be named by longer words—and investigate a potential communicative explanation.

Several theories of communication predict that longer utterances should be associated with more complex meanings (a *complexity bias*). More specifically, they predict that communicative pressures should lead to a complexity bias that operates at the moment of language interpretation. Horn (1984) proposed one such theory. He argued for a relationship between the cost of an utterance and the probability of meaning. He suggested that costlier phrases are associated with less probable meanings. For example, the phrases "I started the car" and "I got the car to start" are denotationally equivalent (they both denote the successful starting of a car), but they differ in cost in terms of their utterance length. Horn's principle accounts for this difference as an asymmetry in the probabilities of meanings: The shorter form refers to the typical (more frequent) starting of a car, and thus the longer form refers to an atypical case, perhaps where the driver encountered difficulty.

Information theory also predicts a complexity bias that operates during online language interpretation. Under this theory, speakers optimize information transfer (in terms of bits) by keeping the amount of information conveyed in a unit of language constant across the speech stream (Aylett & Turk, 2004; Frank & Jaeger, 2008). This "uniform information density" hypothesis predicts as a straightforward consequence that a longer utterance should have a less predictable meaning. Consistent with this prediction, speakers tend to increase the duration of a word prosodically in cases where the word is unpredictable (highly informative) given the linguistic context (Aylett & Turk, 2004).

Recent experimental data provide additional evidence for a complexity bias operating at the moment of language interpretation. For example, in an artificial language learning task, Fedzechkina, Jaeger, and Newport (2012) found that learners tended to use case marking (i.e. longer referential forms) in cases where the sentence meaning was less predictable in the experimental context. Bergen, Goodman, and Levy (2012) provide a more direct test of a complexity bias. In their task, partners were told that they were in an alien world with three objects of different base rate frequencies and three possible utterances of different monetary costs. Their task was to communicate about one of the objects using one of the available utterances. The results suggest that both the speaker and hearer expected costlier forms to refer to less frequent meanings, consistent with Horn's principle.

A number of pieces of evidence thus corroborate proposals that communicative pressures lead to a complexity bias at the moment of language interpretation. In our work here, we asked: Do these communicative pressures have consequences beyond in-the-moment language choices? That is, do these communicative pressures lead to the instantiation of a complexity bias in the lexicon? While this question has not been previously investigated, there is some evidence to suggest that they might: words that are more predictable in their linguistic context are found to be shorter than words that are less predictable in their linguistic context (Piantadosi, Tily, & Gibson, 2011b).

A substantial challenge in studying a complexity bias is the difficulty of defining complexity. We give an intuitive definition guiding our investigation, but we note that our operationalization below is also consistent with other definitions. Imagine a space of possible meanings as compositional semantic primitives. In this space, a more complex meaning would be one with more primitives in it. (In a probabilistic framework, having more units would also be correlated with having a lower overall probability).

We explore a complexity bias through norming and experimental studies. Analyses of norming data suggest that longer English words tend to have more complex meanings. Evidence from three experiments also suggests that a complexity bias guides inference in a novel word learning task, in both adults (Exp. 1 and 2) and preschoolers (Exp. 3). We conclude by discussing hypotheses about the source of these results and the relationship between them.

Complexity bias in the lexicon

To explore whether there is a complexity bias in the lexicon, we asked adults to rate the meanings of English words for their complexity. If there is a complexity bias in the lexicon, we predicted that the meanings of longer words should be rated as more complex, relative to shorter words.

Method

Participants

We recruited 250 participants from Amazon Mechanical Turk, but excluded 10 for earlier participation in related experiments or failure to correctly answer a control question (answering a simple math problem). In this and the subsequent experiments, all reported results remain statistically significant when including these participants.

Stimuli

We selected 500 English words that were broadly distributed in their length. All of these words were also included in the MRC Psycholinguistic Database (Wilson, 1988). This database includes norms for three other psycholinguistic variables: concreteness, familiarity, and imageability. This allowed us to compare our complexity norms to previously measured psycholinguistic variables.

Procedure

Participants first viewed a webpage that described the norming task. The instructions read:

> In this experiment, you will be asked to decide how complex the meaning of a word is. A word's meaning is simple if it is easy to understand and has few parts. An example of a simple meaning is "brick." A word's meaning is complex if it is difficult to understand and has many parts. An example of a more complex meaning is "engine."

They then rated 32 words for their complexity on a 5-point Likert scale. For all participants, the first two items were "ball" and "motherboard" in order to anchor participants on the complexity scale. After the 17th word, participants were asked to complete a simple math problem to ensure they were engaged in the task.

Results

Word length in terms of number of characters was highly correlated with complexity ratings ($r = .66,\ p < .01$).[1] This relationship held for other metrics of length for words where this information was available (phonemes: $r = .69,\ p < .01$; syllables: $r = .63,\ p < .01$).

Next we examined whether there was a systematic relationship between word length and complexity, controlling for other variables. We created a linear model predicting word length with complexity, controlling for concreteness, imageability and familiarity. We also controlled for the log spoken frequency of each word. Frequency was estimated from a corpus of transcripts from American English movies (Subtlex-us database; Brysbaert & New, 2009).

[1] All code and data available at `http://github.com/mllewis/refComplex-cogsci`

	Estimate	Std. Error	t value	Pr($>|t|$)
(Intercept)	5.6484	0.0707	79.87	0.001
Complexity	1.1550	0.0878	13.16	0.001
Log Frequency	-0.6861	0.0893	-7.68	0.001
Concreteness	-0.2148	0.1683	-1.28	0.203
Imageability	0.2204	0.1611	1.37	0.172

Table 1: Model parameters for linear regression predicting word length in terms of number of characters. Coefficient estimates are standardized, so they can be interpreted as number of characters per standard deviation.

Familiarity and log frequency were highly colinear ($r = .77,\ p < .01$), and so our final model included only complexity, concreteness, imageability, and frequency as predictors. Complexity and frequency reliably predicted word length in terms of number of characters. Concreteness and imageability were not reliable predictors. The estimates for this model are presented in Table 1. The same pattern held for the other two length metrics (phonemes and syllables).

This analysis suggests there is a complexity bias in the lexicon: more complex meanings tend to be encoded in language with longer forms.

Complexity bias at the moment of language interpretation

In addition to a complexity bias in the lexicon, we also predicted that speakers might be able to make use of this regularity to guide reference selection in a novel word learning task. We tested this prediction experimentally in three experiments with adults and children. In each experiment, we presented participants with two objects and a novel word, and asked participants to identify the referent. One of the objects was visually simple and the other visually complex. Across trials, we manipulated the length of the word. If a complexity bias guides reference selection, participants should be more likely to map a long word to a complex object, relative to a simple object. The data support this hypothesis, suggesting that this bias is present in adults (Exp. 1 and 2) and develops in preschoolers (Exp. 3).

Experiment 1a

In Experiment 1, we probed the complexity bias in adults with a forced choice (Exp. 1a) and betting measure (Exp. 1b).

Methods

Participants. We recruited 60 participants from Amazon Mechanical Turk, but excluded 11 for earlier participation in related experiments or failure to correctly answer a control question (identifying a familiar object).

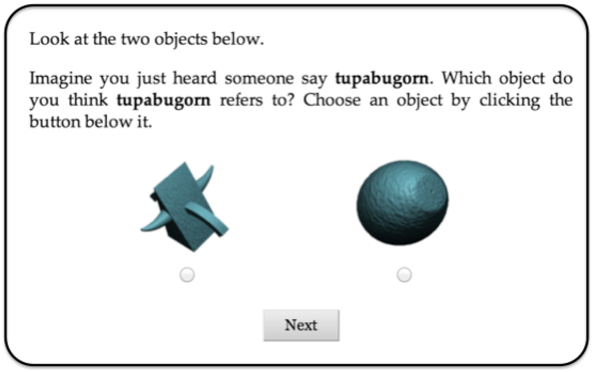

Figure 1: Example trial from Experiment 1a.

Stimuli. The referents were objects composed of a varying number of geometrical shapes. Each geometric shape was a geon — a representation that has been argued to be the primitive unit in the visual system for object recognition (Biederman, 1987). The simple objects were composed of a single geon and the complex objects were composed of five geons. There were eight simple items and eight complex items. The actual stimuli were adopted from Hayward and Tarr (1997). The linguistic stimuli were novel words. There were seven short words composed of two syllables (e.g., "tupa," "gabu," "fepo") and seven long words composed of four syllables (e.g., "tupabugorn," "gaburatum," "fepolopus").

Procedure. Participants viewed a webpage that showed two objects and a word, and were asked to select the referent (see Fig. 1). Each participant completed four critical trials followed by a control trial. In the control trial, participants were asked to identify the referent among two known objects (e.g. cactus and sandwich).

We manipulated two factors: word length and object complexity. On each trial, a simple and a complex object were presented. Word length was manipulated between participants, such that each participant saw only one word length across all four trials.

Results Participants selected the complex referent significantly more often in the long word condition, relative to the short word condition ($\chi^2(1) = 6.71$, $p < .01$, $d = .46$; Fig. 2).

Experiment 1b

Methods

Participants. We recruited 90 participants from Amazon Mechanical Turk. Five were excluded for earlier participation in related experiments or failure to correctly answer the control question.

Stimuli. The stimuli were identical to Experiment 1a.

Design. The procedure was identical to Experiment 1a except for the new measure. In this version of the experiment, participants were asked to provide bets 0-100 indicating their judgement, with a constraint that the bets on the two objects summed to 100. In this experiment, we also introduced baseline trials by manipulating the referential alternatives available across participants. Participants viewed either two simple objects (baseline), two complex objects (baseline), or a simple and a complex object (critical).

Results In the critical alternatives condition, participants selected the complex referent significantly more often when they saw a long word compared to short word ($t(134) = 2.44, p < .05, d = .42$; Fig. 2). There was no linguistic difference in the baseline conditions (complex-complex: $t(106) = -1.07, p = .29$; simple-simple: $t(94) = 0.88, p = .38$).

Experiment 2

The results from Experiment 1 suggest that adults have a complexity bias when interpreting the meaning of a novel word. In Experiment 2, we further explored this bias using more naturalistic stimuli for the objects. Instead of geons, we presented participants with pictures of real objects that had been normed for visual complexity.

Methods

Participants. Ninety-two adults were recruited from Amazon Mechanical Turk. Thirty-three participants were excluded for participating in related experiments.

Stimuli. The object stimuli were taken from a set of images normed for complexity. In the norming task, we presented participants with 10 randomly selected novel objects from the full set of 60. For each object, participants were asked, "How complicated is this object?," and then indicated their response using a slider scale. Ratings were collected from 60 participants. A second sample of 60 participants gave ratings that were highly correlated with those of the first sample, $r = .94$. Figure 3 shows all images sorted by complexity.

In the present experiment, 11 images from the bottom quartile of the normed objects and 15 from the top quartile were selected as the simple and complex objects (Fig. 3). The linguistic stimuli were the same as in Exp. 1.

Procedure. The procedure was identical to the betting procedure used in Exp. 1b, with the exceptions that each participant completed only a single critical trial and there were no baseline conditions.

Results As in Experiment 1, participants selected the complex referent significantly more often in the long word condition compared to the short word condition ($t(57) = 3.54, p < .001, d = .92$; Fig. 2).

Experiment 3

Experiments 1 and 2 suggest that adults have a complexity bias when inferring the meaning of a novel word. Given evidence that a complexity bias is present in the lexicon, the ability to adopt a complexity bias in the moment of reference could potentially help constrain referential uncertainty for

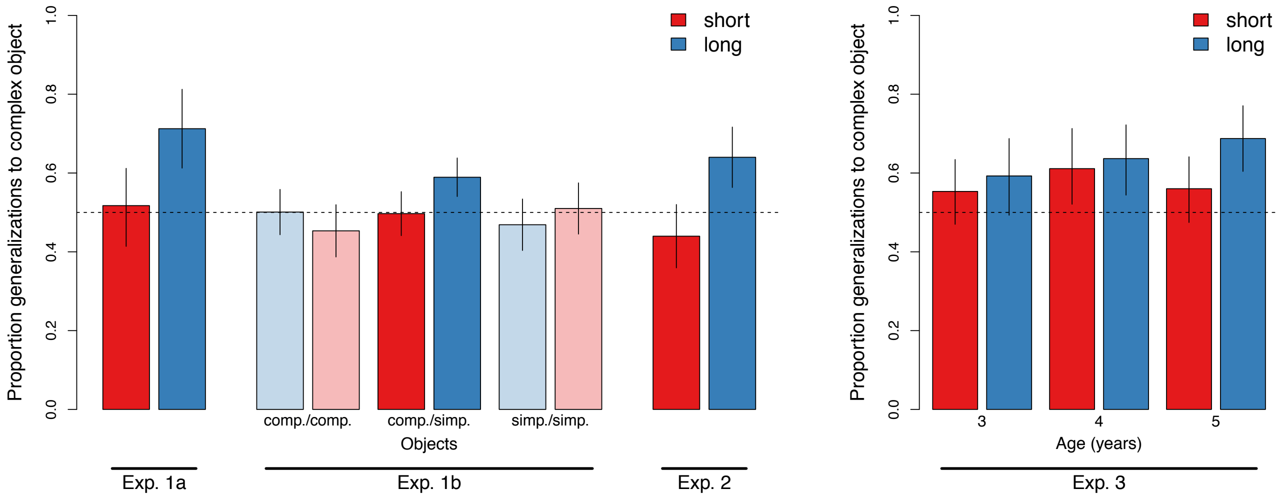

Figure 2: Mean proportion selections (Exp. 1a and 3) or mean bets (Exp. 1b and 2) to complex object for adults (left) and children (right). Error bars represent 95% CIs as computed via non-parametric bootstrap (Exp. 1a) and parametric 95% CIs (Exp. 1b, 2, 3). In Experiment 1b, transparent bars represent the two baseline conditions (two complex object alternatives and two simple object alternatives).

young language learners. In Experiment 3, we explored whether preschool age children adopt a complexity bias in an iPad adaptation of Experiment 2.

Methods

Participants. We recruited 108 children (60 girls) from a nursery school at Stanford University and the San Jose Children's Discovery Museum (36 3-year-olds, *M*= 3;8; 36 4-year-olds, *M*= 4;5; 36 5-year-olds, *M*= 5;5).

Stimuli. The object and linguistic stimuli were identical to those used in Experiment 2.

Procedure. The procedure was identical to Experiment 2 with a few changes necessary to adapt the experiment for children. Children first completed a training phase to gain familiarity with the iPad's touchscreen. They were then introduced to a puppet named Furble and told that "he speaks a different language from us. He has some toys that will look familiar, but he has lots of funny-looking toys, too. I need you to help me pick Furble's toys. Furble's going to say a word, and you're going to pick which toy you think the word is for." Children were then shown a display with two images, and the experimenter asked the child in child-directed prosody to select a referent using one of three naming frames ("Can you find the [label]?," "Where's the [label]?," or "Can you pick the [label]?").

We manipulated word length within subject. Children completed 12 trials: four trials with long words, four trials with short words, and four trials with familiar words. Familiar word trials involved a familiar word and two familiar objects. These were included to familiarize children with the task of reference selection. Order of trials was randomized with the constraint that two familiar trials always appeared first.

Results Figure 2 (right) illustrates the proportion of children selecting the complex referent in the short and long word conditions. Five percent of trials were excluded in cases where the child had difficulty operating the iPad, or where there were technical difficulties. Across the three age groups, a complexity bias emerged only in 5-year-olds.

To test the reliability of this difference, we ran a generalized linear mixed model predicting object selection as an interaction between condition (long or short word) and age with random effects of participant and trial number. There was a main effect of age ($\beta = -0.40$, $p < .01$), indicating that children were overall more likely to select the complex referent with increasing age. There was also a reliable interaction between age and condition ($\beta = 0.38$, $p < .05$). The effect of condition was not significant.

We ran a series of paired *t*-tests to examine response differences between conditions for each age group (3s, 4s, and 5s). There was a reliable difference between conditions for the 5-year-old age group ($t(35) = 2.35, p < .05$, $d = .48$), but not the younger groups. These results suggest that older, but not younger, preschoolers have an adult-like bias to map longer words onto more complex referents.

General Discussion

A number of theories of communication predict that longer utterances should be associated with more complex meanings in the moment of language interpretation. We explored this prediction at the level of words. Using two sources of data — correlational evidence within actual spoken language and experimental evidence with novel words — we found evidence that longer words tend to be associated with more complex

Figure 3: Novel objects sorted by mean complexity rating. Top left object corresponds to the lowest-rated object, and the bottom right object corresponds to the highest-rated object. Objects from the bottom (red) and top (blue) quartile were used as stimuli in Exp. 2 and 3.

referents.

While both sources of evidence reveal a complexity bias in the pattern of observable data, each has different implications for the origins of the bias. In particular, each piece of evidence points to the emergence of a complexity bias at a different timescale (McMurray, Horst, & Samuelson, 2012). The norming study suggests that the bias may be grammaticalized in the lexicon (language change timescale). The experimental work with children suggests that a complexity bias could shape children's learning through the process of word learning (language acquisition timescale). And, the experimental work with adults suggests that a complexity bias operates during individual episodes of language use (pragmatic timescale).

How, if at all, are complexity biases at these three timescales related? Three broad hypotheses are consistent with our findings: A complexity bias (1) is innately given, (2) emerges as an artifact of the emergence of language, or (3) emerges from communicative pressures. Below we outline these alternatives in more detail.

1. The Innate Bias Hypothesis. One explanation of our data is that humans have an innate bias to map long words to complex referents. Under this hypothesis, a cognitive bias would lead to a complexity bias in the moment of language use. Over time, this behavioral regularity might become a probabilistic rule, becoming "grammaticalized" into the language. A challenge for this hypothesis is accounting for why a behavioral bias was not observed in young preschoolers in Experiment 3. Nonetheless, it has been argued in the case of other innate constraints that early age of onset is not a necessary condition for innateness (e.g., Markman, 1992).

2. The Efficient Naming Hypothesis. A second class of hypotheses posits that the complexity bias in the lexicon emerged independently of communicative pressures. To understand one variant of this hypothesis, consider the following fable: At the beginning of linguistic time, names were assigned to objects by length, starting with the shortest. Objects were named in the order that they were observed. Since more frequent (i.e. higher probability and hence likely less complex, see Introduction) objects tended to be observed earlier, these objects received shorter names, relative to the less frequent objects that were encountered later. This story provides one possible account of the emergence of a complexity bias over the language change timescale.

Under this hypothesis, there are several ways to account for a pragmatic in-the-moment complexity bias. One possibility is that the lexical bias and the in-the-moment bias are the result of independent causal processes: the lexical bias may be the result of an efficient naming strategy, while the in-the-moment complexity bias may be the product of general pragmatic reasoning.

An alternative possibility is that the in-the-moment bias emerged from a generalization, or overhypothesis (Kemp, Perfors, & Tenenbaum, 2007), based on observations about the lexicon. That is, given experience with a lexicon that contains a regularity to map longer words to more complex meanings, learners might have induced a complexity regularity about the lexicon. Thus, when faced with a novel word, speakers might apply this bias as a probabilistic heuristic about the meaning of the word. An overhypothesis account of the behavioral data has the advantage that it is able to account for the development in an in-the-moment complexity bias in preschoolers. Under this account, preschoolers might not show this behavioral bias because they have not yet observed enough data to induce a complexity overhypothesis.

3. The Communicative Pressure Hypothesis. A final possibility is that humans are predisposed to consider the intentions of others (Tomasello et al., 2005). This predisposition leads to pragmatic reasoning in the moment of language use. As argued by Horn (1984), one type of inference that might be guided by pragmatic reasoning is that a costlier (i.e. longer) utterance is more likely to refer to a more complex meaning. Thus, under this hypothesis, domain-general pragmatic reasoning may underly an in-the-moment complexity bias. Over time, this in-the-moment bias may become grammaticalized, leading to a complexity bias in the lexicon.

A variant of this hypothesis is that an in-the-moment behavioral bias might be the product of both an underlying pragmatic inference and an overhypothesis about a complexity regularity in the lexicon. This possibility is similar to one account of a different, well-studied bias in word learning (Lewis & Frank, 2013): the mutual exclusivity bias. The mutual exclusivity bias is the tendency for children to map a novel word onto a novel object. In this work, we suggest that the mutual exclusivity behavior may emerge from both an in-the-moment pragmatic inference (the speaker would have used the known word to refer to a known object if that was the intended referent, and so the novel word must refer to the novel object) and an overhypothesis about the structure of the lexicon (a 1-1 mapping between words and concepts). We argued that processes at both timescales may jointly contribute to mutual exclusivity behavior.

In conclusion, our work provides the first analysis of the relationship between utterance length and referent complexity at the level of words. We find evidence suggesting that a complexity bias is present in the lexicon and that speakers make use of this regularity in a word learning task. On grounds of parsimony, we favor the Communicative Pressure Hypothesis as an account of these data. However, future work is needed to more fully understand the relationship between complexity biases at different timescales.

Acknowledgments

We would like to thank the staff and families at the Bing Nursery School and San Jose Children's Discovery Museum.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*, 31–56.

Bergen, L., Goodman, N. D., & Levy, R. (2012). That's what she (could have) said: How alternative utterances affect language use. In *Proceedings of the Thirty-Fourth Annual Meeting of the Cognitive Science Society.*

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94*, 115.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, *109*, 17897–17902.

Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society.*

Hayward, W. G., & Tarr, M. J. (1997). Testing conditions for viewpoint invariance in object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1511.

Hockett, C. (1960). The origin of speech. *Scientific American*, *203*, 88-96.

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. *Meaning, Form, and Use in Context*, *42*.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*, 307-321.

Kemp, C., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*, 1049–1054.

Lewis, M., & Frank, M. C. (2013). Modeling disambiguation in word learning via multiple probabilistic constraints. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society.*

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity.

McMurray, B., Horst, J., & Samuelson, L. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*, 831.

Piantadosi, S., Tily, H., & Gibson, E. (2011a). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291.

Piantadosi, S., Tily, H., & Gibson, E. (2011b). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*, 3526–3529.

Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H., et al. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–690.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*, 6–10.