The length of words reflects their conceptual complexity

Molly L. Lewis

Department of Psychology, Stanford University

Michael C. Frank

Department of Psychology, Stanford University

Address all correspondence to Molly L. Lewis, Stanford University, Department of Psychology, Jordan Hall, 450 Serra Mall (Bldg. 420), Stanford, CA, 94305. Phone: 650-721-9270. E-mail: mll@stanford.edu

Abstract

Are the forms of words systematically related to their meaning? The arbitrariness of the sign has long been a foundational part of our understanding of human language. Theories of communication predict a relationship between length and meaning, however: Longer descriptions should be more conceptually complex. Here we show that both the lexicons of human languages and individual speakers encode the relationship between linguistic and conceptual complexity. Experimentally, participants mapped longer words to more complex objects in comprehension and production tasks and across a range of stimuli. Explicit judgments of conceptual complexity were also highly correlated with implicit measures of study time in a memory task, suggesting that complexity is directly related to basic cognitive processes. Observationally, judgements of conceptual complexity for a sample of real words correlate highly with their length across 80 languages, even controlling for frequency, familiarity, imageability, and concreteness. While word lengths are systematically related to usage—both frequency and contextual predictability—our results reveal a systematic relationship with meaning as well. They point to a general regularity in the design of lexicons and reinforce the importance of cognitive constraints on language evolution.

Keywords: xx

Introduction

Human languages are systems for encoding information about the world. A defining feature of a symbolic coding system is that there is no inherent mapping between the form of the code and what the code denotes (Peirce, 1931)—the color red holds no natural relationship to the meaning 'stop', the numeral *3* holds no natural relationship to three units, and in language, the word 'horse' looks or sounds nothing like the four-legged mammal it denotes. The arbitrariness of the linguistic sign has long been observed as a fundamental and universal property of natural language (Saussure, 1916, 1960; Hockett, 1960). And, despite the growing number of cases suggesting instances of non-arbitrariness in the lexicon (Schmidtke, Conrad, & Jacobs, 2014, for a review), there is clear evidence for at least some degree of arbitrariness in language based only on the observation that different languages use different words to denote the same meaning (e.g. the word for horse in English is "horse" but is "at" in Turkish).

Importantly, however, the arbitrary property of language is true only from the perspective of the outside analyst. From the perspective of the language user, language is highly non-arbitrary; there are strong constraints on how language is used in particular instances of communication. A rich body of theoretical work has explored regularities in the use of particular forms to refer to particular types of meanings in context—the study of *pragmatics* (Grice, 1975; Horn, 1984; H. H. Clark, 1996). Broadly, this work argues that language users assume certain regularities in how speakers refer to meanings, and through these shared assumptions, the symmetry of the otherwise arbitrary character of language is broken. For example, consider a speaker who intends to refer to a particular apple on a table. Because language is *a priori* arbitrary, there are a range of ways the speaker could convey this meaning (e.g, "the apple," "the banana," "the green apple,""the green apple next to the plate," etc.), but the speaker is constrained by pragmatic pressures of the communicative context. If the listener also speaks English, the phrase "the banana" will be an unhelpful way to refer to the apple. Furthermore, if there is only one apple on the table, the phrase "the green apple" will be unnecessarily verbose given the referential context. These constraints

might lead a speaker to select "the apple" as the referential phrase, because it both allows the listener to correctly identify the intended referent while also minimizing effort on the part of the speaker.

In the present paper, we examine whether principles of communication influence the otherwise arbitrary mappings between words and meanings in the lexicon. In particular, we explore a communicative regularity first observed by Horn (1984): pragmatic language users tend to consider the effort that speakers have exerted to convey a meaning. For example, the utterance "Lee got the car to stop" seems to imply an unusual state of affairs. Had the speaker wished to convey that Lee simply applied the brakes, the shorter and less exceptional "Lee stopped the car" would be a better description. The use of a longer utterance licenses the inference that there was some problem in stopping—perhaps the brakes failed—and that the situation is more complex. Do we reason the same way about the meanings of words, breaking the symmetry between two unknown meanings by reference to length? In other words, is a "tupabugorn" more likely to be a complex, unusual object than a "ralex"? We examine whether this regularity—a *complexity bias*—is present in the lexicon.

The present hypothesis is motivated by the possibility that language dynamics take place over different timescales, and these different dynamics may be causally related to each other. Minimally, two timescales are relevant to the present hypothesis. At the shorter timescale are the minutes of a single communicative interaction—*the pragmatic timescale*. At the longer timescale is language change, which takes place over many years—*the language evolution timescale*. We consider the possibility that communicative pressures at the pragmatic timescale may, over time, influence the structure of the lexicon at the language evolution timescale. There are other reasons why a regularity like a complexity bias might emerge in the structure of the lexicon, however, and we consider some of these alternative possibilities in the General Discussion.

[paragraph about what we mean by complexity?]

The plan of the paper is as follows. We first review prior work suggesting that

communicative principles are reflected in the structure of the lexicon. We then review work related to accounts of our particular linguistic feature of interest—variability in the length of forms. Next, we present four sets of studies that explore a complexity bias in the lexicon. In Studies 1 and 2, we experimentally test whether participants are biased to map a relatively long novel word onto a relatively more complex object, using artificial objects (Study 1) and novel, real objects (Study 2). In Study 3, we explore the underlying cognitive construct of complexity in a reaction time task. Finally, in Study 4, we examine a complexity bias in natural language through a corpus analysis of 80 languages.

Pragmatic equilibria in the lexicon

Several broad theories of pragmatics include a version of two distinct pressures on communication: the desire to minimize effort in speaking (*speaker pressure*) and the desire to be informative (*hearer pressure;* Zipf, 1936; Horn, 1984). Importantly, these two pressures tradeoff with each other: the optimal solution to the speaker's pressure is a single word that can refer to all meanings, while the optimal solution to the hearer's pressure is a verbose, minimally ambiguous phrase. The utterance that emerges is argued to be an equilibrium between these two tradeoffs.

At the timescale of language evolution, there a number of cases in which these pragmatic equilibria are reflected in the lexicon. One way these equilibria are reflected is in the size of the semantic space denoted by a particular word. From the hearer's perspective, Horn argues there is a pressure to narrow semantic space (Horn, 1984). This reflects the idea that the hearer's optimal language is one in which every possible meaning receives its own word. One example of this is the word "rectangle." This word refers to a quadrilateral with four right angles. A special case of a "rectangle" is a case where the four sides are equal in length, which has its own special name, "square." Consequently, the term "rectangle" has been narrowed to mean a quadrilateral with four right angles, where the four sides are *not* equal. From the speaker's perspective, there is a pressure for semantic broadening. This is because the speaker's ideal language is one in which a single

word can refer to a wide range of meanings. An example of this is the broadening of brand names to refer to a kind of product. For example, "kleenex" is a name of a product name for facial tissues, but has taken on the meaning of facial tissues more generally.

The opposition of these two semantic forces predicts an equilibrium in the organization of semantic space that satisfies the pressures of both speaker and hearer. A body of empirical work has tested this prediction by examining the organization of particular semantic domains cross-linguistically (Regier, Kemp, & Kay, 2014). Languages show a large degree of similarity in how they partition semantic space for a particular domain, which is likely due to universal cognitive constraints. But, they also show a large degree of variability and these different systems can be shown to all approximate an equilibrium point between speaker and hearer pressures.

Kemp and Regier (2012) demonstrate this systematicity in the semantic domain of kinship. For each language, they developed a metric of the degree to which Horn's speaker and hearer pressures (in their terminology: communicative cost and complexity, respectively) are satisfied. A language that better satisfies the hearer's pressure is one that is more complex, as measured by the length of the description of the system in their representational language. A language that better satisfies the speaker's pressure is one that requires less language to describe the intended referent. To understand this, consider the word "grandmother" in English: this word is ambiguous in English because it could refer to either the maternal or paternal mother, and so identifying one in particular is more costly in English than in a language that encodes this distinction lexically. They find that the set of attested languages is a subset of the range of possible languages, and this subset partitions the semantic space in a way that is near the optimal tradeoff between pragmatic pressures. This type of analysis has also been done for the domains of color (Regier, Kay, & Khetarpal, 2007), light (Baddeley & Attewell, 2009), and numerosity (Xu & Regier, 2014).

A second phenomenon that is predicted by these forces is cases where there are multiple meanings associated with a word from a context-independent perspective, or cases of lexical ambiguity. Lexical ambiguity is present in both open-class words like "bat" (a baseball instrument

or a flying mammal) and closed-class quantifiers like "some" ("at least one and possibly all" or "at least one but not all"). Lexical ambiguity is toleratedbecause the meaning is usually easily disambiguated by context. When the word "bat" is uttered while watching a baseball game, the mammal usage of the word is very unlikely. The presence of this type of ambiguity can be viewed as an equilibrium between the two pragmatic pressures. If the meaning of a word can be disambiguated by the referential context, then it would violate the speaker's pressure to minimize effort by keeping track of two distinct words.

Indeed, recent work by Piantadosi, Tily, and Gibson (2011a) reveals systematicity in the presence of lexical ambiguity in language. They argue that ambiguity results from a speaker based pressure to broaden the meaning of a word to include multiple possible meanings. In particular, they suggest that this pressure should lead to a systematic relationship between the presence of ambiguity and the cost of a word. According to their argument, costly words (in terms of length, frequency, or any metric of cost) that are easily understood by context violate the speaker's principle to say no more than you must. Consequently, there should be a pressure for these meanings to get mapped on to a different, less costly word. This word may happen to already have a meaning associated with it, and so the result is multiple meanings being mapped to a single word. For example, in the case of the word "bat," a speaker could instead say "baseball bat." But, because this referent is easily disambiguated in context from the mammalian meaning, Horn's speaker principle leads to a pressure to use the shorter form. This leads to a testable prediction that shorter words should tend to be more ambiguous. Through corpus analyses, Piantadosi et al. (2011a) find this precise relationship between cost and ambiguity. They find a linear relationship between word length and ambiguity across English, Dutch and German: Shorter words are more likely to have multiple meanings.

An additional case of this lexical ambiguity is found in words that have very little context-independent meaning, known as indexicals or deictics (?, ?). These words get their meaning from the particular referential context of the utterance, and are therefore highly

ambiguous from a context-independent perspective. There are many types of indexicals that are present to varying degrees across languages. An example of a temporal indexical form is "tomorrow." The context-independent meaning of this word is something like "the day after the day this word is being uttered in." Critically, abstracted from any context, this word has little meaning; it is impossible to interpret without having knowledge about the day the word was uttered. This phenomenon is also present in person pronouns (e.g. "you" and "I") and spatial forms, like "here" and "there." As for lexical ambiguity, this type of ambiguity is a predicted equilibrium point from Horn's principles: If the hearer can recover the intended referent from context, the speaker would be saying more than is necessary by using an overly-specific referential term (e.g., "December 18th, 2014" vs."tomorrow"). Language structure reflects this pressure through lexicalized ambiguity in the form of indexicals.

Finally, the relationship between the meanings of different words can be seen as a consequence of pragmatic principles. A number of theorists have noted a bias against two words mapping onto the same meaning — that is, a bias against synonymy (Saussure, 1916, 1960; ?, ?; Horn, 1984; E. Clark, 1987; ?, ?). This bias is an equilibrium between Horn's speaker and hearer principles. Recall that the optimal language for a hearer is one in which each meaning maps to its own word — exactly a language biased against synonymy. It turns out that the speaker's pressure also biases against synonymy. The optimal language for the speaker is a language where a single word maps to all meanings. But, a case where multiple words map to a single meaning is also undesirable because the speaker must keep track of two words. So, for both the speaker and the hearer, there is pressure to avoid synonymy. Thus, when a listener hears a speaker use a second word for an existing meaning, the hearer infers that this could not be what the speaker intended because this would violate the speaker's principle. The result is an assumption that the second word maps to a different meaning. This pattern is reflected in language structure by a one-to-one pattern in the lexicon — that is, a structure in which each word maps to exactly one meaning and each meaning maps to exactly one word.

As one kind of evidence for this one-to-one structure in the lexicon, Horn (1984) points to a phenomenon called *blocking*. Blocking refers to cases in which an existing lexical form blocks the presence of a different, derived form with the same root. Consider the following examples:

(a) fury furious *furiosity

(b) *cury curious curiosity

In both (a) and (b), forms that would be expected, given the inflectional morphology in English, are not permitted. This is presumably because they would have the same meaning as the existing form because they have the same root. Examples such as this provide some evidence for a one-to-one structure in language, but a one-to-one structure is a particularly difficult linguistic regularity to test empirically. Nonetheless, it is an important regularity because it licenses certain inferences in interpreting the meaning of words. In particular, the cognitive representation of a one-to-one regularity has been posited as an explanation of children's bias to map a novel word onto a novel object (Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003).

Together these phenomena—semantic organization, ambiguity, and one-to-one structure—provide three cases in which equilibria that are predicted by theories of communication at the pragmatic timescale are reflected in the structure of the lexicon at the language evolution timescale. While this commonality does not entail causality, it is suggestive of a causal relationship between the two timescales. Next, we turn to accounts at both the pragmatic and language evolution timescale for our particular phenomenon of interest: variability in language length.

Accounts of language length variability

Language forms vary along many dimensions, but a salient dimension is length: words and entire utterances can have dramatically different phonetic lengths. At the pragmatic timescale, several theories of communication predict that longer expressions should be associated with less predictable or typical meanings than their shorter counter parts.

One such theory is Horn's theory of communication (1984). It is often possible in language to use two different utterances to refer to the same meaning (in truth functional terms), and often these utterances differ in length. Horn presents the following example:

(a) Lee stopped the car.

(b) Lee got the car to stop.

Both (a) and (b) have the same denotational meaning (the successful stopping of a car), but they differ in length ((b) has two extra words). Horn argues that this asymmetry leads to an inference on the part of the listener that the two differ in meaning. The logic of this inference is identical to the lexical structure case above. The listener hears a speaker use a more costly phrase to express a meaning that could have been expressed in a less costly way. The listener thus infers that this other meaning could not be what the speaker intended because this would violate the speaker's principle to say no more than is necessary. Horn adds an additional layer to this argument. He suggests that no only do these two forms differ in meaning, but that they map onto meanings in a systematic way. In particular, he argues that the longer form gets mapped on to the more marked meaning, while the shorter form refers to the unmarked meaning. The notion of 'markedness' is underspecified here, but an intuitive definition is related to complexity: more marked things are more conceptually complex, while less marked things are more conceptual simple. Thus, in the above example, (a) would refer to a simple, average case of car stopping, while (b) might refer to case where something complex or unusual happened, perhaps because Lee used the emergency brake.

The source of the particular mapping between forms of different lengths and meanings of different degrees of markedness is unclear. This is because, in principle, there are multiple equilibrium points in the mapping between form and meaning. Assuming a one-to-one constraint on the mapping, there are two possible equilibria: {short–simple, long–complex} or {short–complex, long–simple}. Both satisfy the constraint that each form gets mapped to a unique meaning. So how do speakers arrive at the {short–simple, long–complex} equilibrium? This is a difficult result to derive from models of pragmatic reasoning. ? (?) successfully derive this result

as a consequence of the fact that {short–simple, long–complex} is a more optimal mapping for the speaker (the indirect result of Zipf's Principle of Least Effort). Another possibility relies on iconicity: hearers have a cognitive bias to map more complex sounding forms to meanings that are similarly complex.

A second theory to predict that longer forms should be associated with less predictable meanings is uniform information density (UID). UID p

Bergen, Goodman, and Levy (2012) provide experimental evidence for this type of implicature in a communication game. In their task, partners were told that they were in an alien world with three objects and three possible utterances of different monetary costs. They operationalize the idea of markedness or complexity as frequency, such that participants were instructed that each of the three different objects had three different base rate frequencies associated with them. Participants' task was to communicate about one of the objects using one of the available utterances. If they successfully communicated, they received a reward. The results suggest that both the speaker and hearer expected costlier forms to refer to less frequent meanings. This study provides one piece of evidence that Horn's predicted equilibrium between word length and meaning emerges in coordination games.

At the timescale of language evolution, there is some indirect evidence that this same bias is present in the lexicon. These approaches use the linguistic context of a word as an indirect measure the complexity of meaning. The idea is that words that are highly predictable, given the linguistic context, have more complex meanings, while words that are less predictable given the linguistic context, have less complex meanings. Piantadosi, Tily, and Gibson (2011b) measured the relationship between the predictability of words in context and the length of words. Across 10 languages, these two measures were highly correlated: words that were longer were less predictable in their linguistic context on average. This result held true even controlling for the frequency of words. Additional evidence for this relationship comes from examining pairs of words that have very similar meaning, but differ in length (e.g. "exam" vs. "examination;"

Mahowald, Fedorenko, Piantadosi, & Gibson, 2012). Through corpus analyses, they find that the longer forms are used in less predicable linguistic contexts. They also find in a behavioral experiment that speakers are more likely to select the longer word in unsupportive contexts. This body of work points to a systematic relationship between word length and meaning, when complexity is operationalized as predictability in the linguistic context.

One proposed account of this variability is spoken frequency. ? (?) argued that words that were more frequent would tend to be shorter, as a consequence of satisfying the speaker's pressure to minimize effort. While frequency may be related to

A separate body of work also finds a relationship between length and aspects of meaning. [markedness]

Thus, at the pragmatic timescale, there is a well-motivated prediction that less predictable meanings should be described with longer utterances. At the language evolution timescale, there is some indirect evidence that longer words refer to less predictable meanings, but no work directly tests this prediction. This was the goal of the present studies. In the set of four studies described below, we test the hypothesis that longer words refer to less predictable or complex meanings—a complexity bias—and find evidence for this bias both experimentally with novel words and observationally in the lexicons of 80 natural languages.

Study 1: Complexity bias with artificial objects

As a first step in exploring a complexity bias, we manipulated the complexity of objects and asked participants to infer their names. Object complexity was manipulated by varying the number of primitive parts the objects were composed of. If participants have a complexity bias, we predicted they should be more likely to map a longer novel word onto an object composed of more parts, compared to an object with fewer parts. In Experiment 1a, we first conducted a norming study to verify our intuitions that the number of object parts correlated with explicit judgements of complexity. In Experiment 1b, we used these normed stimuli in a simple word mapping task,

revealing a robust complexity bias. Experiment 1c replicated Experiment 1b with randomly concatenated syllables.

Experiment 1a: Object complexity norms

Methods

Participants. In this and all subsequent experiments, participants were recruited on Amazon Mechanical Turk and received US $0.15-0.30 for their participation, depending on the length of the task. 60 participants completed this first experiment.

Across all experiments, some participants completed more than one experiment. The results presented here include the data from all participants, but all reported results remain reliable when excluding participants who completed more than one study. Participants were counted as a repeat participant if they completed a study using the same stimuli (e.g., completed both Experiment 1a and 1b with geons).

Stimuli. As object primitives, we used "geon" shapes which are argued to be primitives in the visual system under one theory of object recognition (Biederman, 1987). We created a set of 40 objects containing 1-5 geon primitives (Figure 1)[1].

Procedure. We presented participants 12 objects from the the full stimulus set one at a time. For each object, we asked "How complicated is this object?," and participants responded using a slider scale anchored at "simple" and "complicated." Each participant saw two objects from each complexity condition, and the first two objects were images of a ball and a motherboard to anchor participants on the scale.

Results and Discussion

Number of object parts was highly correlated with explicit complexity judgement ($r = .93$, $p < .0001$; $M = .47$, $SD = .18$). Figure 2a shows the mean complexity rating for each of the 40

---

[1] All stimuli, experiments, raw data and analysis code can be found at REMOVED FOR BLIND REVIEW.
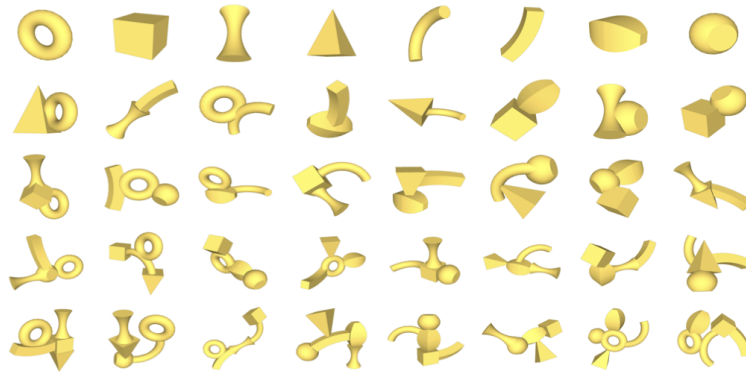
Figure 1. Stimuli in Experiment 1, where each row corresponds to a complexity condition. The complexity condition is determined by the number of "geon" parts the object contains (1-5).

objects as a function of their complexity condition. This result suggests that we can use manipulations of visual complexity as a proxy for manipulations of conceptual complexity.

Experiment 1b: Mapping task

Methods

   Participants. 750 participants completed the experiment.

   Stimuli. The referent stimuli were the set of 40 objects normed in Experiment 1a. The linguistic stimuli were novel words either 2 syllables (e.g., "bugorn") or 4 syllables (e.g., "tupabugorn") long. There were 8 items of each syllable length.

   Procedure. We presented participants with a novel word and two possible objects as referents, and asked them to select which object the word named ("Imagine you just heard someone say *bugorn*. Which object do you think *bugorn* refers to? Choose an object by clicking the button below it.").

   Within participants, we manipulated word length and the relative complexity of the referent
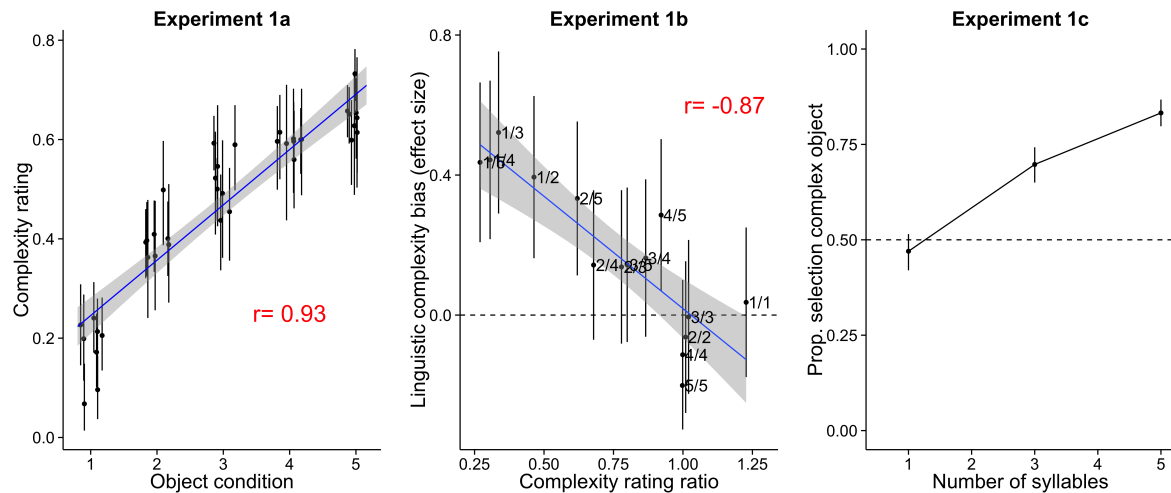
Figure 2. (a) The relationship between number of geons and complexity rating is plotted below. Each point corresponds to an object item (8 per condition). The x-coordinates have been jittered to avoid over-plotting. (b) Effect size (bias to select complex alternative in long vs. short word condition) as a function of the complexity rating ratio between the two object alternatives. Each point corresponds to an object condition. Conditions are labeled by the number of geons of the two alternatives. For example, the "1/5" condition corresponds to the condition in which one alternative contains 1 geon and the other contains 5 geons. (c) Proportion complex object selections as a function of the number of syllables in the target label. The dashed line reflects chance selection between the simple and complex alternatives. All errors bars reflect 95% confidence intervals, calculated via non-paramedic bootstrapping in 1a and 1c, and parametrically in 1b.

alternatives. We tested every unique combination of object complexities (1 vs. 2 geons, 1 vs. 3 geons, 1 vs. 4 geons, etc.), giving rise to 15 conditions in total. Each participant completed 4 short and 4 long trials in a random order, where each word was randomly associated with one of the complexity conditions. No participant saw the same complexity condition twice and no word or object was repeated across trials.

Results and Discussion

Across conditions, the more complex object was more likely to be judged the referent of the longer word. For each object condition (e.g., 1 vs. 2 geons), we calculated the effect size for participants' complexity bias—the degree to which the complex object was more likely to be chosen as the referent of a long word, compared to the short word. Effect sizes were calculated using the log odds ratio (Sánchez-Meca, Marín-Martínez, & Chacón-Moscoso, 2003). Effect size was highly correlated with the ratio of object complexities: The greater the mismatch in object complexity, the more the longer word was paired with the more complex object ($r = -.87$, $p < .0001$).

This experiment provides initial evidence for a complexity bias in the lexicon: Given an artificial word and two objects of differing visual complexity, participants are more likely to map a longer word onto a more complex referent, relative to a shorter word.

Experiment 1c: Control mapping task

One limitation of Experiment 1b is that it uses a small set of words as the linguistic stimuli (8 short and 8 long), making it possible that idiosyncratic properties of the words could be driving the complexity bias. In Experiment 1c we sought to test this possibility by using words composed from randomly concatenated syllables rather than items selected from a small list of words.

Methods

Participants. 200 participants completed the experiment.

Stimuli. The referent stimuli were the geon objects composed of either 1 or 5 geons. The novel words were created by randomly concatenating 2, 4, or 6 consonant-vowel syllables (e.g., "nur," "nobimup," "gugotobanid"). The last syllable of all words ended in a consonant.

Procedure. Participants completed six forced-choice trials identical to Experiment 1b. We tested only the "1/5" complexity condition (1-geon object vs. 5-geon object). Word length was manipulated within-participant, such that each participant completed 2 trials for each of the three possible word lengths (2, 4, or 6 syllables).

Results and Discussion

Replicating the "1/5" condition in Experiment 1b, we found that participants were more likely to select a five geon object compared to a single geon object as the number of syllables in the word increased ($\beta = -.44$, $p < .0001$). This suggests that the complexity bias observed in Experiment 1b is unlikely to be due to the particular set of words we selected.

Study 2: Complexity bias with novel real objects

Study 1 provides evidence for a complexity bias using artificial objects. The complexity manipulation in these experiments was highly-transparent, however, making it possible that task demands influenced the effect. We next asked whether this bias extended to more naturalistic objects, where the variability in complexity might be less obvious to participants. In Study 2, we conducted the same set of 3 experiments as in Study 1 using a sample of real objects without canonical labels. We find that the complexity bias observed with artificial geon objects in Study 1 extends to naturalistic objects. In Study 2d, we also find the same bias in language production.

Experiment 2a: Object complexity norms

Methods

Participants. We recruited two samples of 60 participants to complete Experiment 2a.

Figure 3. Study 2 stimuli: naturalistic objects without canonical labels. Each row corresponds to a quintile determined by the explicit complexity judgements obtained in Experiment 2a (top: least complex; bottom: most complex).

Stimuli. We collected a set of 60 objects that were real objects but that did not have canonical labels associated with them (Figure 3).

Procedure. The procedure was identical to Experiment 1a.

Results and Discussion

Complexity judgements were highly reliable across two independent samples ($r = .93, p < .0001; M_1 = .49, SD_1 = .18, M_2 = .44, SD_2 = .18$). Figure 4a shows the relationship between the complexity judgment for each item across the two samples of participants.

Experiment 2b: Mapping task

Methods

Participants. 1500 participants completed the experiment.

Stimuli. The linguistic stimuli were identical to Experiment 1b. The object stimuli were the 60 naturalistic objects normed in Experiment 2a. Five complexity conditions were determined by
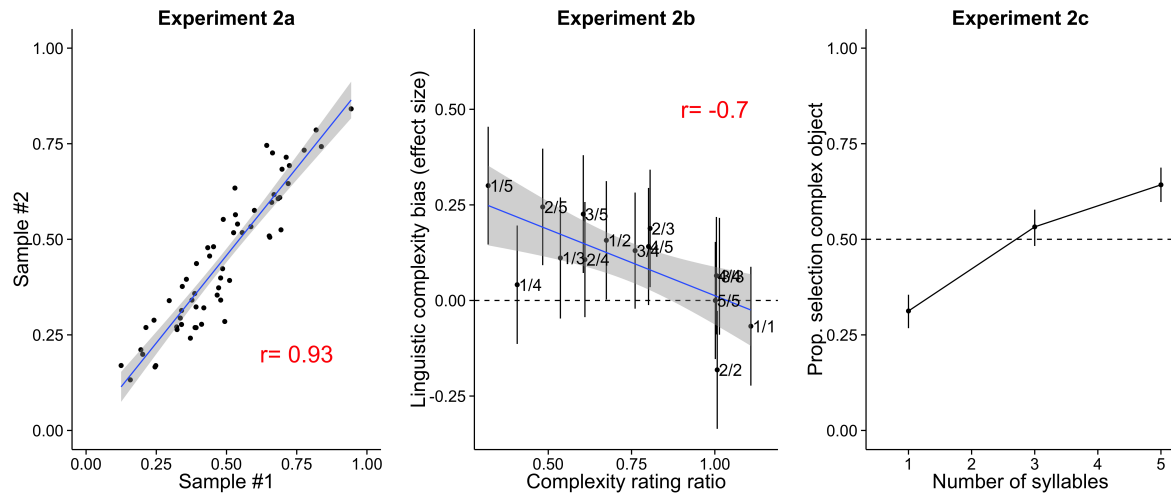
Figure 4. (a) The correlation between the two samples of complexity norms. Each point corresponds to an object ($n = 60$). (b) Effect size (bias to select complex alternative in long vs. short word condition) as a function of the complexity rating ratio between the two object alternatives. Each point corresponds to an object condition. Conditions are labeled by the complexity norm quintile of the two alternatives. (c) The proportion of complex object selections as a function of number of syllables. The dashed line reflects chance selection between the simple and complex alternatives. All errors bars reflect 95% confidence intervals, calculated via non-paramedic bootstrapping in 2a and 2c, and parametrically in 2b.

dividing the objects into quintiles based on the norms.

Procedure. The procedure was identical to Experiment 1b, except for the use of naturalistic rather than the artificial geon objects.

Results and Discussion

As with the artificial objects, effect size was negatively correlated with the complexity rating ratio between the referent alternatives ($r = .70, p < .005$; Fig. 4b). This suggests that the

complexity bias observed with artificial objects extends to more naturalistic objects, consistent with the proposal that this a complexity bias is a characteristic on natural language.

The effect size in Experiment 2b is smaller than in Experiment 1b, however. This may be due to the fact that some of the effect in Experiment 1c was due to task demands associated with the transparent complexity manipulation. Nonetheless, Experiment 2b reveals a robust complexity bias with naturalistic objects.

Experiment 2c: Control mapping task

Methods

Participants. 200 participants completed the experiment.

Stimuli. The objects were 12 objects from the first and fifth quintile of complexity norms. The linguistic stimuli were constructed as in Experiment 1c.

Procedure. The procedure was identical to 1c, except for the different object stimuli.

Results and Discussion

Participants were more likely to select an object from the fifth quintile as opposed to the first quintile when the novel word contained more syllables ($\beta = -.34, p < .0001$; Fig. 4c). This pattern replicates the complexity bias seen in 2b with randomly concatenated syllables.

In the present experiment, participants were overall less likely to select the complex object, compared to the same experiment with artificial objects (Experiment 1c; STATS). One possible reason for this difference may be the fact that some of the simple artificial objects in Experiment 1c may be associated with canonical labels. For example, the sphere single-geon object may have evoked the label "ball." If true, this might lead participants to appeal to mutual exclusivity in their object selections by selecting an object they do not already have a name for—in this case, the more complex object (Markman & Wachtel, 1988). Independent of this main effect, however, we replicate the complexity bias with random syllables in both studies.

Experiment 2d: Production task

Thus far we have observed the complexity bias only in comprehension tasks. In the next experiment, we ask whether the bias extends to language production by presenting participants with an object and asking them to produce a novel label to refer to it. [not sure how to motivate the extension to production]

Methods

Participants. Fifty-nine participants completed the experiment.

Stimuli. The object were drawn from the set of 60 naturalistic objects used in Experiments 2a-c.

Procedure. In each trial, we presented participants with a single object and asked them to generate a novel single-word label to refer to it. The instructions read: "What do you think this object is called? For example, someone might call it a 'tupa' or a 'pakuwugnum.' In the box below, please make up your own name for the object. Your name should only be one word. It should not be a real English word." Each participant completed 10 trials, with five objects from the bottom and top complexity norm quantiles each. Order of objects was randomized.

Results and Discussion

There were 26 productions (4%) that included more than one word. These productions were excluded. Length was measured in terms of log number of characters.

Participants produced novel coinages that were longer for the top quartile of objects ($M = 1.94$, $SD = 0.18$) compared to the bottom quartile ($M = 1.85$, $SD = 0.17$; $t(57) = 3.92$, $p < .001$). We also analyzed length as a function of the complexity norms for each object. Length of production was correlated with the complexity norms: Longer labels were coined for objects that were rated as more complex ($r = .17, p < .0001$). This suggests that the complexity bias is present in both language comprehension and production.

Study 3: Complexity as a cognitive construct

Experiment 3a

artificial object

Methods

Procedure.

Participants.

Stimuli.

Procedure.

Results

Experiment 3b

real objects

Methods

Procedure.

Participants.

Stimuli.

Procedure.

Results

Study 4: Complexity bias in natural language

Experiment 4a

English

Methods

Procedure.

Participants.

Stimuli.

Procedure.

Results

Experiment 4b

google xling

Methods

Procedure.

Participants.

Stimuli.

Procedure.

Results

General Discussion

alternative accounts of complexity bias * iconicity * naming hypothesis

need more direct evidence for a causal link

what is complexity??

References

Baddeley, R., & Attewell, D. (2009). The relationship between language and the environment information theory shows why we have only three lightness terms. *Psychological Science*, *20*(9), 1100–1107.

Bergen, L., Goodman, N. D., & Levy, R. (2012). Thats what she (could have) said: How alternative utterances affect language use. In *Proceedings of the thirty-fourth annual conference of the cognitive science society.*

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition. Hillsdale, NJ: Erlbaum.*

Clark, H. H. (1996). *Using language*. Cambridge university press.

Grice, H. (1975). Logic and conversation. *1975*, 41–58.

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., & Regier, T. (2012). Kinship categories across languages reflect general communicative principles. *Science*, *336*(6084), 1049–1054.

Mahowald, K., Fedorenko, E., Piantadosi, S., & Gibson, E. (2012). Info/information theory: speakers actively choose shorter words in predictable contexts. *Cognition*, *126*, 313–318.

Markman, E., & Wachtel, G. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, *20*(2), 121–157.

Markman, E., Wasow, J., & Hansen, M. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275.

Peirce, C. (1931). *The collected papers of charles s. peirce* (P. W. C. Hartshorne & A. Burks, Eds.). Cambridge: Harvard University Press.

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

Piantadosi, S., Tily, H., & Gibson, E. (2011b). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529.

Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, *104*(4), 1436–1441.

Regier, T., Kemp, C., & Kay, P. (2014). Word meanings across languages support efficient communication.

Sánchez-Meca, J., Marín-Martínez, F., & Chacón-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological methods*, *8*(4), 448.

Saussure, F. (1916, 1960).

In *Course in General Linguistics*. London: Peter Owen.

Schmidtke, D. S., Conrad, M., & Jacobs, A. M. (2014). Phonological iconicity. *Frontiers in psychology*, *5*.

Xu, Y., & Regier, T. (2014). Numeral systems across languages support efficient communication: From approximate numerosity to recursion. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*.

Zipf, G. (1936). *The psychobiology of language*. Routledge, London.

Appendix A: Materials