CHAPTER

6

Does Vocabulary Help Structure the Mind?

GARY LUPYAN
UNIVERSITY OF WISCONSIN-MADISON
DEPARTMENT OF PSYCHOLOGY, UNIVERSITY OF WISCONSIN-MADISON, 1202 W JOHNSON STREET, MADISON,WI 53706 USA, LUPYAN@WISC.EDU

MARTIN ZETTERSTEN
UNIVERSITY OF WISCONSIN-MADISON

LANGUAGE AND THOUGHT: HOW ARE THEY RELATED?

The idea that language shapes thinking seemed plausible when scientists were in the dark about how thinking works (Pinker, 1994, p. 58)

Instead of language merely reflecting the cognitive development which permits and constrains its acquisition, language is being thought of as potentially catalytic and transformative of cognition (Bowerman & Levinson, 2001, p. 13)

Does language reflect the categories of our mind or does it help create them? For centuries, people have been asking versions of this question (see Leavitt, 2011; Lee, 1996 for historical reviews). In the last several decades, this question has become the subject of increased empirical investigation. And yet, rather than moving toward consensus, the question of whether human cognition is transformed by language remains as contentious as ever (Bloom, 2002; Boroditsky, 2010; Carruthers, 2002; Gleitman & Papafragou, 2005; Lupyan, 2012a, 2016; Malt & Wolff, 2010; McWhorter, 2014; Pinker, 1994; Wolff & Holmes, 2011). Why?

We think a root cause of its contentious nature lies in two widespread assumptions: (1) that human concepts reflect objective reality and (2) that learning a word is simply learning a mapping between this objective reality and a sequence of sounds (or visual gestures in the case of signed languages). In the first section we review these assumptions and relate them to the question of linguistic influences on cognition. We next describe several mechanisms by which the words of a language can help structure knowledge and navigate cognitive problems. We will argue that when we learn a word, we do not simply map its meaning onto a pre-existing concept; instead, the learning process contributes to the formation of the conceptual category denoted by the word. We then review the idea of *nameability* – the ease with which an entity can be named – and describe ongoing empirical work investigating how differences in nameability relate to performance on a variety of categorization and reasoning tasks.

Cognitive Priority and Linguistic Priority

It is because thought and language seem so closely linked that language is so often used as a window to thought (Pinker, 2007). Acknowledging a link between language and thought raises the question of priority: "Which comes first? Thought or Language?" (Fodor, 2001). For Fodor and others working within the classical cognitivist tradition (e.g., Fodor, 1975; Mahon & Caramazza, 2009; Pinker, 1994; Snedeker & Gleitman, 2004), the answer is clear. Thought comes first. Language is its expression. A common argument for this position (sometimes referred to as the *cognitive priority hypothesis*) is that it is only possible to learn a word for a concept you already have (see Bowerman, 2000 for discussion and critique). This position is sometimes stated explicitly: "The meanings to be communicated, and their systematic mapping onto linguistic expressions, arise independently of exposure to any language" (Gleitman & Fisher, 2005, p. 133). More commonly, however, the assumption is an implied one. For example, in his polemical essay "The great Eskimo Vocabulary hoax," Pullum (1989) ridicules the claim that languages differ in how they lexicalize snow by arguing that even if such differences in lexicalization were true, they would not be interesting:

> [E]ven if there were a large number of roots for different snow types in some Arctic language, this would not, objectively, be intellectually interesting; it would be a most mundane and unremarkable fact. [H]orse breeders have various names for breeds, sizes, and ages of horses; botanists have names for leaf shapes; interior decorators have names for shades of mauve … If these obvious truths of specialization are supposed to be interesting facts about language, thought, and culture, then I'm sorry, but include me out.
>
> (pp. 278–279)

Pullum's tacit assumption that words map onto pre-existing categories leads him to conclude that it does not – and indeed *cannot* – matter if a distinction is lexically marked. The assumption seems to be that when there is a need to categorize something (horse breeds, shades of mauve, etc.), individuals will learn the relevant categories, and then may go on to develop a vocabulary to facilitate communication about those distinctions. The possibility that words can help people learn the categories in the first place is never considered (Lupyan, 2012b for discussion).[1]

If meanings indeed come first, where do they come from? Fodorian nativism (Fodor, 1975) aside, we can identify two sources of this knowledge. Some meanings (e.g., dog, water, spoon) come from identifying the joints of nature. Once identified, some of these joints are mapped onto words (see Lupyan, 2016; Lupyan & Lewis, 2017 for discussion). Of course, linguistic meanings are not limited to concrete categories. Instead, much – and on some analyses, *most* – of what we talk about is quite abstract (Lupyan & Winter, 2018). For abstract categories such as containment, causality, and time, researchers have often posited innate (or "core") knowledge as the source of meanings that words map onto (e.g., Spelke & Kinzler, 2007).

The opposing view, sometimes referred to as the *linguistic priority hypothesis*, is that our conceptual content and structure draws on – or even requires – experience with natural language. On this view, we have the particular concepts we do, not because they reflect objective categories in the world or because we are endowed with them by our biology, but because these categories have been *constructed* by humans and are transmitted via natural language. Consider the meanings conveyed by words like "game," "furniture," and "Sunday." Clearly, these do not reflect objective joints of nature. Nor do they plausibly reflect innate content. Would a child never exposed to these linguistic terms still go on to have these same concepts? Or does learning these categories depend, in some way, on learning the corresponding words? The most famous proponent of the view that our conceptual structure importantly depends on natural language is Benjamin Lee Whorf:

> The categories and types that we isolate from the world … we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds – and this means largely by the linguistic systems in our minds….
>
> (Whorf, 1940/1956)

In 2019 alone, there were over 600 references to the *Whorf hypothesis* (see also Wolff & Holmes, 2011). But the idea that the close link between language and thought exists because our thoughts in part *derive* from natural language precedes Whorf, and we think a brief historical review is illuminating.

Long before Whorf, John Locke argued that it is precisely *because* our thoughts are so affected by natural language that we must guard against the vicissitudes of language lest it "cast a mist before our eyes and impose upon our understandings" (Locke, 1849, p. 356). Hardly a relativist, Locke nevertheless recognized that even "a moderate skill in different Languages" reveals that "though they have Words, which in Translations and Dictionaries, are supposed to answer one another; [there] is scarce one of ten, amongst the names of complex *Ideas* … that stand for the same precise *Idea*" (Locke, 1849, p. 315). Different languages seem to identify different joints of nature.

Arguing for a much more causal role of language, the philosopher, diplomat, and early linguist Wilhelm von Humboldt rejected the idea that words simply reflect pre-existing categories, writing in 1816 that a word "is so little the sign of a concept that the concept cannot even come into being, much less be fixed, without it" (1816; as cited by Leavitt, 2011, p. 93). The categories imposed by language, argued Humboldt, are "not so much the means to represent truth once established, but rather the means to discover truth previously unknown," and therefore the diversity of languages "is not one of sounds and signs, but a diversity of world views themselves" (Leavitt, 2011, p. 93).

Nearly a century later, William James echoed this idea in his discussion of how one might go about learning to distinguish a claret from a burgundy. At first, wrote James, one might associate the names of these wines with various details of the experience, but "after a while the tables and other parts of the setting, besides the name, grow so multifarious as not to come up distinctly into consciousness, but [the] adhesion of each wine with its own *name* becomes more and more inveterate, at last each flavor suggests [its] own name and nothing else." More than simply *referring* to the pre-existing categories, the names that "differ far more than the flavors … help to stretch these latter farther apart" (James, 1890, p. 511).

Two Arguments Against the Cognitive Priority Hypothesis

The cognitive priority view faces two serious problems. The first is accounting for the cross-linguistic diversity of vocabularies. If words map onto pre-existing concepts, why are there such large differences between the vocabularies of different languages? Some of these differences can be attributed to differences in culture. The development of specific artifacts and institutions would bring with them vocabularies that would be unnecessary in a culture lacking those artifacts and institutions. However, cross-linguistic differences in vocabulary touch on *all* aspects of experience, including universal human experiences such as eating, drinking, carrying, and having sex (Evans & Levinson, 2009; Malt et al., 2015; Wierzbicka, 2009). The diversity revealed by cross-linguistic analysis of semantics is often masked by a trick of typography, as when psychologists and philosophers use capitalized words to stand in for non-linguistic concepts: GIFT the concept vs. "gift" the word. This typographical convention assumes the existence of the posited concept independent of any linguistic experience, assuming at the outset that the cognitive priority view is true.

The second problem is the problem of origin. If concepts come first, where do they come from? For some concrete concepts like TREE, a reasonable answer is that they come from analyzing nature at its joints. But even with such a seemingly straightforward category we quickly run into trouble. What makes a tree? What makes it different from a bush or shrub? The National Park Service offers the following definition: "Generally, trees are over 20 feet tall and have trunks more than 2 inches in diameter at 4.5 feet above the ground" (USNPS, 2018). This hardly sounds like an obvious joint of nature and raises doubts as to whether someone who was never exposed to the word "tree" as used by English speakers would have the very same semantic representation of the concept TREE as someone whose knowledge of trees includes the experience with learning and using the word "tree."

The origin problem becomes more acute for abstract meanings. Though there might be fuzziness around the boundaries, someone who has never encountered the word "tree" would presumably come to have *some* concept of trees based on perceptual experiences alone. But this argument is difficult to maintain for abstract meanings (Lupyan & Winter, 2018). In learning English, we learn words such as "exciting," "pathetic," "miracle," "lucky," "barely," "opinion," "fun," "somewhere," and "meanwhile." An English learner who already speaks a language with close translations of these terms may well map these terms (with varying success) to corresponding terms in a language they already speak. But what about children learning English as a first language? What prior meanings would these words map onto?

Can we solve this problem by appealing to innate knowledge? We think not. Even if infants come into the world with core knowledge in broad domains such as agency, causality, space, number, emotions, and an innate motivation to attend to these domains, it is still a far leap to go from such general knowledge to ~~specific~~ meanings that can be mapped onto the sort of abstract words mentioned above. It is difficult to escape the conclusion that the categories picked out by such abstract words may depend – in a rather strong way – on experience with language itself.

A Challenge to the Linguistic Priority Hypothesis

If learning concepts such as "meanwhile," "someplace," and "fun" require linguistic experience, then how did these words come to be in the first place? This is perhaps the chief critique of the idea that language helps structure our concepts (Bloom, 2002). Learning these words does require certain cognitive and perceptual prerequisites. It is difficult to see how one could learn word meanings like "somewhere" with no prelinguistic notion of space; "meanwhile" with no prelinguistic notion of time; "nostalgic" with no ability to represent emotional states. Indeed, proponents of linguistic priority do not typically claim that children enter the world as blank slates depending on language for all of their mental content. For example, William James argued for the importance of verbal labels in learning perceptual categories, writing that the difference between experiences is "made to seem more substantial by recognizing the terms." At the same time, James acknowledged that the labels are unlikely to do much if a person could not detect any differences between the experiences in the first place: "it is difficult to show coercively that naming … hardly distinguishable [experiences] is essential to their being felt as different at first" (James, 1890, p. 512). Likewise, the starting point for Whorf was not a blank conceptual slate, but rather the aforementioned "kaleidoscopic flux." In less metaphoric terms: innate mental content and perceptual input from the world *under-determine* conceptual structure. What language does, on the linguistic priority view, is to help create order by "proposing" an organizational scheme to the flux. The answer to how words enter a language if they do not map onto pre-existing conceptual structure is that words *create* their own structure. The challenge, of course, is to understand how and when this happens.

Vocabulary as a Joint of Nature

How can words help structure the mind? In a poignant analogy, Clark (1998) compares the relationship between words and concepts to the relationship between trees and the soil in which they grow. "If a tree is seen growing on an island, which do you suppose came first?" asks Clark. It is natural to assume, he acknowledges, that the island "provided the fertile soil in which a lucky seed came to rest" (Clark, 1998, p. 176). But "a revealing exception to this general rule [are mangroves]." Mangrove seeds become trapped and send aerial roots that catch floating soil and various debris, which over time form a small island that traps progressively more soil:

> Throughout this process, and despite our prior intuitions, it is the land which is progressively built by the trees … Something like the Mangrove effect, I suspect, is operative in some species of human thought. It is natural to suppose that words are always rooted in the fertile soil of pre-existing thoughts. But sometimes, at least, the influence seems to run in the other direction.
>
> (p. 176)

We think the influence runs from words to thoughts more frequently than often supposed. We will argue that not only are patterns of lexicalization cognitively relevant, but that even small differences in nameability – the ease with which something can be named in a given language – have surprisingly large cognitive consequences. Such effects are expected on linguistic priority accounts, but difficult to reconcile with a strong cognitive priority view. Before proceeding to the data, let us consider several reasons why it might matter whether a language lexicalizes a certain distinction using a frequent and compact verbal expression.

A Named Distinction is a Marked Distinction

While the reasons *for* lexicalizing a certain distinction are many, and certainly include cultural specialization of the type discussed by Pullum (1989), it is the *consequences* of lexicalization for language learners and users that are of psychological interest. Take color words as an example. Languages vary in the number of lexicalized color terms (Kay, Berlin, Maffi, Merrifield, & Cook, 2011). Although all languages allow us to describe visual properties, some do not have words that pick out differences in hue in particular (Wierzbicka, 2006). In such a language, the question "what color is this?" is not only difficult to answer, but difficult to even pose. Others, like English and Russian, have many color words. These differences in vocabulary stem from various historical factors such as dye production and mass manufacture of objects that can vary arbitrarily in color (the informativeness of phrases like "grab me the blue one" hinges on there being objects that vary in color, but are otherwise functionally identical; something not generally found in the natural world) (Kay & Maffi, 1999). But the question of *why* languages have the number of color words they do is distinct from the question of *what are the consequences* of learning and using a language with a certain color vocabulary (Forder & Lupyan, 2019).

One consequence of English lexicalizing certain basic colors (red, green, blue, etc.) is that all speakers will learn these distinctions in the course of learning English. Some English speakers will go on to learn many more color words beyond these basic ones. But *all* English speakers (even those who are congenitally blind) will, beginning at a young age, learn at least the basic color words, because these words are a core part of modern English. Although all languages can develop color words if needed, this process is a gradual one, unfolding over generations. Someone who is in a situation where it would be useful to refer to a specific hue but who learned a language that lacks color words is out of luck (just as English speakers are out of luck when trying to accurately name an odor; Majid & Burenhult, 2014).

The same reasoning applies to words for numbers, shapes, spatial relations, and thousands of other words, each of which has been shaped by many generations of cultural evolution. It may be within some people's capacity to invent these word meanings on the spot, but with these words in the language already, learners have a far simpler job – to learn the word meanings already used by the community rather than to discover them on their own.[2] Learning the meaning of a word *necessarily* requires learning to distinguish the set of objects/relations/abstract ideas/etc. to which the word applies from the ones that it does not. Although there is nothing preventing a speaker from learning a non-lexicalized distinction, lexicalizing a distinction *ensures* that it is learned by all speakers of the language.

Names Discretize the Continuous

The world of perception and action is analogue. Objects vary continuously in size, color, weight, and position. Object categories, while often seemingly all-or-none, tend to come in degrees. Whatever genetic markers may exist to unambiguously mark that a dog is really a dog, the *perceptual* fact is that some dogs are "doggier" than others. In contrast, the world of language is a world of discrete categories. It is categories all the way down. "Animal" is a category, but so is "dog" and "beagle" (albeit with a progressively narrower extension).[3]

We can and do talk about degrees; we can say "a beagle is *doggier* than a bulldog." But such expressions are still categorical. "Doggier" denotes a positive direction on a not-a-dog-to-dog dimension without specifying the precise value on that dimension. Expressions like "it is green" are clearly categorical, but so are hedges like "it is sort of green." The latter refers to the category of colors that can be plausibly, but not typically, described as green (though in practice such expressions may be more informative about the state of the speaker's knowledge than about the colors in question).

A consequence of this linguistic discretization is that words create equivalences that otherwise may not exist. In referring to a class of spatial relations by the word "on," English *creates* an equivalence class between otherwise rather dissimilar entities: a plate *on* a table, a painting *on* a wall, a handle *on* a door, etc. (Bowerman & Choi, 2001). To reiterate: it is not that representing the relationship between a painting and a wall, or between a handle and a door requires learning the word "on." There may well be other, equally good nonlinguistic ways of highlighting the relationships. The point is that English speakers necessarily learn that a painting and a wall are related in some similar way to a handle and a door; not learning this relationship would mean that they cannot use the word "on" properly.

Names, Dimensionality Reduction, and Compositionality

The meanings of many words can be decomposed into simpler units (if this were not possible, writing dictionaries would be an even more daunting task). Even so, there is something unitary, something chunky, about a meaning conveyed by a word. We can decompose 100 into 10 × 10 just like we can decompose 10,000 into 10 × 1,000, yet to an English speaker, *hundred* feels more unitary than *ten thousand*.[4] The word *hundred* seemingly compresses the more complex meaning of "ten tens" into a single chunk. This point is well made by Levinson:

> We don't have to think about a hundred as "ten tens" when doing mental arithmetic, or aunt as "mother's sister, or father's sister, or father's brother's wife, or mother's brother's wife" when greeting Aunt

> Mathilda … Composing complex concepts gives enormous power to our mental computations, and most of those complex concepts are inherited from the language we happen to speak.
>
> (Levinson, 2003)

As speakers of our native tongue, we learn (i.e., culturally inherit) thousands of "chunks" such as "*hundred*." Might the availability of such chunks facilitate certain cognitive operations?

As an initial test of this idea, we conducted a category-learning experiment in which participants had to learn one of two nearly identical category structures (Figure 6.1) (see also Zettersten & Lupyan, 2019). On each trial, participants saw a category exemplar and had to assign it to one of two categories, at which point they received accuracy feedback. On standard accounts of categorization (e.g., Ashby & Maddox, 2011), learning these categories involves integrating information across two dimensions: the height of the horizontal line along the y-axis and the position of the vertical line along the x-axis. The category structure is thus thought to be determined by these basic perceptual dimensions, which have little to do with language. But there is an alternate way of representing this category space. In Figure 6.1A, many of the shapes can be named. Recognizing that the categories comprise shapes that can be named – roughly, as *T*s and *L*s – allows the learner to collapse the two-dimensional space into a simpler one-dimensional one.[5] If we simply rotate the stimuli 180° (Figure 6.1B), we leave all perceptual features (and logical structure) unchanged, but make it less likely that people recognize any of the shapes as belonging to the T and L "chunks," leaving the problem space two dimensional.

We recruited 70 people to learn an "information integration" as shown in Figure 6.1A or Figure 6.1B. Learners saw each shape individually and were asked to classify it as a member of category 1 or category 2. They then received immediate feedback on whether their response was correct. Participants were not told that they should try to name the shapes or that such an approach is useful. Each learner completed 60 trials (each shape was shown twice). Participants in the condition with harder-to-name stimuli (Figure 6.1A) performed much more poorly than participants in the condition in which some of the stimuli resembled Ts and Ls (Figure 6.1B)

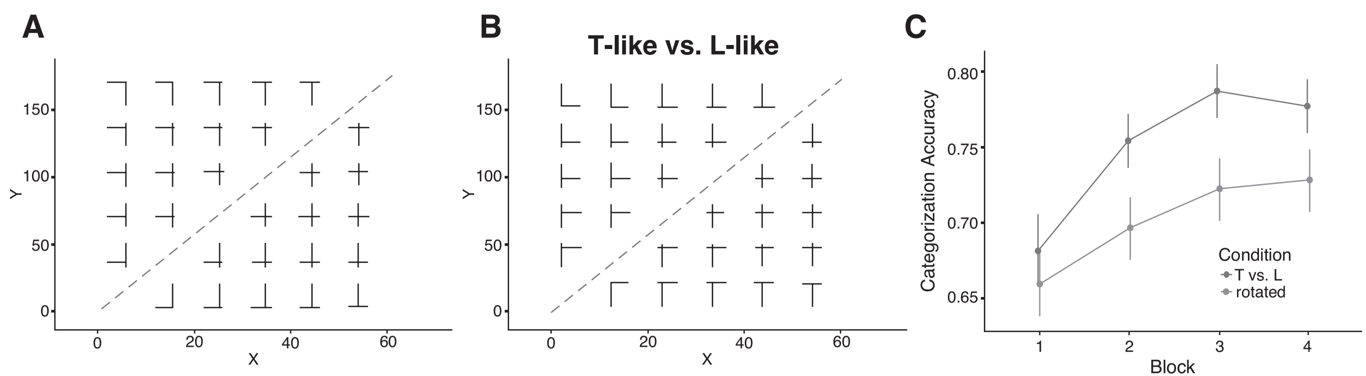

Figure 6.1 (A) A–B Two classic "information-integration" category structures of putatively identical complexity. The category boundary is marked by the dashed line. (C) Participants learn (B) better than (A) because names (T-like vs. L-like) help reduce the dimensionality of (B), but not (A).

enabling participants to represent the stimulus space in terms of a T-like to L-like dimension ($M_{harder\text{-}to\text{-}name}$ = 70%; $M_{easier\text{-}to\text{-}name}$ = 75%, z = 2.32, p = .02; Figure 6.1C). These results hint at how a subtle visual manipulation that makes visual stimuli easier to name can impact a seemingly straightforward category learning task. Data, analyses, and stimuli for many of the experiments and results described in this chapter are available at https://github.com/mzettersten/vocab-mind-2020.

WHEN NAMES ORGANIZE THE FLUX

To *really* know if words help structure our minds requires manipulating people's knowledge of language while holding all else equal – an experiment impossible on both practical and ethical grounds. What we can do, however, is measure and manipulate linguistic factors and examine how these relate to putatively non-linguistic cognitive behaviors. Finding a correlation between a linguistic factor and performance on some cognitive task suggests that the two may be related. Finding that manipulating the linguistic variable selectively affects the putatively nonlinguistic one suggests that language may be a driving factor. The T/L categorization study described above hints at how we can use subtle manipulations to examine influences of language on category learning. Further examples that extend this logic to other domains can be found in Lupyan (2012b, 2016) and Lupyan and Bergen (2016). In this section, we present data – much of it preliminary – testing the hypothesis that *nameability* – the ease with which people can name a certain object or relation – affects people's ability to categorize, reason, and make inferences about those objects and relations. To do this, we first quantify and manipulate nameability, and then measure the consequences of these manipulations.

Any empirical investigation of nameability runs into an immediate challenge. Suppose it is discovered that a less nameable distinction leads to poorer performance when, e.g., learning a new category that relies on making this distinction. Does this mean that the ability to name helps people to categorize or that certain distinctions are inherently difficult, and therefore both difficult to learn *and* less likely to be named? On their own, none of the results we present here can unambiguously distinguish between these two possibilities. In sum, however, we believe the results present a compelling case for the causal power of verbal labels to influence learning and reasoning in adults, and hint at even more significant effects in the development of children's conceptual knowledge. We present these results not as a conclusive proof of the linguistic priority hypothesis or a bullet-proof case to convince the Whorfian skeptic. Instead, we hope to give our reader reason to doubt a strong form of the cognitive priority hypothesis and to provide novel evidence for the influence of linguistic factors on human cognition.

Nameability Defined

We use the term nameability to refer to the ease with which people can name *X* where *X* can stand in for anything: an object, a relation, or an abstract idea. Something is highly nameable if it evokes the same verbal response on various occasions. Nameability is related to the more familiar and well-studied construct of name-agreement – the extent to which different people agree on what *X* should be called. It turns out that disagreement *between* people on what something is called is highly correlated with the time it takes an *individual* to name that thing (Lachman, 1973). That name-agreement, defined at the level of a *group*, systematically predicts performance of *individuals* is not a logical necessity, but it enables us to use various measures of name-agreement in the population as a proxy for what is happening in an individual mind.

To obtain agreement-based measures of nameability, multiple participants are presented with some stimuli and asked to name them. Static images are most often used, but the same procedure can, in principle, be used with any stimuli. There are various ways of computing agreement-based nameability. Most measures focus on naming, for example, by computing what percentage of participants gives the modal response (Brandimonte, Hitch, & Bishop, 1992; Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010; Perry & Lupyan, 2016). Another measure is the entropy of the naming responses (Brodeur et al., 2010; Snodgrass & Vanderwart, 1980), defined as:

$$H = \sum_{i=1}^{k} p_i \log_2 \left(1 / p_i \right)$$

where k is the number of different names given to an item and pi is the proportion of subjects giving each name. In this context, entropy measures how predictable the naming response of one person is if you know the responses to the same stimulus made by other people. If participants all give the same verbal response, the verbal responses are perfectly predictive of each other and the entropy is zero. As the variability in participants' responses increases, they become harder to predict from one another and entropy increases. Higher entropy therefore indicates lower nameability. A similar, measure that focuses on the diversity of responses rather than their predictability per se is Simpson diversity (Simpson, 1949; for recent application to nameability, see Majid et al., 2018; Zettersten & Lupyan, 2019).

Another way of measuring nameability is *naming divergence*, which captures the inconsistency in participants' naming responses:

$$\text{naming divergence} = \frac{\text{number of unique words}}{\text{number of total words}}$$

For instance, if six participants respond to a color patch with the word "purple" and four others respond with "mauve," there are two unique words and 10 total words in the responses yielding a naming divergence of 0.2. If, however, six participants respond with "purple" and four others each respond with a different word ("lavender," "periwinkle," "magenta," and "violet"), we have five unique responses yielding a naming divergence of 0.5.

These examples highlight the limitations of computing naming consistency based on the percentage of participants who give the modal response. In both of the examples above, the dominant name makes up 60% of the responses – meaning that a name agreement measure based on modal responses treats these two cases as equivalent. Both the entropy-based measure and the naming divergence measure capture the fact that there is more consistency in the case where the remaining 40% of participants use the same term than when they use different terms. In the following analyses, we will use the naming divergence measure; we obtain similar results using the entropy-based measure of name agreement.

Name agreement is one dimension along which we can quantify nameability. Another dimension is the *complexity* of the verbal response. ~~All~~

~~else being equal, s~~omething ~~with~~ a longer naming response (measured in number of words or number of clauses) is less nameable than something ~~with~~ a shorter response. More complex verbal expressions are more effortful to produce, but more importantly they are less likely to be *consistently* produced. While it is logically possible that naming consistency could be independent of response length, in practice the two are strongly related. When a language lexicalizes a distinction~~,~~ ensuring that it has a compact verbal expression~~,~~ then, all else being equal, people are more likely to use that term, leading to greater consistency.

A quick Google search makes this point in the domain of color names. Among the colors that English lexicalizes are "yellow," "green," and "blue." We can get a quick sense of their relative frequency by enumerating the number of webpages containing these terms as indexed by Google: 7.13, 14.14, and 15.31 billion, respectively. Expressions with modifiers are, by comparison, much less frequent, e.g., "light blue" (122 million) and "dark blue" (114 million). Like English, Russian lexicalizes yellow, green, and blue. As in English, we find approximately a 1:2 frequency difference between green ("zeleniy": 105 million) and yellow ("zheltiy": 61 million).[6] However, the lexicalization of blueness in Russian differs from its lexicalization in English. Russian does not have a single term that corresponds to the English meaning "blue." Instead, Russian lexicalizes "dark blue" (siniy; синий) and "light blue" (goluboy; голубой). The frequency of "siniy" is, at 105 million, roughly equal to that of "zeleniy" (green), a basic color. The frequency of "goluboy" is, at 61 million, roughly equal to zheltiy (yellow), another basic color. There is no word or phrase in English that denotes a *shade* of blue that has anywhere close to this relative frequency. The category "light blue" is clearly more nameable in Russian than in English.

In the analyses below, we capture this complexity-based sense of nameability by computing the number of content words a participant uses. For more complex stimuli that elicit multi-word responses, we also use a count of clauses.[7] Notice that while agreement-based measures of nameability cannot, by definition, be computed from individual respondents, only groups, this is not the case for complexity-based measures of nameability.

Naming the Difference: Nameability and Finding Solutions to Bongard Problems

Nameability and Complex Problem Solving: The Case of Bongard Problems

Bongard problems are a set of categorization problems developed by Mikhail Bongard (1967), a Russian computer scientist, who was interested in the automation of visual perception. Bongard's initial 100 problems were later popularized in the English-speaking world by Hofstadter (Hofstadter, 1979/1999) who used them as an illustration of the power of the human mind to find commonalities between images, and as test cases for models of human pattern recognition (Foundalis, 2006). Each problem consists of 12 images: six on the left and six on the right. The task is to discover the rule that distinguishes the six images on one side from the six images on the other. From this simple premise, Bongard, Hofstadter, Foundalis, and others have created hundreds of fascinating problems ranging from simple (solvable in a few seconds) to extremely difficult.[8] What is of interest is why some problems are easy and others difficult.

An inspection of Figure 6.2 reveals that the answer often has little to do with perceptual factors. For example, from the perspective of a feature-based visual pattern detector, identifying what the six shapes on the left of Figure 6.2A all have in common is extremely complex (Linhares, 2000). Yet this problem is trivial for people. Consider now the problem in Figure 6.2B. A geometric pattern analyzer that was flummoxed by Figure 6.2A would have no problem here. A simple geometric feature – convexity – separates the shapes on the left from those on the right. The figures on the left are all convex; the shapes on the right are not. Despite the geometric simplicity of this problem, it poses substantial difficulty for our participants. In our data, only about 21% of participants (English-speaking adults) discovered an acceptable solution. Note that although concavity/convexity is lexicalized in English, the terms "concave" and "convex" are not well known by most English speakers. While it is fair to say that "triangle" and "circle" are words one learns in the course of simply learning English, the same cannot be said for "concave" and "convex."

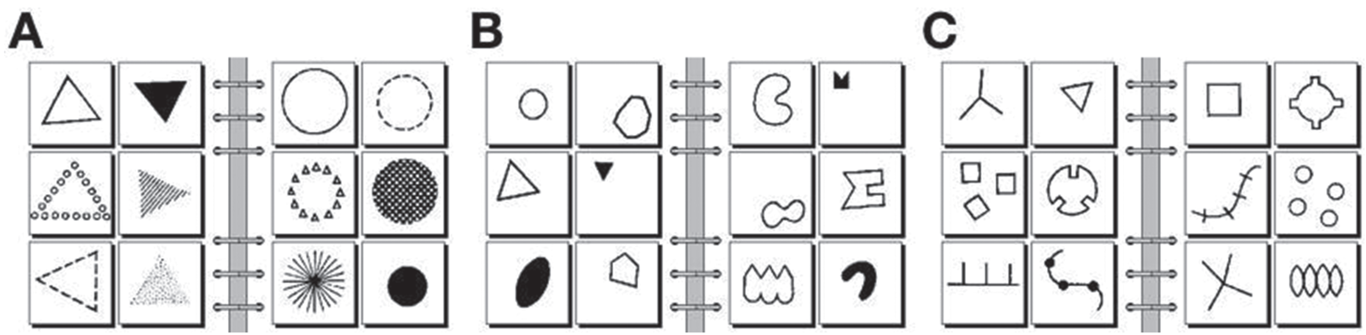

Figure 6.2 Three example Bongard problems. Possible solutions to the problems are (A) triangle vs. circle, (B) convex vs. concave, and (C) "three-ness" vs. "four-ness".

Figure 6.2C provides another instructive example. The figures here are more perceptually complex than in Figure 6.2B, yet there is a readily accessible verbal solution: the figures on the left represent "three-ness" in some way (edges, number of figures, number of lines, etc.), while the figures on the right represent "four-ness." A far higher percentage of participants succeed at this problem (~70%), despite its apparent perceptual complexity. What makes problems A and C so easy, but problem B so hard? We think nameability has something to do with it. Is it merely a coincidence that the rule instantiated by problems A and C lends itself to a verbal expression that is both highly accessible (because of its frequency) and compact, while problem B does not?

To examine the relationship between nameability and ease of solving Bongard problems, we first revisited data from Foundalis' (2006) dissertation and examined whether the solution complexity of the "ideal" solutions to each problem (as formulated by the problem inventors) correlated with solution success. The answer is clear: regardless of whether naming complexity is quantified as number of content words or number of clauses, problems with longer solutions are less likely to be solved (see Figure 6.3). Note that verbal complexity does not map onto perceptual complexity in any straightforward manner in these problems.

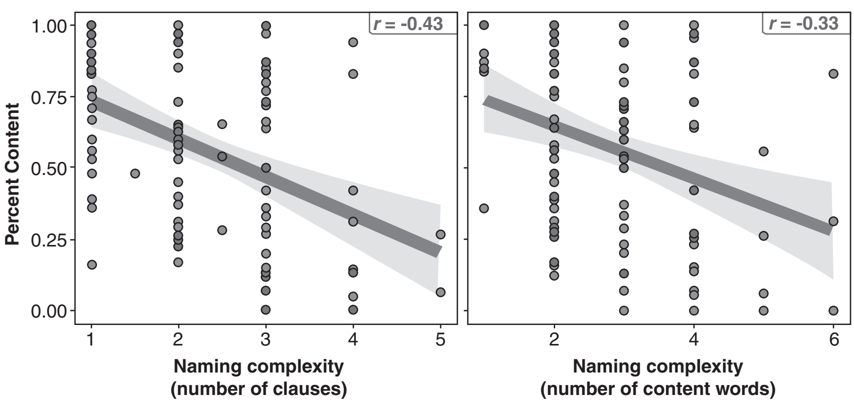

Figure 6.3 Relationship between solution accuracy and the naming complexity of the ideal solution, analyzed using data from Foundalis (2006). Naming complexity was assessed in terms of the number of clauses (left) and the number of unique content words (right).

A problem with this initial analysis is that the solutions whose length we are measuring are the "ideal" solutions according to the experimenters, rather than the solutions people actually give when trying to solve these problems. To examine whether similar relationships are observed between solution success rates and naming complexity, we tested a group of participants (n = 89) on a subset of 16 Bongard problems. Participants' verbal description of their solution was subsequently coded for accuracy. We found that problems with higher average verbal complexity (as measured by the average number of content words used in correct responses) and higher naming convergence (the percentage of unique words used across correct responses) were also more difficult for people to solve (see Figure 6.4; verbal complexity: $z = -4.87$, $p < .001$; naming divergence: $z = -3.00$, $p = .003$). Thus, a powerful predictor of the difficulty of a Bongard problem is the compactness of its verbal description.

These initial analyses suffer from two limitations. First, there is a circularity in relying on verbal descriptions both for determining participants' accuracy and measuring the verbal complexity of normatively correct responses. In the next section, we will discuss new data aimed at both collecting verbal complexity measures independently from the original Bongard problems themselves, and collecting more objective measures of solution accuracy. A second, broader limitation is

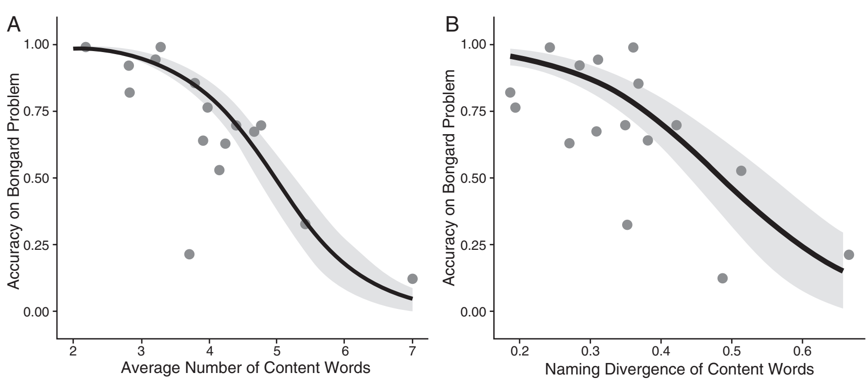

Figure 6.4 The relationship between verbal complexity as measured by (A) average number of content words in solutions provided by our participants (A), (B) naming divergence of content words (see the section "Nameability defined") in these same solutions. Each point represents a Bongard problem.

distinguishing correlation from causation. An alternative explanation for the relationship between nameability and performance is that more "difficult" distinctions – where difficulty is defined on some separate metric – are both more difficult to name and more difficult to solve. If true, then the observed correlations between problem difficulty and nameability do not reflect any causal influence of language on problem solving. We will address this concern in sections "Words as guides to category joints: manipulating nameability in category learning tasks" and "Nameability and geometric reasoning: strengthening the case for a causal link."

Overcoming Circularity: Verbal Complexity Predicts People's Ability to Discover Solutions to Physical Bongard Problems

To help overcome the circularity that arises in using verbal solutions for both measuring accuracy and nameability, we developed a set of simplified Bongard problems that isolated the dimension central to solving the full problem, and then used the nameability of the simplified problems to predict people's performance on the full problems.

Rather than using the original Bongard problems, we used a variant of Bongard problems that depict simple physical events and relationships requiring participants to reason about similarities and differences between these events, and often requiring participants to mentally simulate how an event will unfold in time (Weitnauer & Ritter, 2012).[9] For example, in problem 9 (Figure 6.5A) what makes the scenes on the left different from the scenes on the right is that the two objects will move in the same direction in the scenes on the right, while they will move in opposite directions in the scenes on the left.

We began by asking participants to identify what makes one set of scenes different from the other in the simplified versions of 11 physical Bongard problems (see Figure 6.5B).[10] We collected responses from 85 participants, each of whom provided verbal rules for six of the problems. After coding the correctness of each verbal description, we calculated an average complexity score of the verbal solutions by computing the average number of content words in correct verbal solutions for each of the 11 problems. We then tested a new set of participants (n = 83) on their ability to solve the original physical Bongard problems (see Figure 6.5A). We found that the

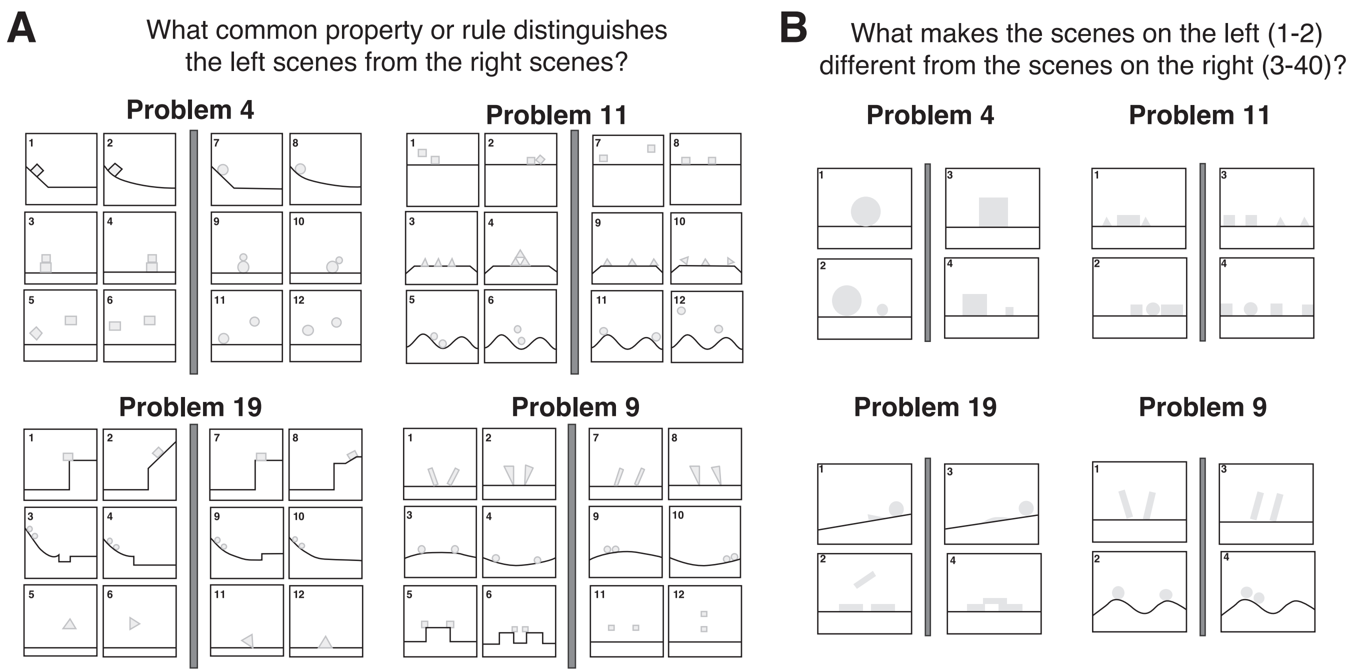

Figure 6.5 (A) Example physical Bongard problems. Possible solutions to the problems are: (problem 4) squares vs. circles; (problem 11b) objects close to one another vs. objects far from one another; (problem 19) at least one object travels through the air vs. all objects always maintain contact with the ground, and (problem 9) objects move in opposite directions vs. items move in the same direction. (B) Simplified versions that we used for collecting nameability data. These simplified versions seek to isolate the dimension most central to solving the full problem.

verbal complexity of the solutions provided for the simplified versions of the problems predicted participants' ability to provide a correct verbal solution to the full versions of the problems: problems with more complex verbal solutions were more difficult to solve, $z = 2.46$, $p = .01$ (see Figure 6.6). The correlation between average accuracy on a problem and verbal complexity was $r = -.60$, $t(9) = -2.26$, $p = .05$ (Baird, Zettersten, & Lupyan, unpublished data).

As a second step to overcoming circularity, we conducted an additional experiment ($n = 202$) in which participants were asked to discover the solution to one of the 11 physical Bongard problems by sorting the scenes into groups. We found that verbal complexity predicted not only their ability to verbalize a correct solution ($z = 2.58$, $p = .01$), replicating the previous result, but also led to greater objective accuracy in classifying new category exemplars, as measured by their ability to sort novel exemplars into the correct category ($z = 2.68$, $p = .007$).

To see why verbal complexity is so strongly related to accuracy in solving these particular problems, consider one reason Bongard problems are difficult in the first place. The key challenge is discovering what the

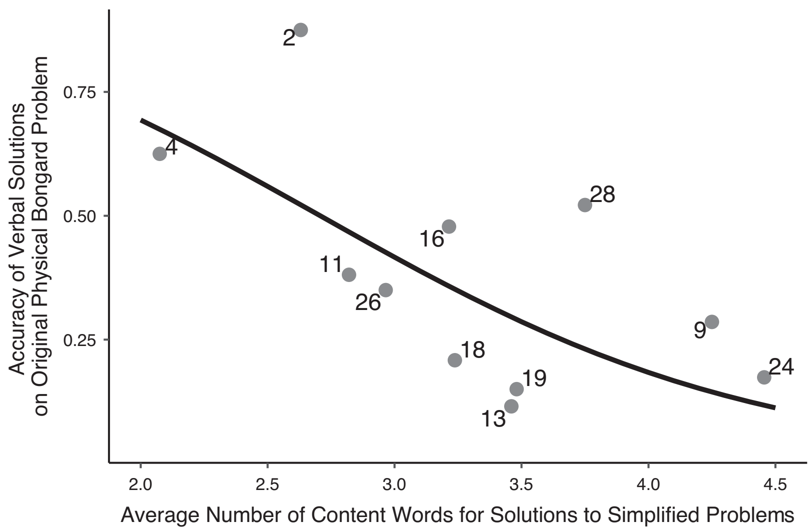

Figure 6.6 Relationship between verbal complexity of the solution (average number of content words in correct solutions) for simplified problems, and accuracy of verbal solutions on the original physical Bongard problems. Each point represents a problem.

relevant dimensions or features are for solving each problem. This is an open-ended task that changes from one Bongard problem to the next. In one problem, size is relevant; in the next, something about the contours of the shapes; the next might require representing each group in terms of a more abstract relation such as "same" and "different" or "threes" and "fours." When do these features come to mind, and why are some easier to discover more than others? This is where we believe language plays a critical role. Whether a feature is discovered may partly depend on how easily it can be formulated as a verbal hypothesis. Once formulated verbally, the hypothesis becomes easy to test against the images. On this account, the difficulty in discovering the rule in Figure 6.2B is, in part, due to the property of convexity being difficult to name for our participants.[11] In the next section, we provide further evidence for the idea that easier-to-name visual features are more likely to be used by people when judging visual similarity.

A Shape by any other Name is not as Similar: Nameability Predicts Similarity Judgments

Does the nameability of features affect the weight that people give them? For instance, are objects more likely to be grouped together if they share a more nameable feature? We tested this question with a set of items with unfamiliar global shapes developed by Roland Fleming (pers. comm.). These items were created in pairs such that for each novel shape there were several outline types (e.g., compare the left and right shapes in each pair in Figure 6.7). Some of these unusual outlines can be compactly described ("curved," "bubbly"), while others do not lend themselves to compact descriptions ("kind of jaggedy splitting thing"). Does this difference in nameability influence how individuals reason about these unfamiliar shapes?

To answer this question, we first collected information on how easily people can describe the properties of shape outlines such as those shown in Figure 6.7. We presented participants with pairs of shapes differing only in the outline type, then asked them to describe the difference (Figure 6.8A). We computed the average number of content words participants (n = 40) used to describe the surface outline differences (Figure 6.8A). We then tested a separate group of participants (n = 50) in a triad task (Figure 6.8B).

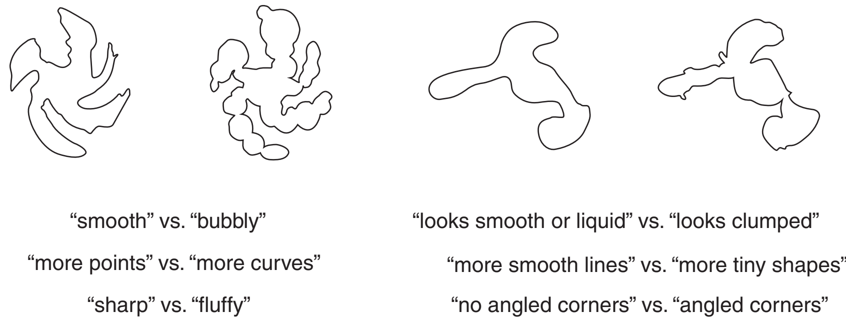

Figure 6.7 Example pairs of shapes and the verbal descriptions people provided to describe what makes the two shapes different.

On each trial, participants were asked to choose which of two images were more similar to a target image. One of the choices always matched the image in global shape but differed in its outline type (shape match), while the other matched the outline type while differing in its outline (surface match). There was a strong correlation between participants' likelihood of matching the images on surface outline (choosing the surface match) and the difficulty of verbally describing the particular surface outline (see Figure 6.8C): easier to describe surface outlines were more likely to be chosen as the feature by which to group items, $r = -.69$, $t(15) = 3.54$, $p < .01$.

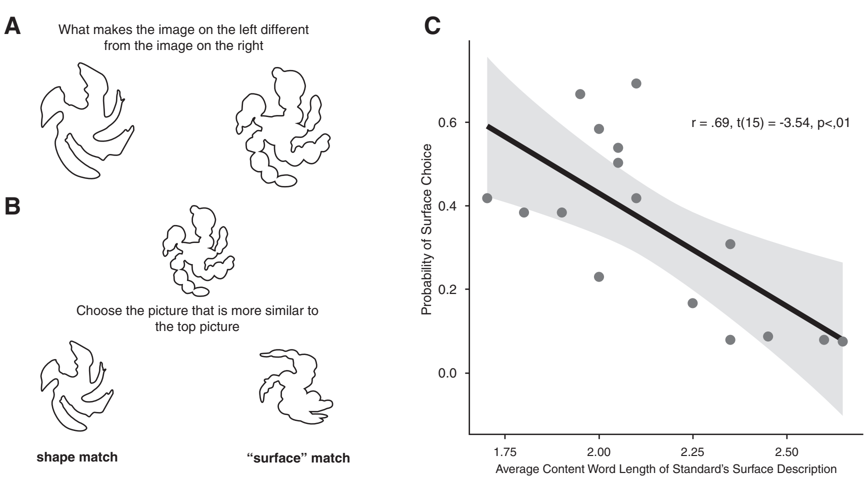

Figure 6.8 The length of the verbal descriptions given for shape "surfaces" predicts the likelihood of "surface choices." Each point represents a different triad.

Words as Guides to Category Joints: Manipulating Nameability in Category Learning Tasks

The previous task shows that how easily people can form verbal descriptions of features is associated with how likely those features are used when grouping together novel images. If nameability influences which features come to mind, might we find that people can learn novel categories more easily if they differ on more nameable features? Or are the similarity judgments we saw in the previous task ephemeral and easily overridden, such that verbally based feature preferences ultimately have little consequence in shaping people's category representations? The experiments we describe next show that nameability can have substantive consequences in categorization tasks in which participants must learn novel categories and are given explicit feedback on their performance.

A lingering concern from many of the studies presented so far is that more "complex" categories are simply more difficult or more complex to verbalize. In recent work (Zettersten & Lupyan, 2019), we sought to test whether nameability affects people's ability to learn novel rule-based categories when holding the underlying conceptual complexity of the categories constant. First, we analyzed data from a large-scale online color-naming study (*N* = 134,727; Munroe, 2010) to determine the ease of naming a broad swath of different colors. We then selected a set of colors that were highly nameable (named according to their modal label by 80%–85% of the population) and a set of colors that were much more difficult to name (modal names used by 6%–10% of participants in the original naming task), while matching the color sets on distinctiveness. We then constructed two different categories with identical structure for the easy-to-name color set and the more difficult-to-name color set (see Figure 6.9). For both categories, a single color was perfectly predictive of category membership, e.g., "red" vs. "brown" for the high nameability condition and "lavender" vs. "olive"[12] for the low nameability condition. Would categories composed of more nameable features be easier to learn than categories composed of less nameable features?

Participants learned the categories with the identical conceptual structure more accurately when the underlying features of the category were more nameable (see Figure 6.9A). That is, they were more likely to learn

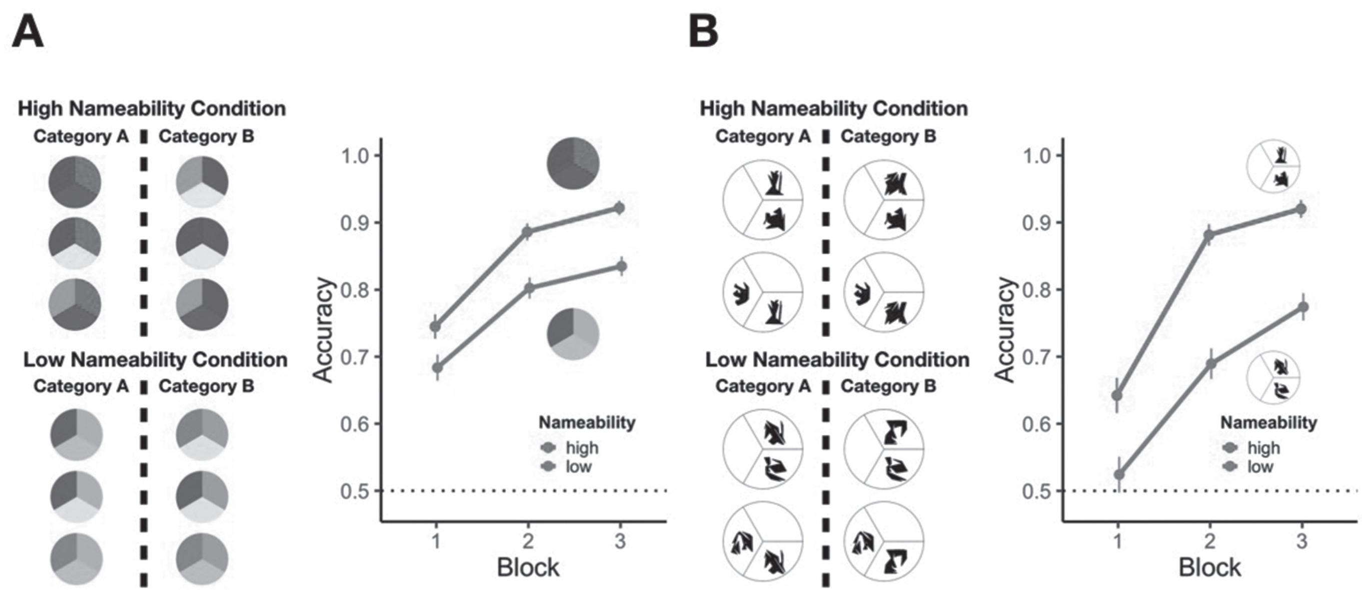

Figure 6.9 Nameability of individual features predicts categorization accuracy for (A) color features and (B) shape features. The left side of (A) and (B) depicts the category structure and example stimuli in the high and low nameability conditions. The right sides show categorization performance for each training block showing superior performance for the more nameable colors (A) and shapes (B).

the categories when the features were red and brown difficult-to-describe lavender and olive colors. This result is not restricted to particular kinds of features or category structures. We observed similar results when testing category learning for rule-based categories composed of more nameable (though still novel) shapes compared to less nameable novel shapes (see Figure 6.9B), and for compositional categories that required combining shape and color information. Together, these findings suggest that controlling for the logical complexity of categories, those composed of more nameable features were easier to learn. By prioritizing some features over others, language can affect the ease with which categorical joints can be carved into the environment. An important limitation of these results is that both Bongard problems and the categorization tasks we described pertain to a subclass of categorization problems – those requiring rule-based solutions rather than the kind of family-resemblance structure that characterize many of the categories learned by young children and non-human animals.

Nameability and Geometric Reasoning: Strengthening the Case for a Causal Link

So far, we have shown that: (1) success on Bongard problems – a type of category induction problem – varies with the ease of verbally expressing the rule/pattern that has to be induced (see the section "Naming the difference: nameability and finding solutions to Bongard problems"); (2) the likelihood that a certain visual feature influences visual similarity is predicted by the ease of naming that feature (see the section "A shape by any other name is not as similar: nameability predicts similarity judgments"); and (3) learning novel categories is greatly facilitated when the categories comprise easy-to-name compared to difficult-to-name features (see the section "Words as guides to category joints: manipulating nameability in category learning tasks").

These three lines of evidence make it clear that nameability strongly predicts performance in a range of categorization tasks. One interpretation ~~of these results~~ is that the results support the linguistic priority thesis. Another interpretation, mentioned above, is that we have it exactly backward. It may be that nameability is predictive because whatever causes conceptual difficulties also causes difficulties in naming. On this view, the causality runs from conceptual difficulty to nameability, rather than the other way around. This possibility would be in line with the cognitive priority thesis.

The data presented in sections "Naming the difference: nameability and finding solutions to Bongard problems"and "A shape by any other name is not as similar: nameability predicts similarity judgments," while showing that nameability is predictive of performance on a variety of tasks, cannot distinguish causal direction. The category-learning data (see the section "Words as guides to category joints: manipulating nameability in category learning tasks) does begin to distinguish them in that there does not appear to be an a priori metric on which the easy-to-name colors and shapes are simpler than the hard-to-name ones (Zettersten & Lupyan, 2019). The case for the linguistic priority thesis can be further strengthened in two ways: (1) by showing that the cognitive difficulty of appreciating certain distinctions varies with how nameable they are in different languages and (2) by showing that manipulating nameability affects performance selectively on items predicted to be most influenced by linguistic experience.

In this section we summarize ongoing work that subjects the linguistic priority thesis to these two tests. The domain we use is geometric reasoning. We chose this domain because it has been explicitly claimed that basic geometric reasoning is independent of language (and culture more generally) and is part of people's core knowledge (Dehaene, Izard, Pica, & Spelke, 2006).

The task used by Dehaene et al. (2006) uses an odd-one-out design to tap into geometric reasoning. People are presented with groups of six figures and asked to select from each group the one that does not belong with the others (the "target"). Figure 6.10 shows two sample trials. The key evidence the authors use to support their argument that geometric reasoning is independent of language comes from a comparison of performance on this task by educated American adults and the Mundurukú. The Mundurukú are an Amazonian indigenous people without formal education and who do not possess a conventional vocabulary for describing the geometric relations in question. Although Americans performed much better on the task overall, there was a strong correlation (r ~ .7–.8)

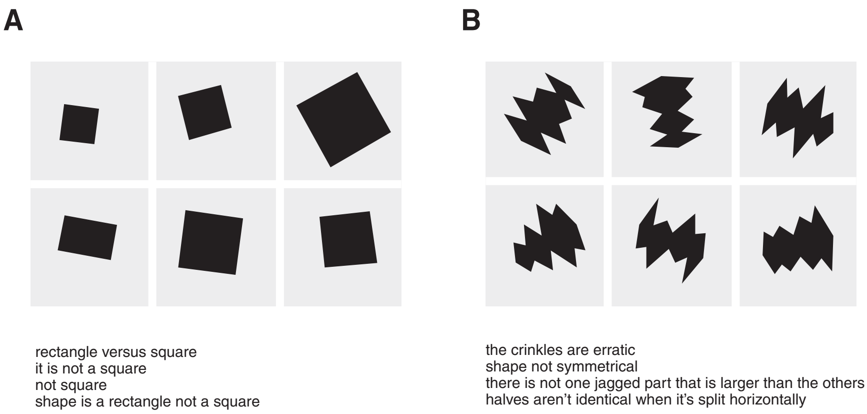

Figure 6.10 Sample geometric reasoning trials with some of the responses produced by participants describing how the odd-one-out shape (target) differs from the non-target shapes.

between item accuracy in the two groups. This high item correlation led Dehaene et al. (2006) to conclude that the Mundurukú shared geometric knowledge ("core geometry") with American subjects despite lacking linguistic and other cultural sources for this knowledge. We used the very same task to support the claim of universality and non-language dependence of geometric reasoning to reach the opposite conclusion. The results we report here are abbreviated; a full report is forthcoming (Lupyan, Wendorf, Rojas-Berscia, & Paul, 2018).

The first thing to note about this kind of odd-one-out task is that choosing the target requires identifying (either explicitly or implicitly) the dimension of variation that is most relevant. On many trials, choosing different dimensions will lead to different answers. For example, someone might identify surface area as the relevant dimension in Figure 6.10A and choose the top-right choice because it is the shape with the largest surface area. The targets designated as correct are defined on dimensions deemed geometrically relevant. Importantly, these dimensions create discrete rather than continuous differences between the target and non-targets. The difference between a square and a rectangle is one of kind: a square has all equal sides; a non-square rectangle does not. The second thing to notice about the task is that some of these dimensions are more nameable than others. For example, virtually everyone describes the distinction between the normatively correct target and the non-targets as "rectangle vs. square" (Figure 6.10A). The distinction in Figure 6.10B – having to do with reflection symmetry – is relatively difficult to name.

The first question we ask is whether nameability predicts solution accuracy. We use a complexity-based definition of nameability (the number of modifiers used on average when describing what makes the target different from the non-targets). Nameability is strongly correlated with performance ($r = -.49$, $p < .01$): the fewer modifiers people needed, the more accurate were the responses of a separate group of participants. We next examined which items showed the largest differences between the accuracy of our subjects (American adults) and the accuracy previously reported for Mundurukú participants. If items such as rectangles-vs.-squares are easy for English speakers *because* they are easy to name, then it is these items that should show the largest difference between the two groups. This is indeed what we found. English nameability was significantly correlated with the *difference* between American and Mundurukú ($r = -.33$, $p = .04$). Americans performed relatively well on the items that were most nameable in English.

We next collected data on two additional populations: (1) congenitally deaf children residing in a Chinese special school for the deaf who were deprived of normal language input for most of their childhood, compared to the performance of children with normal language input, and (2) the Shawi, an indigenous group of horticulturalist traders from Northwestern Amazonia who speak a Kawapanan language. The Shawi we tested varied in formal education and knowledge of Spanish. Our results replicate Dehaene et al's (2006) finding of substantial correlations in performance (r's $> .6$) on this task even among these very disparate populations. However, children with impaired language input performed substantially worse (M = .50) than children with normal language input (M = .75; $t = 4.1$, $p < .01$). The performance of the former was predicted by proficiency with Chinese sign language. The Shawi performed poorly (M = .41) though, like the Mundurukú, considerably above chance (chance = .17). The Shawi's performance was strongly modulated by their knowledge of Spanish. Importantly, neither the Mundurukú's nor the Shawi's responses were predicted by English nameability, suggesting that geometric relations that are easy to name are not universally accessible, but *become* easy when compact verbal descriptions are available. Evaluations of Chinese and Shawi nameability measures are in progress. An additional prediction, which we do not have sufficient data to test at present, is that the items on which children and adults differ the most should also be those that are most nameable by adults.

Finally, we examined what happens when we manipulate language in our English-speaking adults by either asking people to verbally justify their chosen answer (a way of upregulating the use of language/verbal strategies), and interfering with language by having participants repeat "a b c" while doing the task on half of the trials. Overt naming improved accuracy ($t = 3.70$, $p < .01$). This improvement could not be attributed to merely greater effort spent on the task. Conversely, verbal interference impaired performance ($t = 2.76$, $p = .01$). While overt naming increased performance for hard-to-name items ($r = -.37$, $p = .02$), verbal interference selectively impaired performance on the normally easy-to-name items ($r = .35$, $p = .03$).

As in the section "Nameability and geometric reasoning: strengthening the case for a causal link," we believe the role that is played by language in this task is one of facilitating hypothesis formation. Presenting English speakers with five rectangles and a square – objects that are highly nameable in English – makes it easy to pose the hypothesis that the relevant distinction is between rectangles and squares. This hypothesis becomes less available when the distinction is less nameable, either because the lexicalized distinction is not readily available in the language (or simply unknown to the participant, as in the case of low-frequency terms like "concave"/"convex"), or because it was *made* less available by interfering with language during the task.

CONCLUSION

There is no doubt that humans, like other animals, enter the world with numerous biases that guide and constrain the conceptual knowledge we go on to develop. And yet, the sheer variety of ways there are to be human is a testament to the incredible flexibility of our species (Henrich, 2015; Prinz, 2014). Our success in adapting to such varying environments requires the ability not only to learn from others, but to maintain an ever-growing repository of information to which children become exposed (we even have a word for this process: enculturation). Although the centrality of language to this process is widely acknowledged (Pinker, 2010), the role that language plays in structuring our minds is nevertheless frequently denied (sometimes by the very same people; cf. Pinker, 1994).

We have argued for a constructive view of language in human cognition. The environmental experience of any one individual greatly underdetermines any one conceptual scheme. Our language offers us a system of categories, most of which have undergone extended cultural evolution, and many of which we acquire "for free" in the course of learning a language. The ways in which words help create categories are varied (see the sections "Two arguments against the cognitive priority hypothesis," "Two challenges to the linguistic priority hypothesis," and "Vocabulary as a joint of nature"): they include both offline mechanisms such as cohering otherwise disparate entities during word learning, and online mechanisms such as helping to posit hypotheses and performing in-the-moment dimensionality reduction.

Many of the findings we described here are preliminary. None of them on their own unequivocally support the thesis that the vocabulary we learn as part of learning a language helps structure the mind. Yet taken together, we believe results like the ones reported here are difficult to reconcile with a strong cognitive-priority perspective. To us, these findings hint at the wealth of other effects that may be revealed through a systematic study of the effect of learning and using words on our conceptual structure. If simply making a distinction slightly more or less nameable can have the kinds of effects we describe, what might this mean for the more protracted developmental differences experienced by people learning languages with substantially different vocabularies? What might be the downstream effect of spending one's childhood immersed in different linguistic environments? Clark's (1998) analogy of the mangrove is useful here. Just as the effect of a mangrove seed on its landscape is small at first, so we think the effect that words have on the conceptual landscape are likely to magnify over development. At the same time, as one becomes more fluent with the vocabulary of a language, it becomes increasingly difficult to appreciate ~~(and to study!)~~ its influence on our cognition. ~~It is difficult to appreciate that w~~ithout the seeds ~~that are~~

~~the~~ words like "blue," "triangle," and "hundred," our ability to reason in fundamental domains such as color, geometric relations, and numbers may be very different from what we know it to be.

ACKNOWLEDGMENTS

The writing of this chapter was supported in part by NSF-1734260 and NIH-R21HD092867 to the first author. We thank Nicholas Baird for help with developing the experiments and collecting the data reported in the section "Overcoming circularity: verbal complexity predicts people's ability to discover solutions to physical Bongard problems," Justin Sulik with preliminary analyses of the Bongard problems, and Roland Fleming for sharing the stimuli shown in Figures 6.7 and 6.8.

NOTES

[1] Some have referred to claims that the presence or absence of a word in a language has interesting consequences for cognition as the "No Word for X" fallacy. The blog *Language Log* aggregates such claims (mostly from the popular media) in their "No Word for X" archive (http://languagelog.ldc.upenn.edu/nll/?p=1081).

[2] Learners are almost assured of succeeding in learning these words because if too many cannot, the unlearned meanings would not be transmitted to the next generation of speakers. Finding that some languages lexicalize certain distinctions is therefore prima facie evidence of the learnability of these distinctions by a large majority of the speech community.

[3] Proper names – Maggie, the Eiffel Tower – are also categories, narrower still. Although they denote specific individuals, the denotation extends in space and time and those experiences constitute categories. Important classes of linguistic terms that do not denote categories are logical terms such as "and," "or," "not," indefinite pronouns like "somebody" and "neither," and highly relational words like "same".

[4] Chinese uses a simple term for 10,000 (万), which, we expect, makes 10,000 a better "chunk" in Chinese than in English.

[5] Note that representing the category distinction in terms of T vs. L will not necessarily lead to 100% performance, since it might lead participants astray in some of the boundary cases. Nevertheless, we reasoned that grounding the categories in the T/L distinction should lead to higher accuracy.

[6] The large absolute differences in the number of English and Russian simply reflect the dominance of English-language websites in Google's index.

[7] Although the two senses of nameability introduced here – naming consistency and naming complexity – often agree, we have found that naming consistency measures are more predictive when the naming task constrains participants' responses to 1–2 words, and becomes less predictive as the length of responding increases. In general, we focus on naming consistency when nameability data was based on short verbal responses, and on naming complexity when verbal responses were more open ended (such as in the case of Bongard problems below).

[8] Foundalis comments in his dissertation that Hofstadter did not write down solutions to many of the problems he developed, and Foundalis ultimately found the solutions to many of these problems only with the help of responses from readers of his webpage (see http://www.foundalis.com/res/diss_research.html).

[9] To view the full set of the physical Bongard problems created by Weitnauer and Ritter (2012) along with their solutions, see https://github.com/eweitnauer/PBPs/tree/master/pngs/all-with-sol. We selected 11 of these physical Bongard problems for our experiments, using the first 12 example images for each problem (see Figure 6.5A). The numbers in Figures 6.5 and 6.6 correspond to the numbering used by Weitnauer and Ritter (2012).

[10] Many of these scenes were short animations that demonstrated the unfolding of the event over time, e.g., https://github.com/mzettersten/vocab-mind-2020/blob/master/pbp/stimuli/namingTask/pbp09_naming.gif.

[11] Bongard's original 100 problems are ordered roughly by difficulty (with a few notable exceptions to highlight the ease with which people make solve certain perceptually difficult problems such as Figure 6.2C). It is curious then that the concave/convex problem is presented very early at number 4. It is conceivable that the difficulty that English speakers have with this problem is not mirrored by Russian speakers (i.e., Mikhail Bongard himself and the original audience of his book). The Russian word for "concave" ("вогнутый") is relatively rare, but the word for "convex" ("выпуклый"), literally "bulging", is much more frequent than the English tranlation equivalent. For example, it is more than twice as frequent as the Russian word for "triangle" ("треугольник"). We speculate that the relative ease of naming convexity in Russian may enable Russian-speakers to do better on this problem. Bongard may have listed it early on because the solution was more obvious to him.

[12] Although these are modal names, "olive" and "lavender" are produced by only about 10% of respondents.

REFERENCES

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161. doi:10.1111/j.1749-6632.2010.05874.x

Bloom, P. (2002). *How children learn the meanings of words*. A Bradford Book.

Bongard, M. M. (1967). *Pattern recognition*. Hayden Book Co., Spartan Books.

Boroditsky, L. (2010). How the languages we speak shape the ways we think: The FAQs. In M. J. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics* (p. forthcoming). Cambridge University Press.

Bowerman, M. (2000). Where do children's word meanings come from? Rethinking the role of cognition in early semantic development. In L. Nucci & E. Saxe (Eds.), *Culture, thought, and development* (pp. 199–230). Lawrence Erlbaum. https://books.google.com/books?hl=en&lr=&id=TsZ5AgAAQBAJ&oi=fnd&pg=PA173&ots=wQ71Nw1nVD&sig=uG37FVCKBMuJP3-Z2VamRRthWbs

Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge University Press.

Bowerman, M., & Levinson, S. C. (2001). *Language acquisition and conceptual development*. Cambridge University Press.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992). Verbal recoding of visual-stimuli impairs mental image transformations. *Memory & Cognition*, *20*(4), 449–455.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, *5*(5), e10773. doi:10.1371/journal.pone.0010773

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674.

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 162–183). Cambridge University Press.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. S. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science*, *311*(5759), 381–384. doi:10.1126/science.1121739

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(05), 429. doi:10.1017/S0140525X0999094X

Fodor, J. A. (1975). *The language of thought*. Harvard University Press.

Fodor, J. A. (2001). Language, thought and compositionality. *Mind & Language*, *16*(1), 1–15. doi:10.1111/1468-0017.00153

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105–1123. doi:10.1037/xge0000560

Foundalis, H. E. (2006). Phaeaco: A cognitive architecture inspired by Bongard's problems PhD thesis). Indiana University.

Gleitman, L., & Fisher, C. (2005). Universal aspects of word learning. In J. McGilvray (Ed.), *The Cambridge companion to Chomsky* (pp. 123–142). Cambridge University Press. https://books.google.com/books?hl=en&lr=&id=I6CZ6wpNKeEC&oi=fnd&pg=PA123&ots=ujaHwtyK8O&sig=N-te0BeTbB0IyV7aOmuwBAqwtCk

Gleitman, L., & Papafragou, A. (2005). Language and thought. In K. Holyoak & B. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (pp. 633–661). Cambridge University Press.

Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton University Press.

Hofstadter, D. R. (1999). *Godel, Escher, Bach: An eternal golden braid* (20 Anv). Basic Books.

James, W. (1890). *Principles of psychology* (Vol. *1*). Holt.

Kay, P., Berlin, B., Maffi, L., Merrifield, W. R., & Cook, R. (2011). *The world color survey* (1st ed.). Center for the Study of Language and Information.

Kay, P., & Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, *101*(4), 743–760. doi:10.1525/aa.1999.101.4.743

Lachman, R. (1973). Uncertainty effects on time to access the internal lexicon. *Journal of Experimental Psychology*, 99(2), 199–208. doi:10.1037/h0034633

Leavitt, J. (2011). *Linguistic relativities: Language diversity and modern thought.* Cambridge University Press.

Lee, P. (1996). *The Whorf theory complex: A critical reconstruction.* John Benjamins Publishing Company.

Levinson, S. C. (2003). Language and mind: Let's get the issues straight! In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 25–46). MIT Press.

Linhares, A. (2000). A glimpse at the metaphysics of Bongard problems. *Artificial Intelligence*, *121*(1), 251–270.

Locke, J. (1849). *An essay concerning human understanding. With the notes and illustrations of the author, and an analysis of his doctrine of ideas.* Tegg.

Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Cognition*, 8(3)54. doi:10.3389/fpsyg.2012.00054

Lupyan, G. (2012b). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. *57*, pp. 255–297). Academic Press. http://www.sciencedirect.com/science/article/pii/B9780123942937000078

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning*, 66(3), 516–553. doi:10.1111/lang.12155

Lupyan, G., & Bergen, B. (2016). How language programs the mind. *Topics in Cognitive Science*, *8*(2), 408–424. doi:10.1111/tops.12155

Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience,34(1), 1–19*. doi:10.1080/23273798.2017.1404114

Lupyan, G., Wendorf, A., Rojas-Berscia, L. M., & Paul, J. (2018). Core knowledge or language-augmented cognition? The case of geometric reasoning. *Proceedings of the 12th International Conference on the Evolution of Language (Evolang12)*.

Lupyan, G., & Winter, B. (2018). Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *373*(1752). doi:10.1098/rstb.2017.0137

Mahon, B., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*, 27–51. doi:10.1146/annurev.psych.60.110707.163532

Majid, A., & Burenhult, N. (2014). Odors are expressible in language, as long as you speak the right language. *Cognition*, *130*(2), 266–270. doi:10.1016/j.cognition.2013.11.004

Majid, A., Roberts, S. G., Cilissen, L., Emmorey, K., Nicodemus, B., O'Grady, L., … Levinson, S. C. (2018). Differential coding of perception in the world's languages. *Proceedings of the National Academy of Sciences*, *115*(45), 11369–11376. doi:10.1073/pnas.1720419115

Malt, B. C., Gennari, S., Imai, M., Ameel, E., Saji, N., & Majid, A. (2015). Where are the concepts? What words can and can't reveal. In E. Margolis & S. Laurence (Eds.), *Concepts: New directions*. MIT Press.

Malt, B. C., & Wolff, P. M. (2010). *Words and the mind: How words capture human experience*. Oxford University Press, USA.

McWhorter, J. H. (2014). *The language Hoax: Why the world looks the same in any language*. Oxford University Press.

Munroe, R. P. (2010, May 4). Color Survey Results. Xkcd. http://blog.xkcd.com/2010/05/03/color-survey-results

Perry, L. K., & Lupyan, G. (2016). Recognising a zebra from its stripes and the stripes from "zebra": The role of verbal labels in selecting category relevant information. *Language, Cognition and Neuroscience, 32(8), 925–943*. doi:10.1080/23273798.2016.1154974

Pinker, S. (1994). *The language instinct*. Harper Collins.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. Viking.

Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 8993–8999. doi:10.1073/pnas.0914630107

Prinz, J. J. (2014). *Beyond human nature: How culture and experience shape the human mind* (1st edition). W. W. Norton & Company.

Pullum, G. K. (1989). The great eskimo vocabulary hoax. *Natural Language & Linguistic Theory*, *7*(2), 275–281. doi:10.1007/BF00138079

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688. doi:10.1038/163688a0

Snedeker, J., & Gleitman, L. (2004). Why is it hard to label our concepts? In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (illustrated edition, pp. 257–294). The MIT Press.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, 6(2), 174–215. doi:10.1037/0278-7393.6.2.174

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96.

USNPS. (2018). Trees and shrubs – North Cascades National Park. https://www.nps.gov/noca/learn/nature/treesandshrubs.htm

Weitnauer, E., & Ritter, H. (2012). Physical Bongard problems. *Artificial Intelligence Applications and Innovations*, 157–163. doi:10.1007/978-3-642-33409-2_17

Whorf, B. L. (1956). *Language, thought, and reality* (J. B. Carroll, Ed.). MIT Press.

Wierzbicka, A. (2006). The semantics of colour: A new paradigm. In C. P. Biggam & C. Kay (Eds.), *Progress in colour studies* (pp. 1–24). John Benjamins Publishing Company. doi:10.1075/z.pics1.05wie

Wierzbicka, A. (2009). All people eat and drink. Does this mean that "eat" and "drink" are universal human concepts? In J. Newman (Ed.), *Typological studies in language* (Vol. *84*, pp. 65–89). John Benjamins Publishing Company. doi:10.1075/tsl.84.05wie

Wolff, P., & Holmes, K. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(3), 253–265.

Zettersten, M., & Lupyan, G. (2019). Finding categories through words: More nameable features improve category learning. *Cognition*, *196*, 104135. doi:10.1016/j.cognition.2019.104135