THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
http://dx.doi.org/10.1080/17470218.2015.1130730

Routledge
Taylor & Francis Group

The paradox of the universal triangle: Concepts, language, and prototypes

Gary Lupyan

Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA

ABSTRACT
For over 300 years, the humble triangle has served as the paradigmatic example of the problem of abstraction. How can we have the idea of a general triangle even though every experience with triangles is with specific ones? Classical cognitive science seemed to provide an answer in symbolic representation. With its easily enumerated necessary and sufficient conditions, the triangle would appear to be an ideal candidate for being represented in a symbolic form. I show that it is not. Across a variety of tasks—drawing, speeded recognition, unspeeded visual judgments, and inference—representations of triangles appear to be graded and context dependent. I show that using the category name "triangle" activates a more prototypical representation than using an arguably coextensive cue, "three-sided polygon". For example, when asked to draw "triangles" people draw more typical triangles than when asked to draw "three-sided polygons". Altogether, the results support the view that (even formal) concepts have a graded and flexible structure, which takes on a more prototypical and stable form when activated by category labels.

ARTICLE HISTORY
Received 27 May 2015
Accepted 11 September 2015

KEYWORDS
Concepts; Language; Categorization; Prototypes; Inference; Language and thought

Geometrically, a triangle is a triangle regardless of its form—equilateral, isosceles, scalene. Its triangularity is unaffected by which way it "points", and certainly not by what it is called. And yet, as we shall see, people's mental representation of triangles appears to very much depend on such factors (cf. Armstrong, Gleitman, & Gleitman, 1983; Lupyan, 2013).

For over 300 years, the humble triangle has served as the paradigmatic example of the problem of abstraction. In a well-known passage, John Locke commented on the

> pains and skill to [form] the general idea of a triangle, [for] it must be neither oblique, nor [right], neither equilateral, [isosceles], nor scalene; but all and none of these at once. [I]t is something . . . that cannot exist; an idea wherein some part of several different and inconsistent ideas are put together. (Locke, 1995, p. 509)

The philosopher George Berkeley went further, arguing that the seeming impossibility of thinking about an abstract triangle stands in contrast with the ease with which we converse about them, concluding that although it may be possible to approximate an abstract triangle by focusing separately on its angles, its sides, and so on, thoughts about triangles are thoughts about particular triangles (Berkeley, 2009). Hume went further still, arguing that "all general ideas are nothing but particular ones, annexed to a certain term" (Hume, 2012, p. 15).

That these musings go beyond philosophical speculation can be verified by glancing at Figure 1, which contains several—geometrically speaking, equally good—triangles. If the representation of "triangle" were truly abstract, we might expect people to judge all of them as equally good triangles and

CONTACT Gary Lupyan lupyan@wisc.edu Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53706, USA.
I thank Sam Brown, Amanda Hammond, Ariel La, Chris Kozak, and Callie Porter-Brown for help with data collection and Emily Ward and Chuck Kalish for helpful discussion.
This work was partially supported by the National Science Foundation (NSF) [grant number BCS#1331293 to the author].

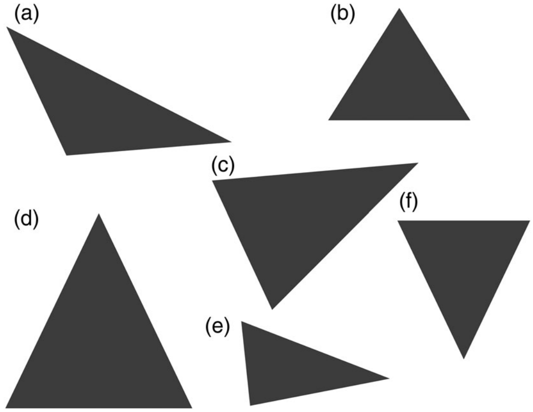

Figure 1 Various triangles. All of them are perfectly triangular geometrically, but not psychologically. A display like this makes an effective demonstration of typicality effects. In an in-class "clicker" poll of 82 students shown Options a–e, 70% chose b as the "best triangle", and 28% chose "d".

be confused by a request to point to the "upside-down triangle". Yet, they do not and are not.

Not only do people almost unanimously judge horizontally oriented equilateral and isosceles triangles to be *better* triangles than scalene and obliquely oriented ones (making remarks like "sexy little equilateral triangles are the trianglest"), up to 15% of educated adults claim that scalene triangles—especially when oriented at an oblique angle—are not triangles at all (Lupyan, 2013, p. 630)!

These observations lead to the paradox of the universal triangle: On the one hand, people can learn facts that apply to *all* triangles—for example, "the interior angles of triangles always add up to 180°" —and perform far more abstract feats like constructing formal proofs. On the other hand, as we shall see, true abstraction is elusive, and thoughts about triangles are consistently influenced by (what should be) irrelevant factors and are systematically affected by subtle differences in eliciting cues.

In Experiments 1A–1B, I show that formally equivalent ways of asking people to draw a triangle lead people to draw systematically different shapes. In Experiment 2, I show that such differences are not pragmatic in nature, arising also in a speeded recognition task with objectively correct responses. Experiment 3 shows that subtle differences in cues affect judgment of simple visual properties. Experiment 4 extends the findings to higher level reasoning, showing that cueing the same knowledge in different ways affects category-based inferences.

Together, the studies suggest that the mental representations constituting the concept of a triangle differ depending on *how* it is activated. Despite our apparent facility to converse about abstractions, the results show that mental representations of abstract entities retain a high level of specificity and context-dependence (cf. Laurence & Margolis, 1999).

Rationale

The experiments that follow test three claims. First, I examine whether despite "knowing" what a triangle is, people's mental representations of triangles may be considerably more specific than suggested by the abstract definitions that people provide. Second, contrary to the idea that the representation of a formal concept like triangle is based on a stable "core" (e.g., Armstrong et al., 1983; Osherson & Smith, 1981), I examine whether the representation is altered by subtle differences in the eliciting context (e.g., "triangle" versus "three-sided polygon"). Third, I test the prediction, derived from past work (e.g., Lupyan, 2012b) that using verbal category labels as eliciting cues—here, the word "triangle"—induces a more prototypical representation of a triangle and greater stability from one time to the next, and one individual to another, than when definitionally identical eliciting cues (e.g., three-sided polygon) are used.

The current studies focus on triangles as a *microdomain*. Triangles serve as a test case for the more general claim that *even* categories with easily enumerated membership conditions—of which triangles are a paradigmatic case—have a protean nature (Barsalou, 1987; Casasanto & Lupyan, 2014; Evans, 2009; Hampton, 2006, 2007).

Before proceeding, it is necessary to make explicit a basic assumption. Empirical study of concepts relies on *behaviour*. The various paradigms that psychologists use to study concepts/semantics/word-meanings—recognition, inference, generalization, explanation, and so on—all assume a pragmatist position: We can understand concepts by the effect they have on our behaviour (James, 1907). This assumption is not shared by all.[1] The results I present and the conclusions I draw

[1]The following exchange occurred at the closing of a round-table debate on the subject of "Fodor's Puzzle" of Concept Learning at the 2005 Meeting of the Cognitive Science Society:

Andy Clark: I think Jerry Fodor is getting away with something by slipping it in very early into the argument. And what comes in very early is may thus be unsatisfying to those who maintain a strict distinction between competence (where "true" concepts supposedly reside) and performance (the mechanisms by which concepts affect behaviour). This point is further elaborated in the section *Arguments from competence versus performance*.

Category labels as eliciting cues

On the classic cognitivist position, concepts are constituted by amodal and abstract (sometimes called "symbolic") states. And according to some, people then "invent words that label their concepts" (Li & Gleitman, 2002, p. 266). On such a view, the very question of how language impacts our conceptual structure becomes ill-posed (see Gleitman & Papafragou, 2005; Gomila, 2011; Pinker, 1994, 2007; see Lupyan, in press for discussion).

The idea that concepts are representational states with a format distinct from perception has been steadily losing ground to an alternative. According to what is often called embodiment or "simulation", concepts—even abstract ones—are grounded in sensorimotor and affective states (Barsalou, 1999; Dove, 2009; Simmons et al., 2007; Zwaan & Taylor, 2006; see Kemmerer, 2010; Pecher, 2013, for overviews). Consider that decoding conceptual content from perceptual states in functional magnetic resonance imaging (fMRI; e.g., see Çukur, Nishimoto, Huth, & Gallant, 2013; Harel, Kravitz, & Baker, 2014, for two recent examples) should not be possible if conceptual content were encoded in a purely abstract format, but fully expected if these states have a distributed representation grounded in perceptual states (see also Kiefer & Pulvermüller, 2012).

The idea that words simply label pre-existing concepts has also been challenged. Converging evidence points to the causal role of language in forming conceptual representations in both children (e.g., Althaus & Mareschal, 2014; Balaban & Waxman, 1997; Casasola, 2005; Dessalegn & Landau, 2008; Loewenstein & Gentner, 2005; Lupyan & Thompson-Schill, 2012), and adults (Barrett, Lindquist, & Gendron, 2007; Lupyan & Casasanto, 2015; Lupyan, Rakison, & McClelland, 2007). The causal involvement of language in categorization is further supported by findings that language impairments such as aphasia impair categorization in nonverbal contexts (Gainotti, 2014; Lupyan & Mirman, 2013, for review). These impairments can be mimicked in healthy adults placed under conditions of verbal interference (Lupyan, 2009) and noninvasive neural stimulation of cortical regions classically associated with language (Lupyan, Mirman, Hamilton, & Thompson-Schill, 2012; Perry & Lupyan, 2014).

I am going to suggest that labels help to transform highly specific representations to ones that are typically associated with functions like general recognition and categorical inference that concepts are thought to fulfil. Let us consider a familiar category like DOG. Assume that thoughts about dogs occur at different levels of abstraction (see also Zwaan, 2014). Direct perceptual experiences fall on the concrete end of this concrete-to-abstract continuum. Any dog we see, smell, hear, touch, is a *specific* dog, having a specific (even if unknown) size, weight, colour, and so on. Anytime we step into the river of perceptual experience, it is a different river (see Connell & Lynott, 2014, for related discussion). This specificity of perceptual experiences is not simply a theoretical construct. Insofar as it is possible to measure the neural state corresponding to the perceptual experience of a poodle, it will tend to be more similar (closer in neural state space) to the state corresponding to perceptual experiences of other poodles as compared to experiences involving German shepherds (see General Discussion for further discussion). For certain tasks, this variability matters. To assess whether a given dog is threatening, it is not enough to simply classify it as a dog. Similarly, in searching for a specific dog, one benefits from precise information about what kind of dog it is. Other contexts call for more abstraction—for example, the dog being referred to by the question "do you own a dog?" or when informed that dogs have tails. How does a representation of such an abstract dog come to be activated? The answer may lie in language.

---

this rejection of pragmatism. The whole argument is just going to go through flawlessly and beautifully, if you agree that there is no essential connection between grasping a concept and being able to do things in the world. There's all sorts of ways to act in the world and learning to do things in the world, which are completely immune to the worries that Jerry Fodor is raising. So that the only way that this argument goes through is if you buy something that is very counterintuitive, that is to say, that there is no essential relation between the concepts that you grasp, and the things that you can actually do in the world. And that's the price that Jerry has to pay for the very lovely argument.

Jerry Fodor: There's no price for me. I'm a Cartesian!

The full transcript can be found at the link below, courtesy of Jesse Snedeker. http://www.wjh.harvard.edu/~lds/pdfs/Niyogi_Snedeker-2005.pdf

While a dog bark is a bark of a specific dog (i.e., carrying with it iconic content), the word "dog" can be used to denote *any* dog. This potential for abstractness that distinguish it from other eliciting cues. The utterance "hey, a dog" leaves unspecified the dog in question in a way that is impossible for any direct perceptual experience of a dog (Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2015). Representations activated by verbal labels (e.g., the word "dog") are therefore predicted to be different from representations activated by nonverbal cues such as a dog bark (even when they are deemed equally informative, familiar, and unambiguous). This is indeed the case. As shown by Lupyan and Thompson-Schill (2012), verbal cues activate familiar concepts more effectively than nonverbal cues, both for familiar and for newly learned categories. Subsequent findings by Edmiston and Lupyan (2015) showed that these differences extended to early eye movements: Following nonverbal auditory cues (dog barks, engine revs, baby cries) people tended to look at the most likely source of the sound even when it was task irrelevant. In contrast, verbal labels led to more categorical patterns of looking times consistent with labels activating a more abstract category representation. Using event-related potentials (ERPs), Boutonnet and Lupyan showed that the label advantage in object categorization arises at the perceptual level. Labels modulate the P100—an early visual response occurring within 100 ms of visual input (Boutonnet & Lupyan, 2015). This result supports the idea that labels are especially effective in activating a more categorical perceptual state.

To further illustrate the role of verbal category labels in forming more abstract perceptual states, consider a very simple neural network model in which a verbal cue ("dog") is associated over time with various dogs. Because the perceptual and motor experiences relating to dogs vary from one time to the next, the associations formed between these experiences and the label become progressively *dissociated* with features of the experiences that vary arbitrarily and *associated* with features that are most diagnostic of the category. When the label is then used as an eliciting cue, what will be activated is a kind of *idealized perceptual state*[2] that highlights the features that best distinguish members and non-members of the named category (see Lupyan, 2012b, for discussion and a computational model). Labels are especially well suited for activating such a state for two reasons. First, they are categorical (i.e., there are big smiles and half-smiles, but the word "smile" is an all-or-none phenomenon; Burling, 1993). Second, unlike all other perceptual cues, labels tend to be *unmotivated* (Edmiston & Lupyan, 2015; Lupyan & Bergen, 2015). Consider the concept of a guitar. Any guitar we can interact with is necessarily a *particular* guitar. Likewise, a picture or a sound of a guitar is specific; it cannot be ambivalent about whether it is, for example, electric or acoustic. Language, on the other hand, can be. Therefore the mental state activated by the word "guitar" may *approximate* a more categorical/idealized perceptual state—a state that is more difficult to achieve without the category name and would perhaps be impossible to achieve in the absence of language entirely.

Specific predictions

We can now relate the view of concepts and category names sketched above to the paradox of the universal triangle and the specific experiments that follow. First, if people represent triangles abstractly, it should not be difficult to get people to treat all triangles equivalently. While some triangles may be judged as more typical, and categorized more quickly (Armstrong et al., 1983), such responses should be easily overcome and should not figure into people's reasoning about triangles (cf. Lupyan, 2013). If people represent triangles as abstractions it should also make little difference *how* one accesses that representation. Thinking of a triangle and thinking of a three-sided polygon is, on this view, the same thing.

If, however, thinking about triangles involves thinking about specific types of triangles, two different predictions follow. First, even in tasks that call for an abstract representation, people should show persistent typicality effects. Second, these typicality effects should be exaggerated when the eliciting cue is a common verbal category label. This latter prediction is tested in a strong way in the experiments that follow. Insofar as a category label (e.g., "triangle") activates a more prototypical/

[2]I will sometimes refer to this state as a *prototype*. On the present formulation, a prototype is not a "thing" but rather a distributed neural activation pattern that (by definition) overlaps more with patterns elicited by more typical than with those elicited by less typical exemplars (see General Discussion).

idealized perceptual state of triangles, the representation of triangles as elicited by the word "triangle" should induce a stronger typicality/canonicality effect even when compared to pragmatically equivalent cues like "three-sided figure" and to logically coextensive cues like "three-sided polygon" (even though all of these cues are, strictly speaking, verbal).[3]

The specific prediction is that verbal labels, especially basic-level category name, ought to elicit conceptual information in a different way from what are ostensibly the "same" concepts elicited without using the category name. While previous work supported this conclusion by examining common categories like dogs and guitars (Boutonnet & Lupyan, 2015; Edmiston & Lupyan, 2013; Lupyan & Spivey, 2010; Lupyan & Thompson-Schill, 2012), the current studies are the first to examine whether a formal category—triangles—is activated differently by formally equivalent cues.

On the account outlined above, the paradox of the universal triangle is resolved not by positing abstract symbols, but by taking the conceptual representation of triangles to be an idealized perceptual state, which can be flexibly transformed by context to accommodate different kinds of triangles.

Defining a triangle

I begin by examining whether people can articulate the normative definition of a triangle (a three-sided polygon). One-hundred and three participants were recruited from Amazon Mechanical Turk and were simply asked to define a triangle. Three research assistants unaware of the hypotheses in question classified the definitions into six non-mutually exclusive categories. Modal responses were used in case of disagreements.

1. Lines/sides, e.g., "three-sided shape", "three-sided figure".
2. Angles/corners/points, e.g., "shape that has 3 corners".
3. Closure, e.g., "three sides, three angles, closed object".
4. Straightness of sides, e.g., "3 straight lines that all connect at their ends".
5. Relative angle size/side length, e.g., "A shape with three equal sides".
6. Orientation, e.g., "A shape with 3 points, one at the top, two at the bottom".

The breakdown of the responses is shown in Figure 2. As evident, the majority of people defined a triangle as a figure/shape having three sides and/or three angles. Of the 92 people who mentioned lines/sides, 25 also mentioned angles. Of the 35 who mentioned angles, 10 only mentioned lines/sides. Importantly, almost no one made any reference to relative side ratios or orientation. If one takes the definitions that people provide as reflecting an underlying conceptual representation, one might conclude that the concept of triangle is indeed quite abstract. One might conclude that it is a universal triangle. The experiments that follow challenge this conclusion.

EXPERIMENT 1A: DRAWING TRIANGLES AND *TRIANGLE* TRIANGLES

The foray of psychology into introspection (Danziger, 1980) has taught us that actions speak louder than words. If someone insisted that all cars have four wheels and yet consistently identified the Elio Motors three-wheeler as a car, we would conclude that their concept of a car generalizes to three-wheelers.[4] Similarly, the behaviour of someone who states that a triangle is a shape with three sides, yet draws only isosceles triangles, suggests that the representation generating this behaviour might be partial to certain kinds of triangles. If we moreover find that people drew systematically *different* triangles depending on whether they were asked to draw a triangle or a figure with three sides—cues that

[3]Of course our participants know that three-sided polygons are triangles and, as discussed below, tend to interpret phrases like "three-sided figure" as denoting a triangle. But reflective knowledge and the ability to strategically translate from the first phrase to the second does not mean that the two will elicit the same representational state, just as knowing that dogs bark does not mean that a bark and the word "dog" activate the same representational state (Lupyan & Thompson-Schill, 2012). That all participants know and use the term "triangle" make the experiments that follow an especially strong test of the stated hypotheses because any observed differences exist *despite* participants' ability to translate between these functionally equivalent expressions. If these putatively coextensive cues activate detectibly different representations, then one may infer the differences to be even larger in the case when a suitable category name is available and in cases where a name is not available.

[4]Indeed, Elio Motors faces a fascinating legislative situation: Their vehicle is federally classified as a motorcycle because it does not have four wheels, allowing it to meet more lenient emission standards set for motorcycles. At the same time, the company has been (successfully) lobbying to obtain an exemption on motorcycle licence or helmet use requirements for the drivers of the vehicle on the basis that it is really a car.

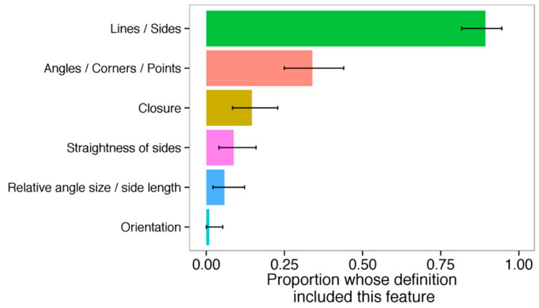

Figure 2 Proportion of participants ($n = 103$) whose definitions of "triangle" mentioned a given feature. Numbers do not add to 1 because the features were not mutually exclusive. Error bars depict binomial confidence intervals. To view this figure in colour, please visit the online version of this Journal.

should, on the traditional account, activate the very same abstraction—we might conclude that the concept of triangle is not as abstract or as stable as its definition suggests. Drawing is, of course, only one way to probe a conceptual representation, but it is where we begin.

Method

Participants

Seventy-three participants from the University of Wisconsin-Madison subject pool were recruited in exchange for course credit. A separate group of 20 participants were subsequently recruited from Amazon Mechanical Turk to make geometric judgments of each drawing.

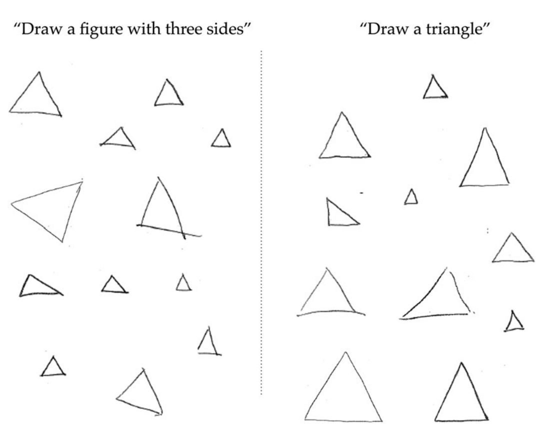

Figure 3 A sampling of triangles drawn in Experiment 1A in response to the two prompts. To view this figure in colour, please visit the online version of this Journal.

Materials and procedure

Drawing task. After completing a paper-based debriefing questionnaire of an unrelated study, participants were asked to draw a shape. About half ($n = 33$) read an instruction to "Please draw a triangle". The remainder ($n = 40$) received the instruction to "Please draw a figure having three sides".

Subjective judgment task. Participants were presented with each of the 69 drawn triangles, one at a time, in random order, and were asked to determine whether it had 0, 2, or 3 equal sides and to determine whether the base looked flat (yes, maybe, no).

Results

All but four participants drew a single triangle. Of the four who did not (all from the three-sided figure condition), three drew three-dimensional shapes with triangular bases, and one drew two triangles. These participants were excluded from further analysis.

The results were analysed in two ways: (a) through objective quantification of the geometry of the drawn shapes, and (b) through subjective judgments of the shape geometry.

Objective analysis

The dimensions of each triangle were quantified by locating the vertices of each shape in a digitized drawing. Measures of all sides and angles were then automatically computed using simple trigonometry. Based on the prediction that the different instructions would cause people to draw triangles differing in orientation and/or symmetry, three primary measures of interest were computed:

1. Deviation from the equilateral was defined as the mean absolute difference of each angle from 60°

$$\frac{\sum_{\theta=1}^{3} |\theta_i - 60|}{3}$$

2. Deviation from the isosceles was defined as the minimum absolute difference of each pair of angles:

$$\min \begin{cases} |\theta_1 - \theta_2| \\ |\theta_2 - \theta_3| \\ |\theta_1 - \theta_3| \end{cases}$$

3. Orientation of the drawing was measured as the angle of the base relative to the horizontal.

Two additional measures were added post hoc: the absolute difference of the two base angles where the base was defined as the most horizontal side (a small difference is indicative of triangles that look like arrows; e.g., Figures 1c, 1f), and size of the drawing (average length of the sides).

Sample drawings are shown in Figure 3. Logistic regression was used to predict the likelihood of each drawing being generated from each of the two prompts based on the three predictors listed above. This was preferred to the conventional linear regression/analysis of variance (ANOVA) analyses that predict differences in the shapes as a function of condition because it allowed for simultaneous modelling of multiple measures and because the distributions of the geometric measures were highly non-normal, making them ill-suited as dependent variables in a linear regression analysis.

Overall, people's drawings were very far from a random sampling of all possible triangles. The drawings deviated from the equilateral by an average of 8.5° ($SD = 5.0°$), from isosceles by 4.8° ($SD = 3.7°$), and from the horizontal by 5.6° ($SD = 8.0°$). So, overall, people's drawings were highly constrained to be roughly equilateral/isosceles and were quite close to having horizontal bases.

The data were analysed using a logistic regression predicting the drawing condition from the geometric factors summarized above. Triangles drawn to the *triangle* prompt deviated from the equilateral to the same degree as triangles drawn to the *three sides* prompt, $z < 1$, but were more likely to be isosceles, $b = -0.22$, 95% confidence interval (CI) = [−0.43, −0.05], $z = -2.28$, $p = .02$, and were more horizontally oriented, $b = -0.18$, 95% CI = [−0.36, 0.057], $z = -2.35$, $p = .02$, than triangles drawn to the *three sides* prompt. These coefficients can be interpreted as log odds: For every degree a drawing deviated from the isosceles, it was $e^{-.22} = .80$ times (i.e., 20%) less likely to be drawn to the triangle prompt. For every degree a drawing deviated from the canonical horizontal orientation, it was $e^{-.18} = .84$ (16%) less likely to be drawn to the *triangle* prompt.

Expressing the analysis above in a slightly different form, if we define an isosceles triangle as a triangle with two angles deviating by at most 4 degrees, 76% of the triangles drawn to the *triangle* prompt are isosceles, but only 61% of triangles drawn to the *three sides* prompt are isosceles. If we define canonically oriented triangles as those having orientations of less than 4° from the horizontal, 82% of the triangles drawn to the triangle prompt were canonically oriented. Only 50% of the triangles drawn to the *three sides* prompt were canonically oriented.

Triangles drawn to the *triangle* prompt had a slightly, but non-significantly, smaller absolute difference between the two base angles ($M_{triangle} = 6.9°$, $M_{three\ \ sides} = 10.3°$), $b = -0.04$, 95% CI = [−0.113, 0.010], $z = -1.45$, $p = .15$. This result, although not significant, is in the predicted direction, as identical base angles correspond to a more canonical isosceles triangle. Triangles drawn to the *triangle* prompt were also marginally larger with an average side length of 205 pixels, compared to an average side length of 150 pixels for those drawn to the *three-sided* prompt, $b = 0.006$, 95% CI = [0.0005, 0.0120], $z = 1.92$, $p = .06$.

Subjective judgments

Objective differences in shape do not necessarily translate to perceived (i.e., subjective) differences. It is therefore informative to find out whether subjective judgments of side-equality and orientation made by naive participants could distinguish the shapes drawn to the two prompts.

Not surprisingly, people's categorical judgments of base orientations were highly correlated with actual base angles, $r = .76$, $p < .001$. People's likelihood of judging the triangle as equilateral was well predicted by its actual deviation from the equilateral, $r = -.60$, $p < .001$, and the likelihood of judging it as isosceles or scalene was predicted by its deviation from the isosceles, $r = .41$, $p < .001$. This agreement was confirmed by linear mixed-effect models on the non-aggregated responses.

To test whether people's judgments differentiated triangles drawn to the two prompts, the discrete responses were analysed using mixed-effects logistic regression predicting the original drawing condition from participant judgments (fixed factor). The regression included the subject as a random intercept. Triangles drawn to the *triangle* prompt were judged to have more equal sides, $b = 0.10$, 95% CI = [0.003, 0.205], $z = 2.10$, $p = .04$, and were perceived to have flatter (more horizontal) bases, $b = -0.65$, 95% CI = [−0.78, −0.52], $z = -9.81$, $p < .001$.

Discussion of Experiment 1A

People drew more canonical triangles when asked to draw a triangle than when asked to draw a figure with three sides. Although it is clear from the results that people took the instruction to draw a three-sided shape as implying a closed two-dimensional figure with three sides—that is, a three-sided polygon—when directly queried as to whether all "three-sided shapes are triangles", 12 of 52 (23%) of newly recruited adults responded that this was not the case, and when asked to elaborate, many cited shapes with curved sides as a three-sided shape that was not a triangle, suggesting that for some people "three-sided figure" is not co-extensive with three-sided polygon. This is a likely explanation for why some people (10%) in the *three sides* condition drew non-triangles, but does not explain why the remaining 90% drew triangles that were systematically more scalene and oblique than those drawn to the instruction to draw a "triangle".

Still, to address the concern that a three-sided figure and a triangle are not truly coextensive, I ran Experiment 1B, which sought to make the two instructions even more similar.

EXPERIMENT 1B: TRULY COEXTENSIVE CUES PRODUCE DIFFERENT DRAWINGS

The primary aim of Experiment 1B was to pit a "draw a triangle" condition against a definitional prompt that cannot be construed to mean anything but a two-dimensional three-sided polygon. A secondary aim was to extend Experiment 1A to a more diverse sample and more variable testing conditions.

Method

Participants

Two hundred adults of varying ages, residing in the US, were recruited from Amazon Mechanical Turk for the drawing task. A separate group ($n = 40$) were subsequently recruited to rate each drawing on typicality.

Materials and procedure

Drawing task. Participants were presented with a blank 320×240-pixel canvas and were asked to draw a figure using their mouse. A random half ($n = 100$) were asked to "draw a triangle". The other half were asked to "draw a three-sided polygon". Pilot testing revealed that many people were unfamiliar with the term "polygon"; the instruction was therefore followed by a definition: "A polygon is a closed shape with straight sides".

Typicality judgments. Participants were simply asked to rate each drawing on its typicality as a triangle on a 7-point Likert scale from 1 (very unusual) to 7 (very typical). Each participant was shown 50 shapes from a quasi-randomly chosen subset of the shapes drawn in the task above such that half of these were drawn to the *triangle* prompt and half drawn to the *three-sided polygon* prompt. Shape presentation was counterbalanced such that each shape was rated by 10 participants.

Results

Objective analysis

A series of logistic regressions identical to those used in Experiment 1A indicated that the triangles drawn to the *triangle* and *three-sided polygon* conditions did not differ reliably in deviation from the equilateral or the isosceles, $zs < 1$, but differed in orientation: $b = -0.11$, 95% CI = [−0.19, −0.05], $z = -3.14$, $p = .002$. *Triangles* had mean deviations of 3.2° off horizontal. *Three-sided polygons* had mean deviations of 8.1° off horizontal. Although the triangles did not differ on the primary metric used to measure deviation from the isosceles, there was a reliable difference between the conditions in the absolute difference between the two base angles, with *triangles* having a smaller difference than *three-sided polygons*, $b = -0.03$, 95% CI = [−0.06, −0.01], $z = -2.59$, $p = .01$. This difference remained significant when the deviations from the equilateral and from the isosceles were partialled out, $b = -0.04$, 95% CI = [−0.07, −0.01], $z = -2.20$, $p = .03$, indicating that triangles drawn to the *triangle* prompt were likely to resemble vertically pointing arrows.

Triangles drawn to the *triangle* prompt were also significantly larger, having average sides of 128 pixels compared to *three-sided polygon* drawings (115 pixels), $b = 0.008$, 95% CI = [0.001, 0.015], $z = 2.10$, $p = .04$.

Subjective typicality judgments

Rather than have participants judge the geometry of the triangles as in Experiment 1A, I sought to test more directly the hypothesis that people drew more typical triangles when prompted to draw a triangle than when prompted to draw a three-sided polygon. The reason for focusing on typicality judgments is that such ratings are highly predictive of classification likelihood. Not only do people rate some triangles as being less typical than other triangles, but such typicality ratings are highly predictive of classifying the shapes as triangles (Lupyan, 2013). Finding that people think that *triangle* triangles are better triangles than *three-sided polygon* triangles would provide convergent evidence that representations cued by the putatively identical instructions are, in fact, systematically different.

Shapes drawn to the *triangle* prompt were rated as more typical than shapes drawn to the *three-sided polygon* prompt ($M_{\text{triangle}} = 4.6$; $M_{\text{three-sided polygon}} = 4.2$), a difference that was highly reliable by a linear mixed-effects model predicting typicality from the drawing prompt with subject and item random intercepts, $b = 0.41$, 95% CI = [0.11, 0.77], $t = 2.66$, $p = .008$.

People's typicality judgments were based solely on the visual form of the drawings. The raters had no idea that the figures were drawn to different prompts. The analyses below test whether people's typicality ratings were predicted by the orientation and deviations from the isosceles and equilateral. Because people tended to use different ends of the scale, in the analyses that follow, participants' responses were normalized (*z*-scored).

Typicality was negatively correlated with deviations from the canonical (horizontal) orientation ($r = -.23$, $p = .001$), with deviation from equilateral ($r = -.26$, $p < .001$), and with deviation from isosceles ($r = -.20$, $p = .003$). Combining the measures into a single regression model with subject and item random intercepts revealed that typicality was simultaneously predicted by the base angle and either the deviation from the isosceles, $b = -0.02$, 95% CI = [−0.03,−0.0003], $t = -2.00$, $p = .047$, or deviation from the equilateral, $b = -0.02$, 95% CI = [−0.03, −0.004], $t = -2.64$, $p = .009$ (the deviation from the isosceles measure did not account for unique variance once deviation from the equilateral was included in the model). Typicality was also strongly predicted by the absolute difference between the two base angles, $b = -0.02$, 95% CI = [−0.024, −0.010], $t = -4.67$, $p < .001$: arrow-like triangles were judged to be more typical.

People judged triangles drawn to the *triangle* prompt as more typical than triangles drawn to the *three-sided polygon* prompt. This difference provides converging evidence that the instruction to "draw a triangle" did not just cause people to draw systematically different triangles—it caused people to draw more typical, "better" triangles as confirmed by people's typicality judgments being systematically higher for canonically oriented and equilateral/isosceles triangles/arrow-like triangles.

Discussion of Experiments 1A–1B

When asked to draw a triangle, people drew triangles. When asked to draw a shape with three sides, or when asked to draw a three-sided polygon, people also drew triangles. But the *kind* of triangles people drew to these putatively identical instructions were systematically different. Despite defining triangles as three-sided figures, shapes with three angles, and other abstractions, people's depictions of triangles were highly constrained to those that past work (Lupyan, 2013) has shown to the most typical. More importantly, this typicality bias was considerably stronger when people were asked to draw a *triangle* than when they were asked to draw a figure with three sides or a three-sided polygon. *Triangles* were more symmetric and had more horizontally oriented bases (and were somewhat larger in size) than *three-sided figures* and *three-sided polygons.* These differences were reflected in objective geometric measurements, subjective appearance, and typicality judgment of naive raters.

While the *definitions* of triangles that people provide clearly abstract away perceptual details and dutifully report the necessary-and-sufficient conditions, the representations brought to bear on this production task appear to be both more specific and highly sensitive to the eliciting cue. The present finding that the category name activates a more canonical/typical representation of triangles supports and extends previously reported results that showed category names like "dog" to activate a more typical form of the concept than equally informative nonverbal cues (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012; see also Holmes & Wolff, 2013, for a related finding in the domain of spatial relations). The results of the typicality rating task in Experiment 1B further corroborate this effect, showing that people think that *triangle* triangles are more typical triangles than *three-sided polygon* triangles. This typicality difference has an important consequence for *alignment*. Because there are fewer ways to be typical than atypical, representations depicting typical forms will tend to be more similar to one another (see also Experiment 2).

Tentatively then, the results support the account that (a) the representations of even a formally defined category like triangle reflect formally irrelevant, but perceptually relevant, properties; and (b), category names help to form a kind of idealized perceptual state—a prototype of sorts. These results are difficult to reconcile with accounts positing that people's representations of triangles reflect definitional properties and show that logically coextensive descriptions do not activate the same representational state, at least in the context of a drawing task.

EXPERIMENT 2: CATEGORY LABELS INDUCE A TYPICALITY GRADIENT IN SPEEDED RECOGNITION

One interpretation of the differences in the drawings produced to the different prompts used in Experiments 1A–1B is that they are pragmatic in origin. Perhaps participants in the three-sides/three-sided polygon conditions reasoned that had the experimenter wanted them to draw a "regular" triangle, they would have just said so, and the instruction to draw, for example, a "figure with three sides" implied drawing not just a triangle, but a weird triangle. Experiment 2 sought to address this concern while also testing the predictions laid out in the introduction in the context of a within-subject speeded recognition task in which participants were variously cued with the word "triangle" or the phrase "three sides". If the two cues elicit a systematically different representation with "triangle" activating a more canonical state, we ought to see these differences reflected in people's recognition times despite identical pragmatics.

Method

Participants

Eleven University of Wisconsin-Madison undergraduate students participated for course credit. One participant was eliminated for having chance-level performance on "no" trials and was replaced.

Materials and procedure

Participants completed a speeded cue–picture verification task. On each trial, participants heard an auditory cue and then saw either a triangle or a rectangle. Their task was simply to press a "yes" button if the cue matched the shape and the "no" button if they did not match.

On a random half of the trials, the auditory cue was a category name: "triangle" or "rectangle". On the remaining trials the cue was "three sides" or "four sides", which, in the context of the study, was completely co-extensive with triangles and rectangles (i.e., all "three-sided shapes" were three-sided polygons).

The auditory cue correctly predicted the upcoming shape 80% of the time (i.e., 80% of trials were "yes" trials; see full trial structure below). Each trial began with a 750-ms fixation period followed by the auditory cue. The shape was presented 750 ms following cue offset and remained visible until a response was made. Responses were followed by auditory feedback: a bleep for correct responses and a buzz for incorrect responses.

The polygons were green, presented on a black background, and subtended approximately 4° to 11° of visual angle depending on their geometry. They appeared at one of the vertices of an imaginary rectangle subtending approximately 25° in width and 8° in height. There were two sub-types of rectangles: wide (aspect ratios 1.1 to 1.9) and tall (aspect ratios 0.1 to 0.9). There were three sub-types of triangles: equilateral, isosceles, and scalene. Finally, the polygons were presented at various orientations: one third of the shapes had horizontally oriented bases; the remaining two thirds were presented in an oblique orientations (−20° to 20° off the horizontal in multiples of 10°).

Because triangles were the theoretically relevant category, they comprised the majority (80%) of the trials. The full trial structure—all factors within subject—was as follows: Cue–Shape Match (yes/no) × Cue Type (label or descriptor) × Shape Category (rectangle or triangle) × Shape Sub-Type ("wide" vs. "tall" rectangles; equilateral, isosceles, or scalene triangles) × Orientation of the Shape. Each participant completed a total of 360 trials.

Results

Overall accuracy was 99% for "yes" trials and 95% for "no" trials, with equivalent accuracy across all conditions. The analyses focus on response times (RTs) of correct responses: responding "yes" after hearing "triangle" or "three sides" followed by seeing a triangle. Errors and trials below 200 ms or above 1000 ms were excluded (6.3% of total). The data were analysed using a linear mixed-effects model with cue condition (triangle vs. three sides), orientation (canonical vs. non-canonical), and triangle type (equilateral, isosceles, scalene) as fixed factors, and subject as a random intercept.[5] Triangle type was coded as an ordered factor (−.5 = equilateral; 0 = isosceles; .5 = scalene) in accordance with previous findings that equilateral triangles are more typical than isosceles, which are in turn more typical than scalene (Lupyan, 2013). All other

[5]The data were aggregated prior to analysis to have a single RT mean for each Subject × triangle-type (equilateral/isosceles/scalene) × Orientation (canonical vs. non-canonical) to ensure that each trial type was weighed equally by the regression.

predictors were centred (sum-coded) as well. People recognized more typical triangles more quickly as shown by a linear effect of triangle type, $b = 16.7$ ms, 95% CI = [6.8, 25.5], $t = 3.31$, $p = .001$. This effect was moderated by cue type such that the effect of triangle type was stronger on the trials when people were explicitly cued by the word "triangle" than when people were cued by "three sides", $b = -24.8$ ms, 95% CI = [−44.5, −5.1], $t = -2.46$, $p = .02$. People's responses were independent of the orientation, $t < 1$, and the orientation did not enter into any higher order interactions. There was a reliable three-way interaction between cue, triangle-type, and orientation, $b = -41.25$ ms, 95% CI = [−80.7, −1.82], $t = -2.05$, $p = .04$. The meaning of this interaction is made clear by plotting the canonically oriented and non-canonically oriented trials separately (Figure 4).

For canonically oriented triangles, hearing "triangle" induced a steep typicality gradient, $b = 43.6$ ms, $t = 2.65$, 95% CI = [11.4, 75.8], $p = .02$, which was entirely absent when people instead heard "three sides", $b = -1.79$ ms, $t < 1$. In contrast, when the triangle was presented in a non-canonical orientation the cue no longer interacted with triangle type, $t < 1$ (Figure 4, right panel). Trials on which the triangle was presented in a non-canonical orientation yielded numerically smaller effects of triangle type across both cueing conditions, $b = 12.4$ ms, 95% CI = [1.1, 23.7], $t = 2.15$, $p = .04$. Examining the two cueing conditions separately showed a marginal effect of triangle type on "triangle" trials, $b = 14.4$ ms, 95% CI = [−1.2, 30.19], $t = 1.81$, $p = .09$, and a marginal effect of triangle type on the "three sides" trials, $b = 10.3$ ms, 95% CI = [−1.2, 21.9], $t = 1.75$, $p = .10$.

Discussion

Slightly different ways of cueing what is, on a traditional account, a unitary triangle concept, impacted people's ability to recognize a subsequently presented triangle. A parsimonious interpretation of this type of cueing effect is that shorter recognition RT indicates greater representational overlap between the cue and the target that follows—that is, the cue "readies" the system for processing representationally similar inputs. The finding that hearing "triangle" and "three sides" led to systematically different patterns of recognition times means that the two cues activated people's knowledge of triangles—at least the knowledge necessary for visual recognition—in subtly, but systematically different, ways.

The finding that hearing "triangle" induced a stronger typicality bias, and did so in a dynamic trial-by-trial fashion, supports the claim that what is activated by the word "triangle" is more consistent with specific types of triangles. Rather than accessing an abstract representation that "stands in" for all triangles, the label elicits a representation more similar to some triangles than others.

The flat slope of the RTs on the "three sides" trial in Figure 4, left panel, might be taken to mean that hearing "three sides" activated a representation that abstracted across triangle type. An alternative is that hearing "three sides" activates *different triangles* at *different times* for *different people*, and it is this that produces the flat slope. Supportive evidence for this alternative comes from examining the internal consistency of responses following the two cue types. I calculated an internal consistency measure (Cronbach's alpha) for the six trial types (equilateral, isosceles, Scalene × Canonical/Non-Canonical Orientation) across the participants when cued by "triangle" compared to when cued by "three sides". People's RTs on the "triangle" trials had a high level of consistency across subjects, Cronbach's alpha = .85. In contrast, consistency on the *three sides* trials was much lower, Cronbach's alpha = .36. This difference in consistency was not accompanied by an observable difference in *variability* of RTs for the various trial types between the "triangle" and "three sides" conditions—coefficient of variation$_{\text{triangle}}$ = 26, coefficient of variation$_{\text{three sides}}$ = 24—a nonsignificant difference ($p > .6$).

The relatively high internal consistency of the responses on the *triangle* trials supports the idea that "triangle" activates a more consistent mental state from one subject to another and from one time to the next. In contrast, the relatively low internal consistency of the "three sides" trials suggests that this cue activated more divergent representations between participants as well as within participants from one trial to another.

This greater consistency between representations activated by a category label has been previously discussed in the context of the label-feedback hypothesis (Lupyan, 2012a, 2012b) according to which labels are especially effective in selectively activating the most typical and diagnostic features of the category.[6]

[6]The resulting representation may be called a prototype: a distributed pattern of neural activity that is more overlapping with some depictions of triangles (the typical/canonical ones) than others (the atypical/non-canonical ones).

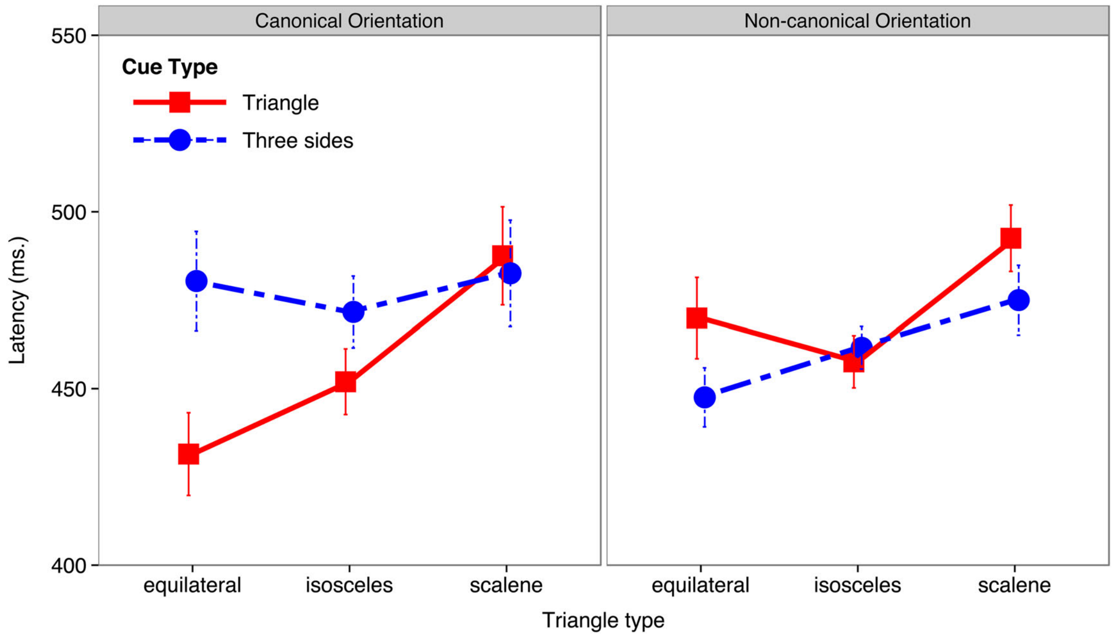

Figure 4 Results of Experiment 2, showing correct response times for the three types of triangles, the two intermixed trial-by-trial cues, and orientations. Error bars show standard errors with between-subject variance removed (Morey, 2008). To view this figure in colour, please visit the online version of this Journal.

Labels produce higher consistency because there are fewer ways to be typical than atypical. Different instantiations of a concept activated by the category label will therefore tend to be more similar to one another (i.e., more aligned) than the representations activated by nonverbal means, or in this case, the actual category label versus a circumlocution (Edmiston & Lupyan, 2015; Lupyan & Thompson-Schill, 2012).

EXPERIMENTS 3A–3B: CALLING A TRIANGLE A "TRIANGLE" AFFECTS SHAPE JUDGMENTS

Experiment 2 showed that hearing "triangle" activated representations of typical triangles, inducing a typicality gradient, while hearing "three sides" a few seconds later—despite signalling participants to do the very same task—activated more idiosyncratic representations. If the category label activates a more typical/canonical representation then such differences may be observed not only in typicality judgments or times to recognize various category exemplars, but even in an unspeeded visual reasoning task with an objectively correct answer.

In Experiment 3A participants were shown triangles of various geometries and were asked to indicate how many equal sides each triangle had. If the "triangle" label activates a more typical representation, it may shift the representation of a currently displayed triangle to a more typical form (Lupyan, 2008a) leading people to systematically misjudge its geometry.

EXPERIMENT 3A

Method

Participants

Eighteen undergraduate students were recruited from the University of Wisconsin-Madison participant pool and were tested online at their leisure.

Materials and procedure

Each participant was shown 156 triangles, one at a time, and was asked to determine whether each had 0, 2, or 3 equal sides.

All 156 triangles shown to each participant were either equilateral (55%) or isosceles (45%). The geometry of the latter was varied parametrically with the apex angle varying from 50° to 59°. The shapes also varied in the orientation of the base (−15° to 15°), with 23% perfectly horizontal. The triangles were ~100 × 100 pixels in size and were visible until a response was made.

The critical manipulation was the exact form of the question presented to the participants. About half ($n = 10$) were asked "How many equal sides does this triangle have?". The remainder ($n = 8$) were asked "How many equal sides does this three-sided figure have?".

Results

The data were analysed using a mixed-effects logistic regression with apex angle and cue type as predictors and subject and item random intercepts. Overall accuracy was 65%, and marginally higher in the *triangle* condition ($M_{\text{triangle}} = 69\%$; $M_{\text{three sides}} = 60\%$), $z = 1.74$, $p = .08$ (Figure 5A). The relatively low accuracy resulted from the difficulty in distinguishing between equilateral and near-equilateral isosceles triangles. Figures 5B–5D show the pattern of errors between the two instructional conditions. Participants only rarely ($M = 7.3\%$) classified triangles as scalene. These errors occurred at equivalent rates between the two instruction conditions, $z < 1$. However, people who were asked to judge the number of equal sides of a *triangle* had a much lower threshold for classifying the triangle as equilateral (i.e., responding "three") and a correspondingly higher threshold for responding "isosceles" than people who were asked to judge the *three-sided figure* (cf. Figures 5C and 5D).

To formally test this claim, I ran a series of mixed-effects logistic regression models. The first two analyses simply show that participants' responses were well behaved: The likelihood of classifying a shape as an isosceles triangle predictably decreased as the triangle approached equilaterality (60° apex), $b = -0.40$, 95% CI = [−0.43, −0.36], $p < .0001$. Conversely, the likelihood of classifying a shape as an equilateral triangle increased as the shape approached equilaterality, $b = 0.58$, 95% CI = [0.53, 0.63], $z = 21$, $p < .0001$.

Of more interest, the likelihood of classifying the shape as an isosceles or an equilateral triangle depended on how the shape was named. When cued to judge *triangles* people were more likely to judge near-equilateral (as well as truly equilateral) triangles as actually equilateral, as reflected in a reliable Cue × Apex Angle interaction, $b = -0.04$, 95% CI = [−0.049, −0.022], $t = -5.29$, $p < .001$. For example, a triangle with an apex between 58° and 59° was 46% likely to be judged as equilateral when cued by *three sides*, but 59% likely to be judged as equilateral when cued by *triangle*. So, simply referring to the shapes as "triangles" led people to judge near-equilateral triangles as more equilateral.

EXPERIMENT 3B

Recall that in Experiment 1A naive participants judged triangles drawn to the "triangle" prompt to have more equal sides and to be more horizontally oriented than triangles drawn to the "three-sided figure" prompt. This difference reflected an objective difference in the geometry of the two groups of triangles. But notice that in simply asking the question, participants were prompted to construe the shape as a "triangle". The results of Experiment 3A suggest that the use of the word "triangle" may have systematically influenced people's judgments of its geometry.

Experiment 3B tested this hypothesis by examining whether judgments of side equality of triangles drawn as part of Experiment 1A were biased by the use of the "triangle" label that was part of the prompt. This was done by referring to the same shapes as "three-sided figures". Finding that referring to the shapes in this different manner led people to judge them to have different number of equal sides would provide further support to the idea that subtly different cues lead to systematic differences in how triangles are represented—specifically, that using a category name elicits a more canonical/prototypical state.

Method

Participants

Twenty participants were recruited from Amazon Mechanical Turk.

Materials and procedure

Participants were asked to judge each of the 69 hand-drawn triangles from Experiment 1A. The judgment task was identical to that in Experiment 1A except instead of being asked "How many equal sides does this triangle have?" participants were asked "How many equal sides does this three-sided figure have?". As in Experiment 1A, participants were also asked to estimate the angle of the base.

Results

The data were analysed using a mixed-effects logistic regression predicting the drawing condition from the same geometric factors as those used in Experiment 1A. The analytic approach was identical to that used in Experiment 1A's *Subjective Judgments* section. Participants' ratings reflected the difference in geometry

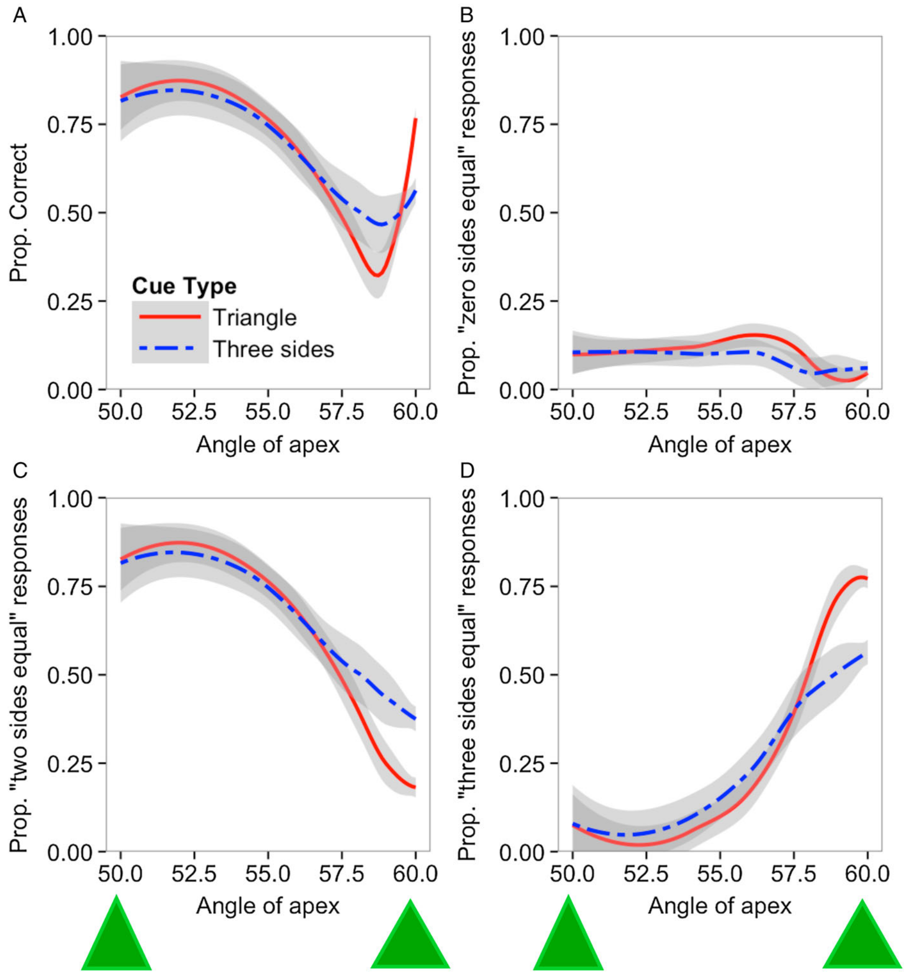

Figure 5 Results of Experiment 3A. Lines show Loess-smoothed regression estimates. (A) Total proportion correct judgments of number of equal sides as a function of apex angle and prompt. (B) Likelihood of judging the triangle to be scalene—never a correct response. (C) Likelihood of judging the triangle as isosceles. (D) Likelihood of judging the triangle as equilateral. The triangles at the bottom show examples of apex angles of 50° and 60°. Shaded error bands show ±1 standard error of the regression estimate. To view this figure in colour, please visit the online version of this Journal.

between triangles drawn to the triangle prompt, which were judged to have more equal sides (i.e., more likely to be isosceles or equilateral), $b = 0.11$, 95% CI = [0.02, 0.21], $z = 2.49$, $p = .02$, and to be more canonically (i.e., horizontally) oriented, $b = -0.54$, 95% CI = [−0.66, −0.42], $z = -8.64$, $p < .001$. Indeed, the by-item correlations between Experiment 1A and this study were quite high: $r = .86$ for judgments of side equality and $r = .96$ for judgments of the horizontality of the base ($ps < .0001$).

Comparing the results of Experiment 1A to the present study showed that people cued by *triangle* were substantially more liberal with their judgments of how many sides were equal. When asked about "triangles", participants' mean response to the "how many sides are equal" question was 2.07. When they were instead asked "How many equal sides does this three-sided figure have?" (the present study), the mean decreased to 1.81. To test whether this difference (i.e., the difference between Experiments 1A and 3B) was reliable, the responses to the number-of-equal-sides question (0, 1, 2) were entered into a mixed-effects linear model with the *divergence from the equilateral* measure (derived in Experiment 1A), the *question prompt* ("triangle" vs. "three-sided figure"), and their interaction as predictors. The model included random subject and item intercepts.

Referring to figures as "triangles" led people to judge them to have more equal sides, $b = 0.27$, 95% CI = [0.07, 0.47], $t = 2.6$, $p = .01$. This effect was moderated by the divergence from the equilateral measure as revealed by a significant interaction term, $b = -0.016$, 95% CI = [−0.019, −0.002], $t = -2.47$, $p = .01$. The difference between referring to a triangle and a three-sided figure was larger for figures close to being equilateral—the very same pattern as that observed in Experiment 3A. Judgments of the orientation of the base were not affected by calling the shape "triangle" or "three-sided figure", $t < 1$.

Discussion

Experiment 3 showed that the same shape referred to by different names leads people to perceive its geometry in systematically different ways. Shapes called "triangles" were judged to be more equilateral than when the same shapes were called "three-sided figures". These results provide yet more evidence that cues pointing to putatively the same TRIANGLE concept actually activate a subtly different representation with consequences for behaviour. Calling a triangle by its name appears to produce an assimilation effect such that triangles close to being equilateral are assimilated into an equilateral form that is—on the present view—activated by the category name.

EXPERIMENT 4: CALLING A TRIANGLE A "TRIANGLE" AFFECTS CATEGORY-BASED INFERENCES

The previous experiments focused on what is widely acknowledged to be a critical function of concepts—explicit classification (Murphy, 2002; Prinz, 2004). But the concept of triangles is useful not only for explicitly classifying shapes as triangles or not, but for enabling inference (e.g., Chin-Parker & Ross, 2004). The inferences may extend to unobserved members of the category such as when we are informed that *all* triangles have angles that add up to 180°. The inferences may also extend to members of other categories. If told that a certain animal leaves triangular footprints, we can make inferences about the presence of the animal based on how well the shape of the footprints matches a triangle.

The previous experiments showed that a category label (here, the word "triangle") induced stronger typicality effects in production (Experiment 1), speeded recognition (Experiment 2), and un-speeded visual judgments (Experiment 3). Experiment 4 investigated whether using the category name similarly affects triangle representations recruited in the context of an inference task—a task not requiring *explicit* classification of individual triangles. If "triangle" activates a more typical representation of triangles, then inferences based on a category of "triangles" should weigh canonical triangles more than inferences based on the category of "three-sided [shapes]".

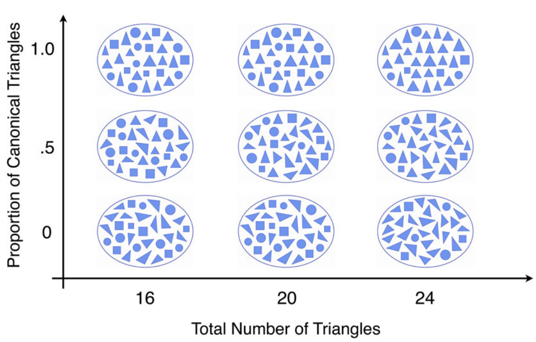

Figure 6 Sample "bacterial colonies" shown to participants in Experiment 4. Displays with 18 and 22 triangles are not shown. To view this figure in colour, please visit the online version of this Journal.

Method

Participants

Twenty-one participants were recruited from Amazon Mechanical Turk. One participant was eliminated for apparently misunderstanding the word "effective" in the instructions to mean "ineffective".

Materials and procedure

Participants were told about a hypothetical drug that was effective against bacteria that had "triangular shapes" or bacteria that had "three sides". Participants were then shown pictures of individual "bacterial colonies" (see Figure 6) and were asked to judge how effective a hypothetical drug would be against each colony. Each picture was 400 × 300 pixels. Participants assigned to the *triangle* condition were told:

> Suppose a drug is discovered that is effective on bacteria that have triangular shapes. How effective do you think the drug will be for a bacterial colony that looks like this?

Participants in the *three sides* condition were given the same instruction except that "triangular shapes" was replaced by "three sides". Everyone responded using a 5-point Likert scale that ranged from 1: "Not effective (will not kill any bacteria)", to 5: "Highly effective (will kill all bacteria)". The stimuli comprising "bacterial colonies" all contained 30 geometric shapes (Figure 6). Of these, between 53% and 80% (16, 18, 20, or 24 shapes) were triangles. Of the triangles, none, half, or all were canonical (horizontally oriented isosceles/equilateral triangles), with the remainder being non-canonical (rotated scalene triangles). Each colony type had three variants, which differed in the positioning of the triangles among the non-triangles. Each participant provided effectiveness estimates for 45 such displays: 5 (triangle number) × 3 (canonicality) × 3 (variants). Each display was visible on the screen until a response was made.

Results

Analyses were conducted using linear mixed-effects models with a subject random intercepts and canonicality (i.e., proportion of canonical triangles) as a random slope (fuller random-effect structures prevented convergence). All predictors were centred. As expected, people's judgments of drug efficacy were strongly predicted by the number of triangles in the display (Figure 7A), $b = 0.13$, 95% CI = [0.12, 0.15], $t = 11.51$, $p < .0001$. Participants in the *triangle* condition thought the drug was more effective overall than participants in the *three sides* condition, $b = 0.38$, 95% CI = [0.10, 0.66], $t = 2.62$, $p = .01$. An interaction between cueing condition and number of triangles, $b = -0.06$, 95% CI = [−0.08, −0.03], $t = -3.97$, $p < .001$, indicated that the difference between cueing conditions was most pronounced for displays with fewer triangles (Figure 7A).

An analysis of canonicality showed a highly reliable quadratic trend: For a given number of triangles, having a mix of canonical and non-canonical triangles (the condition with the most visual diversity) led to lowest drug efficacy judgments, followed by displays with only non-canonical triangles. Displays with all canonical triangles led to the highest perceived efficacy indicating that, overall, participants tended to over-weigh triangles with canonical shapes. The relationship between cueing condition and canonicality was tested using a mixed-effects model with a second-order orthogonal polynomial term (i.e., growth curve analysis; Mirman, 2014). Both the first- and second-order terms were included as random slopes. People's change in perception of the drug's efficacy as a function of canonicality (controlling for total number of triangles) was captured by both the linear, $b = 0.20$, 95% CI = [0.01, 0.39], $t = 2.23$, $p = .03$, and quadratic, $b = 0.28$, 95% CI = [0.15, 0.41], $t = 4.67$, $p = .0001$, components. The interaction with cueing condition was not significant for the linear, $b = 0.15$, 95% CI = [−0.03, 0.32], $t = 1.65$, $p = .11$, or quadratic, $b = 0.11$, 95% CI = [−0.01, 0.22], $t = 1.8$, $p = .09$, component (these interactions became highly significant, $t > 3$, $p < .01$, if the canonicality random slope was omitted).

Although the interaction was not significant with the fuller random-effect structure, planned comparisons showed that the difference between cueing conditions was carried by the displays showing all canonical triangles (Figure 7B). For such displays, people in the *triangle* condition judged the drug to be more effective than people in the *three sides* condition, $b = 0.29$, 95% CI = [0.12, 0.47], $t = 3.27$, $p = .004$.

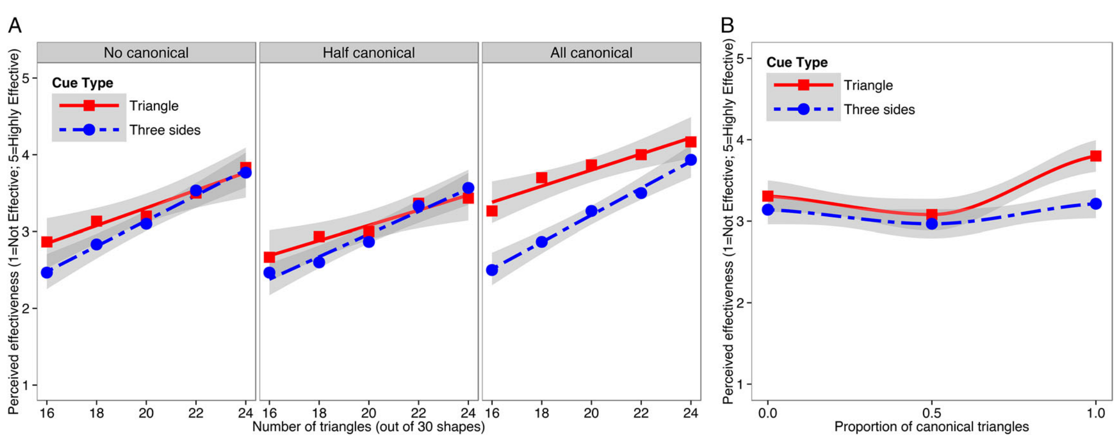

Figure 7 Results of Experiment 4 showing perceived effectiveness of a drug effective against bacteria shaped like triangles vs. bacteria having three-sides as a function of (A) number of triangles, broken down by proportion of triangles that are canonical. (B) Canonicality effect controlling for the number of triangles. Shaded error bands show ±1SE of the regression estimate. To view this figure in colour please visit the online of this Journal

No such difference between conditions was observed when the display contained non-canonical triangles, $b = 0.07$, $t < 1$—a pattern mirroring Experiment 2 in which the difference between the two cueing conditions was largest for the most canonical triangles.

Discussion

The category-based inference task of Experiment 4 corroborates the results of Experiments 1–3 without requiring people to make explicit judgments of individual shapes. Informed that a drug is effective against bacteria "shaped like triangles", people judged it to be more effective than when informed that the drug was effective against bacteria "having three sides". This difference between cueing conditions was especially evident when the displays contained mostly canonical triangles. This result is consistent with the previous findings showing that the word "triangle" appears to selectively activate representations corresponding to such triangles. The interaction between the cue and the number of triangles suggests that the "triangle" label may make the distinction between triangles and non-triangles more salient, especially when the triangles are typical (see Lupyan, 2007, 2008b, for conceptually similar findings using more conventional visual search tasks).

GENERAL DISCUSSION

This paper began by revisiting a three-century-old philosophical dilemma: the paradox of the universal triangle. On the one hand, it seems impossible to conceive a triangle that is not a particular triangle. On the other hand, we are capable of learning facts about all triangles such that they have angles that add up to 180° and—it would seem—are able to converse about triangles in the abstract.

Taking the triangle as a paradigmatic case of a formal concept, I examined its mental representation in the service of a production task (Experiments 1A–1B), speeded recognition (Experiment 2), unspeeded visual judgments (Experiments 3A–3B), and category-based inferences (Experiment 4). The results show that people's representation of triangles in all these tasks were not pure abstractions, but corresponded to canonical/prototypical kinds of triangles (see Barsalou, 1999; Hampton, 2006; Prinz, 2004). And so one reason why we do not flinch at propositions like "upside-down triangle" is that our thoughts about triangles tend to be about *particular* triangles, namely the typical ones.

A consistent finding in the current studies is that explicit use of the category label—the word "triangle"—led to a further exaggeration of typicality effects, as expected if the label is an especially effective way of inducing a more prototypical representation (Lupyan, 2008a, 2008b; Lupyan & Thompson-Schill, 2012). For example, people drew more typical triangles when asked to draw a "triangle" than to draw a "figure with three sides" (Experiment 1A) or a "three-sided polygon" (Experiment 1B). A difference in eliciting a representation of triangles by different means was also found in a speeded-recognition task (Experiment 2) in which hearing the word "triangle" led to more consistent patterns of RTs in recognizing triangles than hearing the equally informative phrase "three sides". Experiment 3 extended this result to unspeeded visual judgments. Calling a triangle a "triangle" led people to judge it as being more equilateral and hence more typical. Experiment 4 extended the claim that using a category label activates a more prototypical state, into the domain of category-based inference.

The finding that even formally equivalent cues—eliciting contexts—like "triangle" and "three-sided polygon", appear to evoke systematically different representations speaks to the flexibility with which people represent a concept even as apparently clear-cut as a triangle. On this perspective, our ability to think of all kinds of triangles speaks not to the abstractness of the underlying representation, but the ability to effectively *modulate* the representation—a process in which language is hypothesized to play a key role (Gomila, 2011; Lupyan, 2012b; Lupyan & Bergen, 2015).

With context playing such a ubiquitous role in the representation even of simple formal categories, attempts to pin down a stable context-independent representation of a concept—*the* concept of triangle—becomes fruitless. In a sense, all concepts are ad hoc concepts (Barsalou, 1983; Casasanto & Lupyan, 2014; Elman, 2011; Spivey, 2008). By activating a (more) prototypical state, a category name helps to stabilize these representations, aligning them into a more similar form from one individual to another and from one time to the next. Such alignment may help to facilitate communication and reasoning from specifics to generics.

This view strongly contrasts with the position taken by classical cognitive science in which the mind is viewed as a symbol-processing device with conceptualization—recognizing a given object or relation as a member of a larger class—implemented via activation of a requisite symbol (e.g., Watson, 1995, for discussion). Instead, the results are consistent with concepts as idealized perceptual states, which verbal category labels are especially well-suited to elicit.

Prototypes for the modern age

I have made numerous references to the term "prototype". This term is used in cognitive science with surprising frequency. For example, it appears, on average, on every fourth page of Margolis and Laurence's *Concepts: Core Readings* (Margolis & Laurence, 1999); about as often as the term "concept" itself. As repeatedly pointed out, even by Rosch herself (1978), the existence of prototype effects do not by themselves constitute a particular processing model for categories. Rather, prototype effects form a bound on a theory of concepts. Theories of concepts that do not naturally accommodate prototype effects are unlikely to be correct.

So, what are prototypes, where do they come from, and what do they have to do with concepts? In a landmark paper, ambitiously titled "On the generation of abstract ideas" Posner and Keele (1968) showed that people exposed to visual patterns derived from a common template became sensitive to the template (what they called a prototype: a pattern that minimized the distance to all the observed exemplars) even though it was never shown to the learners.[7] The findings demonstrated in a straightforward way the naturalness with which people extract regularities from family-resemblance-type categories.

Countless studies have since examined prototype effects (e.g,. Hampton, 2006; Minda & Smith, 2001; Nosofky, 1986; Spencer & Hund, 2002).[8] In many studies, prototypes have been implemented as mathematical formalisms (e.g., Osherson & Smith, 1981, for discussion) without making contact with the stuff of which mental representations are thought to be made: high-dimensional neural firing patterns. Contemporary cognitive neuroscience helps us bridge this gap.

We can think of a prototype as corresponding to a distributed pattern of neural activity that effectively distinguishes between category members and nonmembers. To say that robins and sparrows are more prototypical birds than are penguins (or, in the language of exemplars, to say that robins and sparrows share more features than robins and penguins) simply means that the distributed neural code representing a robin is more similar to the code representing a sparrow than the code representing a penguin. Similarity is defined as the distance in neural state-space. While this idea is far from new (Churchland, 1993, 1998; Rumelhart & McClelland, 1986), we can now actually begin to relate distances in psychological space to distances in neural state space (Haxby et al., 2011; Kriegeskorte & Kievit, 2013; Kriegeskorte et al., 2008; Raizada & Connolly, 2012; Weber, Thompson-Schill, Osherson, Haxby, & Parsons, 2009). Indeed, the effectiveness of contemporary imaging techniques such as adaptation and multivoxel pattern analysis in revealing the representation of human knowledge in the brain is strong evidence that neural patterns subserving performance on a variety of tasks are distributed and graded. If they were not, these methods would not work as well as they do.

The present work shows that representations of even formal categories—categories having easily stated and known conditions for membership—are nevertheless highly sensitive to subtle differences in their eliciting cues. Representations of triangles activated by the word "triangle" and by the phrase

[7]Experiments on prototype effects have sometimes been seen to be arguing that a learned category is represented *as the prototype* (Osherson & Smith, 1981; cf. Laurence & Margolis, 1999). Posner and Keele's (1968) actual findings showed that although people showed sensitivity to the unobserved visual template, it was judged to be *less* familiar than actually observed exemplars—a result clearly at odds with the view that the template is *the* learned representation of the category. The idea that embracing prototype effects implies committing to a theory in which the prototype serves as the representation of the category reflects a misunderstanding. In Rosch's own words:

"To speak of a prototype at all is simply a convenient grammatical fiction; what is really referred to are judgments of degree of prototypicality. Only in some artificial categories is there by definition a literal single prototype. . . . For natural-language categories, to speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation (Rosch, 1978, p. 40).

[8]Much of this work involved contrasting "prototype" and "exemplar" models (see Mack, Preston, & Love, 2013, for recent formulation). Although often presented in opposition, the strengths of both formalisms are neatly incorporated into connectionist/connectionist-inspired architectures (Kruschke, 1992; Love, Medin, & Gureckis, 2004; Rogers & McClelland, 2004) in which a prototype is an emergent rather than a fixed entity. It may be tempting to interpret the present results in terms of people accessing different exemplars in different contexts—for example, accessing an exemplar of "equilateral triangle" when hearing or reading "triangle". But it is difficult to see how invoking exemplar theory in this way offers a genuine alternative explanation to what is proposed here.

"three-sided polygon" are systematically different. This type of effect is an especially strong case of context, and context effects are of course endemic to human cognition (Barsalou, 1987; Casasanto & Lupyan, 2014; Tversky & Kahneman, 1974). What is relevant about a piano in the context of playing one may not be relevant in the context of moving one (Tabossi & Johnson-Laird, 1980). On the present view, conceptual representations reside in a high-dimensional space, representational flexibility is accomplished by warping the representation such that, for example, sound-related features are highlighted in the context of playing a piano and abstracted over in the context of moving one. The graded and distributed nature of concepts for which I argue (cf. Fodor, 1998) easily accommodates the flexibility with which people deploy knowledge as a function of the context that elicits it.

Where does the prototypical structure for triangles come from?

The present findings make it clear that some triangles are more privileged than others in people's representations. But what makes horizontally oriented isosceles and equilateral triangles typical? The answer is unlikely to simply be frequency of exposure if only because people are likely to see triangles in all kinds of orientations (i.e., triangles do not have a "natural" orientation). It is also of note that the most frequent triangle type in people's *formal* exposure to triangles—the trigonometrically friendly right triangle—is actually not especially salient. For example, of the 269 drawn drawings in Experiments 1A–1B, fewer than 10% had an angle within 4° of 90°. Such triangles are also not rated as particularly typical (Lupyan, 2013).[9] In contrast, horizontally oriented equilateral and isosceles triangles appear to form almost the entirety of illustrations in children's books, likely because book illustrators share the canonicality biases shown by the participants here. A possible source may lie in a general preference for simplicity/minimization of description length (Attneave, 1955; Chater, 1999). Fully specifying an oblique scalene triangles requires knowing four bits of information: two side lengths, one angle, and the orientation. Specifying a horizontally oriented isosceles triangle requires two bits of information: the apex angle and the length of one side. Specifying an upright horizontally oriented equilateral triangle requires just one bit: the length of a single side. While such an account may form part of the explanation for the typicality gradient, this account does not generalize to canonicality effects in other shapes. For example, when asked to draw a rectangle, not only do people overwhelmingly draw horizontally oriented rectangles (width > height), but the modal aspect ratio approximates the golden ratio ($\Phi \approx 1.62$; Lupyan, unpublished data).

An additional—and frequently overlooked—source of canonicality may lie in contrast categories (Cohen & Nosofsky, 2000; Dry & Storms, 2010; Goldstone, Steyvers, & Rogosky, 2003). Consider a task in which participants are shown pictures of chairs and tables and are asked to classify each picture as quickly as possible. The pictures identified most quickly will *not* be the most typical chairs and tables, but the most un-chair-like tables and the most un-table-like chairs—that is, the items farthest from the category boundary (Lupyan, 2008a). Taking identification times as indicative of typicality means that typicality depends on the contrast categories (see also Barsalou, 1987). In judging how typical a given chair is of chairs in general, the representation of the chair category may be one that *best discriminates between the target category and commonly contrasting categories*. For a chair, this may be other pieces of furniture. For a dog, this may be other pets. For a triangle, this may be other simple geometric shapes. On this view, the most typical/canonical triangle may correspond to a shape that best contrasts with non-triangles. Although there are infinitely many non-triangles, some are more likely to serve as contrast categories, and it is these that are most weighed in the process that gives rise to the so-called "prototype".[10] In sum, at this point it cannot be said with certainty why canonical triangles are what they are, though the discussion above hints at some ways of finding out.

[9]This under-representation of right triangles may well disappear if people are asked to draw "maths triangles" or are even primed with a mathematical context, a finding that would further corroborate the flexible nature of formally defined categories.

[10]In a recent presentation, Greene (2014) discussed a task in which participants were asked to indicate what objects never occur as part of certain scene categories. For example, "What do you never find in a kitchen?" Such questions would appear to be completely under-determined as there is an infinite number of objects one does not find in a kitchen. Yet for many categories, participants had a shockingly high amount of agreement (e.g., no toilets in a kitchen, no skyscrapers in a field). On the view advocated here, these responses are indicative of implicit contrast categories in action: If a common contrast category of kitchen is "bathroom", and "toilet" is a common object in the latter, then it is a salient non-member of the former.

Why language matters

Just as we come to (quite automatically) categorize a chair as something to sit on, as language users we also see a chair as a "chair", that is, it as something that can be *named*. On one view, language provides us with a convenient way to *talk* about a concept and to draw attention to things in the world, but language does not play an interesting role in either concept learning or concept use (e.g., Gleitman & Papafragou, 2005; Gomila, 2011; Li & Gleitman, 2002; Pinker, 1994, for discussion). On an alternate view (Lupyan, 2012b; Lupyan & Bergen, 2015; Lupyan & Clark, 2015), language is a uniquely powerful system for activating a certain kind of representational state. It so happens that this state comes close to what people mean when they talk about concepts. Consider a recent study (Çukur et al., 2013) in which participants were asked to passively view videos, or to monitor them for presence of humans or vehicles. The study showed that representations in virtually all parts of cortex (including early visual cortex) were warped by the task. Attending to humans caused a compression of semantically distant categories and an expansion of the task-relevant categories. Attending to vehicles produced a similar shift in neural state space (compression of categories semantically far from vehicles and expansion of vehicle-related categories). Such results show the distributed/graded nature of concepts as represented by the brain, and the flexibility with which these representations change according to task demands. The two active conditions —requiring participants to attend to vehicles or people —varied in their goal states. But how were these goal states activated? But how were people instructed to attend to vehicles? The answer is that people were *told* to attend to vehicles. Is language simply a convenient means to communicate a goal state? Consider communicating the same instruction without language. What would one do? Show a picture of a car? But that's a particular type of vehicle. Show many different kinds of cars? Throw in some trucks, motorcycles, and bicycles? But how would one prevent the subject from concluding that the instruction is actually to attend to wheels (thus false alarming to pictures of wheels and failing to respond to wheel-less cars)? What about "attend to humans?" Would one attempt to approximate this instruction by showing someone pictures of faces? Full bodies? Groups of people?

Communicating a categorical goal state is difficult (and perhaps even impossible) in the absence of language.[11] Without a label, the participant may become trapped in the land of the concrete. The label transcends this concreteness, eliciting (at least an approximation of) a categorical state (Edmiston & Lupyan, 2015). In the above example, prior experience of associating "human" and "vehicle" with a variety of perceptual states is leveraged to rapidly induce a representational state that is an idealization of any specific perceptual encounter.

One may contest that it is possible to train non-human animals to perform this very task by associating a particular cue—the image of a circle, say—with various vehicles. Following such training, the cue will serve to activate a kind of vehicle template (prototype). As language users, people come to the lab already equipped with thousands of such cues: words (Lupyan & Clark, 2015)!

On this view, language is not just a powerful way of communicating a goal state, but is critically involved in eliciting the representation of the goal state in the first place. This view implies that lack of words ought to have cognitive consequences, particularly with respect to forming abstract categories. Indeed, categorization—particularly rule-based categorization—is compromised by even minor language impairments induced in healthy adults (e.g., Lupyan, 2009; Perry & Lupyan, 2014), and more notably in cases of acquired language disorders such as aphasia (Baldo, Bunge, Wilson, & Dronkers, 2010; Baldo et al., 2005; Lupyan & Mirman, 2013; see Gainotti, 2014; Murray, 2012; Vignolo, 1999, for reviews). Learning new lexical distinctions, or making previously known lexical distinctions more accessible, has the opposite effect: a reification of categories (e.g., Boutonnet & Lupyan, 2015; see Lupyan, in press; Lupyan & Bergen, 2015, for reviews).

FURTHER CONSIDERATIONS

Arguments from competence versus performance

In rejecting the classical definitional theory of concepts, some have posited a distinction between a conceptual core—an abstract symbolic level specifying membership rules—and an identification procedure that accounts for prototype/typicality/context effects. As articulated by Osherson and Smith (1981, p. 57), "The core is concerned with those aspects of a

[11]A typical adult faced with this procedure might quickly conclude that what is being demanded is the *category*, rather than any specific instance. It is my contention that a *rapid* realization of this sort is mediated—implicitly or explicitly—by verbal labels. Absent the labels, the participant may be guided much more by overall similarity to provided exemplars (Perry & Lupyan, 2014).

concept that explicate its relation to other concepts, and to thoughts, while the identification procedure specifies the kind of information used to make rapid decisions about membership" (see also Armstrong et al., 1983; Gleitman, Armstrong, & Connolly, 2012). One may contest that the representations brought to bear in current tasks are not *constitutive* of the triangle concept and that people's behaviour reflects the workings of an identification procedure or reflects "mere" performance. The problem with this critique is that it discounts data while placing explanatory power in a "competence" (or "core") that does not lend itself to empirical investigation. Of course, absence of evidence is not evidence of absence, and perhaps someday we will find thus-far elusive context-invariant symbols in the brain (Gallistel & King, 2009). But such claims need to be based on empirical data rather than on a construct of competence not open to empirical investigation.

But people can think abstract thoughts

If people's representations of formal categories show profound typicality effects in a wide range of tasks, what explains people's ability to think abstractly? For example, how do people learn and apply rules that hold equally well for all members of a category? The answer is: not very well. So, not only do classification judgments—both timed and untimed—vary as a function of typicality, but so do category-based inferences. Not only do people systematically fail to classify atypical (scalene, oblique) triangles as triangles, but when explicitly told that all triangles have angles that add to 180° and asked to select shapes that have this property, systematically leave out non-canonical triangles (Lupyan, 2013).[12]

Such errors are telling. They are exactly the kinds of errors expected if people's reasoning is based on the sorts of continuous representational spaces that characterize perception. On the present account, the central question becomes how to implement abstract reasoning and other symbolic operations using the distributed, continuous representations that underlie our knowledge. I have argued here that the answer may lie in the use of language (and perhaps other external symbolic systems) to transform these continuous representations into the more discretized and abstract representations that symbolic operations require.

CONCLUSION

Although every *real* triangle is clearly equilateral, isosceles, or scalene, the question of whether the mental representation of a triangle is also thus constrained has been a longstanding pre-occupation. On the classical cognitivist view, the triangle would appear to be the paragon of simplicity. With easily enumerated necessary and sufficient conditions, the triangle is the ideal candidate for symbolic representation. Nothing should be simpler than to conceive of triangles as abstractions. And yet, across a variety of tasks—drawing, recognition, visual judgments, inference—people show profound typicality effects that are neither predicted nor easily accommodated by this classical view. Insofar as such performance sheds light on how concepts are represented in the mind (and if it does not, then what would?), the results point to triangles being represented as idealized perceptual states—what some have called prototypes. These prototypes are flexibly elicited by verbal labels, a conclusion supported by the consistent findings that even co-extensive cues like "triangle" and "three-sided polygon" elicit systematically different representations. These results inform and constrain theories of concepts and speak to the role of language in human cognition.

References

Althaus, N., & Mareschal, D. (2014). Labels direct infants' attention to commonalities during novel category learning. *PLoS ONE*, *9* (7), e99670. Retrieved from http://doi.org/10.1371/journal.pone.0099670

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition*, *13*(3), 263–308. Retrieved from http://doi.org/10.1016/0010-0277(83)90012-4

Attneave, F. (1955). Symmetry, information, and memory for patterns. *The American Journal of Psychology*, *68*(2), 209–222. Retrieved from http://doi.org/10.2307/1418892

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*(1), 3–26.

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language*, *113*(2), 59–64. Retrieved from http://doi.org/10.1016/j.bandl.2010.01.004

[12]Similar observations were made much earlier by Tversky and Kahneman in their work on the availability heuristic (e.g., Tversky and Kahneman, 1974), and it is likely that there is considerable overlap in the mechanisms underlying both types of effects.

Baldo, J. V., Dronkers, N. F., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. Y. (2005). Is problem solving dependent on language? *Brain and Language*, *92*(3), 240–250.

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences*, *11*(8), 327–332. Retrieved from http://doi.org/10.1016/j.tics.2007.06.003

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization* (pp. 101–140). Cambridge: Cambridge University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *The Behavioral and Brain Sciences*, *22*(4), 577–609; discussion 610–660.

Berkeley, G. (2009). *Principles of human knowledge and three dialogues*. Oxford: Oxford University Press.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *The Journal of Neuroscience*, *35*(25), 9329–9335.

Burling, R. (1993). Primate calls, human language, and nonverbal communication. *Current Anthropology*, *34*(1), 25–53.

Casasanto, D., & Lupyan, G. (2014). All concepts are Ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *Concepts: New directions* (pp. 543–566). Cambridge: MIT Press.

Casasola, M. (2005). Can language do the driving? The effect of linguistic input on infants' categorization of support spatial relations. *Developmental Psychology*, *41*(1), 183–192. Retrieved from http://doi.org/10.1037/0012-1649.41.1.188

Chater, N. (1999). The search for simplicity: A fundamental cognitive principle? *The Quarterly Journal of Experimental Psychology: Section A*, *52*(2), 273–302.

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 216–226.

Churchland, P. M. (1993). State-space semantics and meaning holism. *Philosophy and Phenomenological Research*, *53*(3), 667–672. Retrieved from http://doi.org/10.2307/2108090

Churchland, P. M. (1998). Conceptual similarity across sensory and neural diversity: The fodor/lepore challenge answered. *The Journal of Philosophy*, *95*(1), 5. Retrieved from http://doi.org/10.2307/2564566

Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same-different judgments. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1549–1569.

Connell, L., & Lynott, D. (2014). Principles of representation: Why you can't represent the same concept twice. *Topics in Cognitive Science*, *6*(3), 390–406. Retrieved from http://doi.org/10.1111/tops.12097

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience*, *16*(6), 763–770. Retrieved from http://doi.org/10.1038/nn.3381

Danziger, K. (1980). The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences*, *16*(3), 241–262.

Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science*, *19*(2), 189–195. Retrieved from http://doi.org/PSCI2066

Dove, G. (2009). Beyond perceptual symbols: A call for representational pluralism. *Cognition*, *110*(3), 412–431. Retrieved from http://doi.org/10.1016/j.cognition.2008.11.016

Dry, M. J., & Storms, G. (2010). Features of graded category structure: Generalizing the family resemblance and polymorphous concept models. *Acta Psychologica*, *133*(3), 244–255. Retrieved from http://doi.org/10.1016/j.actpsy.2009.12.005

Edmiston, P., & Lupyan, G. (2013). Verbal and nonverbal cues activate concepts differently, at different times. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2243–2248). Austin, TX: Cognitive Science Society.

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100.

Elman, J. L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon*, *6*(1), 1–33. Retrieved from http://doi.org/10.1075/ml.6.1.01elm

Evans, V. (2009). *How words mean lexical concepts, cognitive models, and meaning construction*. Oxford: Oxford University Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong* (1st ed.). Oxford , New York: Oxford University Press.

Gainotti, G. (2014). Old and recent approaches to the problem of non-verbal conceptual disorders in aphasic patients. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *53*, 78–89. Retrieved from http://doi.org/10.1016/j.cortex.2014.01.009

Gallistel, C. R., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience* (1st ed.). Chichester: Wiley-Blackwell.

Gleitman, L., Armstrong, S. L., & Connolly, A. C. (2012). Can prototype representations support composition and decomposition? In M. Werning, W. Hinzen, & E. Machery (Eds.), *Oxford handbook of compositionality*. New York: Oxford University Press. Retrieved from http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199541072.001.0001/oxfordhb-9780199541072-e-20

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

Goldstone, R. L., Steyvers, M., & Rogosky, B. J. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition*, *31*(2), 169–180.

Gomila, T. (2011). *Verbal minds: Language and the architecture of cognition* (1st ed.). Amsterdam , Boston: Elsevier.

Greene, M. (2014). *Human estimates of object frequency are frequently over-estimated*. Presented at the Fourteenth Annual Meeting of the Vision Sciences Society.

Hampton, J. A. (2006). Concepts as prototypes. In *The psychology of learning and motivation: Advances in research and Theory* (Vol. *46*, pp. 79–113). London: Elsevier.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science*, *31*(3), 355–384. Retrieved from http://doi.org/10.1080/15326900701326402

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences*, 201312567. Retrieved from http://doi.org/10.1073/pnas.1312567111

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., … Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, *72*(2), 404–416. Retrieved from http://doi.org/10.1016/j.neuron.2011.08.026

Holmes, K. J., & Wolff, P. (2013). Spatial language and the psychological reality of schematization. *Cognitive Processing*, *14*(2), 205–208. Retrieved from http://doi.org/10.1007/s10339-013-0545-5

Hume, D. (2012). *A treatise of human nature*. CreateSpace Independent Publishing Platform.

James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*. Retrieved from http://www.gutenberg.org/ebooks/5116

Kemmerer, D. (2010). How words capture visual experience: The perspective from cognitive neuroscience. In B. C. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience* (1st ed., pp. 289–329). New York, NY, USA: Oxford University Press.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825. Retrieved from http://doi.org/10.1016/j.cortex.2011.04.006

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: Integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. Retrieved from http://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. Retrieved from http://doi.org/10.1016/j.neuron.2008.10.043

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), *Concepts: core readings* (pp. 3–81). Cambridge, Mass: A Bradford Book.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*(3), 265–294.

Locke, J. (1995). *An essay concerning human understanding*. Amherst, N.Y: Prometheus Books.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, *50*(4), 315–353. Retrieved from http://doi.org/10.1016/j.cogpsych.2004.09.004

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*(2), 309–332.

Lupyan, G. (2007). Reuniting categories, language, and perception. In D. S. McNamara & J. G. Trafton (Eds.), *Twenty-Ninth Annual Meeting of the Cognitive Science Society* (pp. 1247–1252). Austin, TX: Cognitive Science Society.

Lupyan, G. (2008a). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, *137*(2), 348–369.

Lupyan, G. (2008b). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, *108*(2), 566–577.

Lupyan, G. (2009). Extracommunicative functions of language: Verbal interference causes selective categorization impairments. *Psychonomic Bulletin & Review*, *16*(4), 711–718. Retrieved from http://doi.org/10.3758/PBR.16.4.711

Lupyan, G. (2012a). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Cognition*, *3*(54). Retrieved from http://doi.org/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). Waltham, MA: Academic Press. Retrieved from http://www.sciencedirect.com/science/article/pii/B9780123942937000078

Lupyan, G. (2013). The difficulties of executing simple algorithms: Why brains make mistakes computers don't. *Cognition*, *129*(3), 615–636. Retrieved from http://doi.org/10.1016/j.cognition.2013.08.015

Lupyan, G. (in press). The centrality of language in human cognition. *Language Learning*. http://doi.org/10.1111/lang.12155

Lupyan, G., & Bergen, B. (2015). How language programs the mind. *Topics in Cognitive Science*. Retrieved from http://doi.org/10.1111/tops.12155

Lupyan, G., & Casasanto, D. (2015). Meaningless words promote meaningful categorization. *Language and Cognition*, *7*(2), 167–193. Retrieved from http://doi.org/10.1017/langcog.2014.21

Lupyan, G., & Clark, A. (2015). Words and the world: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279–284.

Lupyan, G., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex*, *49*(5), 1187–1194. Retrieved from http://doi.org/10.1016/j.cortex.2012.06.006

Lupyan, G., Mirman, D., Hamilton, R. H., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition*, *124*(1), 36–49. Retrieved from http://doi.org/10.1016/j.cognition.2012.04.002

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083.

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253. Retrieved from http://doi.org/10.3758/APP.72.8.2236

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology: General*, *141*(1), 170–186. Retrieved from http://doi.org/10.1037/a0024904

Mack, M. L., Preston, A. R., & Love, B. C. (2013). Decoding the brain's algorithm for categorization from its neural implementation. *Current Biology*, *23*(20), 2023–2027. Retrieved from http://doi.org/10.1016/j.cub.2013.08.035

Margolis, E., & Laurence, S. (1999). *Concepts: Core readings*. Cambridge, MA: A Bradford Book.

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 775–799.

Mirman, D. (2014). *Growth curve analysis and visualization using R* (Revised ed.). Boca Raton: Chapman and Hall/CRC.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: The MIT Press.

Murray, L. (2012). Attention and other cognitive deficits in aphasia: Presence and relation to language and communication measures. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, *21*(2), S51–S64 Retrieved from http://doi.org/10.1044/1058-0360(2012/11-0067)

Nosofky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition*, *9*(1), 35–58. Retrieved from http://doi.org/10.1016/0010-0277(81)90013-5

Pecher, D. (2013). The perceptual representation of mental categories. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 358–373). Oxford: Oxford University Press.

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66–72. Retrieved from http://doi.org/10.1016/j.bandl.2014.05.005

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

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

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, *77*(3P1), 353–363.

Prinz, J. J. (2004). *Furnishing the mind: Concepts and their perceptual basis* (New ed.). Cambridge, MA: The MIT Press.

Raizada, R. D. S., & Connolly, A. C. (2012). What makes different people's representations alike: Neural similarity space solves the problem of across-subject fMRI decoding. *Journal of Cognitive Neuroscience*, *24*(4), 868–877. Retrieved from http://doi.org/10.1162/jocn_a_00189

Rogers, T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: Bradford Book.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, volumes 1 and 2*. Cambridge, MA: MIT Press.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810. Retrieved from http://doi.org/10.1016/j.neuropsychologia.2007.05.002

Spencer, J. P., & Hund, A. A. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology: General*, *131*(1), 16–37.

Spivey, M. J. (2008). *The continuity of mind*. New York: Oxford University Press.

Tabossi, P., & Johnson-Laird, P. N. (1980). Linguistic context and the priming of semantic information. *The Quarterly Journal of Experimental Psychology*, *32*(4), 595–603. Retrieved from http://doi.org/10.1080/14640748008401848

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. Retrieved from http://doi.org/10.1126/science.185.4157.1124

Vignolo, L. A. (1999). Disorders of conceptual thinking in aphasia. In G. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental neuropsychology* (pp. 273–288). Hove: Psychology Press.

Watson, R. A. (1995). *Representational Ideas: From Plato to Patricia churchland*. Springer Science & Business Media.

Weber, M., Thompson-Schill, S. L., Osherson, D., Haxby, J., & Parsons, L. (2009). Predicting judged similarity of natural categories from their neural representations. *Neuropsychologia*, *47*(3), 859–868. Retrieved from http://doi.org/10.1016/j.neuropsychologia.2008.12.029

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234. Retrieved from http://doi.org/10.1016/j.tics.2014.02.008

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. Retrieved from http://doi.org/10.1037/0096-3445.135.1.1