Meaningless words promote meaningful categorization

GARY LUPYAN and DANIEL CASASANTO

Language and Cognition / *FirstView* Article / June 2014, pp 1 - 27
DOI: 10.1017/langcog.2014.21, Published online: 13 June 2014

Link to this article: http://journals.cambridge.org/abstract_S1866980814000210

How to cite this article:
GARY LUPYAN and DANIEL CASASANTO Meaningless words promote meaningful categorization . Language and Cognition, Available on CJO 2014 doi:10.1017/langcog.2014.21

Language and Cognition (2014), Page 1 of 27. doi:10.1017/langcog.2014.21

Meaningless words promote meaningful categorization*

GARY LUPYAN
Department of Psychology, University of Wisconsin-Madison

AND

DANIEL CASASANTO
Department of Psychology, University of Chicago

(*Received* 27 *August* 2013 – *Revised* 07 *March* 2014 – *Accepted* 25 *March* 2014)

ABSTRACT

On traditional accounts, word meanings are entries in a mental lexicon. Nonsense words lack such entries, and are therefore meaningless. Here, we show that UNDER SOME CIRCUMSTANCES nonsense words function indistinguishably from conventional words. The 'nonsense' words *foove* and *crelch* led participants to select systematically different clusters of adjectives and were reliably matched to different species of alien creatures (e.g., 'crelches' were pointy and narrow and 'fooves' were large and fat). In a categorization task in which participants learned to group two species of aliens primarily on the basis of roundness/pointiness, these novel labels facilitated performance as much as conventional words (e.g., *round*, *pointy*). The results expand the scope of research on sound symbolism and support a non-traditional view of word meaning according to which words do not have meanings by virtue of a conventionalized form−meaning pairing. Rather, the 'meaning' of a word is the effect that the word form has on the user's mental activity.

KEYWORDS: category learning, sound symbolism, iconicity, mental lexicon

1. Introduction

Can nonsense words be meaningful? On traditional accounts of language, word meanings are entries in a mental lexicon (e.g., Jackendoff, 2002). Therefore, words not listed in the mental lexicon cannot be meaningful. Even non-traditional theories of language hold that words have meanings by virtue

---

[*] Address for correspondence: Gary Lupyan, 1202 W. Johnson St. Rm. 419, University of Wisconsin-Madison, Madison, WI 53706. email: lupyan@wisc.edu

of conventionalized 'form–meaning pairings' (Evans, forthcoming, 2009; Taylor, 2012). Yet, sometimes, 'nonsense' words appear to convey meaning, even when they are novel, and even though they have not been paired with any meanings by convention. Consider Lewis Carroll's nonsense poem *Jabberwocky* (2010 [1871]). The opening two lines – "Twas brillig and the slithy toves / did gyre and gimble in the wabe", while certainly open for interpretation, are hardly meaningless. The nonsense words of *Jabberwocky* are made meaningful by a combination of phonological cuing and syntactic and distributional information. So, *slithy* is used in an adjectival frame and has phonological neighbors *lithe* and *slimy*. It modifies *tove*, which, due to its syntactic position and phonology, is likely to be a noun, and moreover, an animate being with the ability to *gyre and gimble*, which in turn connote some type of spinning and tumbling. This confluence of cues appears to be sufficient to lead people to endow these nonsense words with meanings (Cabrera & Billman, 1996; Johnson & Goldberg, 2013).

Another example of apparently nonsensical words being interpreted as meaningful even in the absence of syntactic cues is the phenomenon of sound symbolism. For example, both adults and children consistently match nonsense words like *baluma* and *tukeetee*, or *bouba* and *kiki* to rounded and angular shapes, respectively (Kohler, 1947; Maurer, Pathman, & Mondloch, 2006; Ramachandran & Hubbard, 2001). Although the origin of such sound-shape correspondences is a matter of some debate (see Nuckolls, 1999; Perniss, Thompson, & Vigliocco, 2010; Spence, 2011; for reviews), it is now well established that such correspondences exist and that they can guide behavior in both explicit and implicit paradigms (Aveyard, 2012; Kovic, Plunkett, & Westermann, 2010; Nygaard, Cook, & Namy, 2009; Parise & Spence, 2012; Westbury, 2005).

The present study builds on this rich literature in three ways. First, we sought to find out if nonsense words can guide category learning. Rather than testing for effects of sound-symbolism on learning the mapping between a word and an object or definition (Imai, Kita, Nagumo, & Okada, 2008; Kantartzis, Imai, & Kita, 2011; Nygaard et al., 2009), we were interested in the effects of words on constructing the categories themselves (Lupyan, Rakison, & McClelland, 2007), that is, on the process of homing in on the features and dimensions important for distinguishing different classes of stimuli. Second, we directly compared the contribution of novel words and conventional (familiar) words to category learning, with the aim to see if, under some circumstances, the meanings activated by the novel words are as reliable as those activated by familiar words, and as useful to learning new categories. Third, we conducted a series of control experiments to determine whether the effect of novel words on category learning can be explained by strategic activation of real words that are phonologically similar to the novel words.

To preview our findings: we first show that the nonce words *foove* and *crelch* guide people's behavior in predictable ways. When asked to consider these words as referring to shapes, people consistently matched *foove* to adjectives like *round* and *fat*, and *crelch* to adjectives like *pointy* and *narrow*. We also show that just as people can be told to pick out the 'pointies' out of a group of novel creatures ('aliens'), so people can be told to pick out the 'crelches' to similar effect. After these preliminary experiments, we conducted a category-learning study (following Lupyan et al., 2007), aimed at testing whether the labels 'foove' and 'crelch' can guide the learning of novel categories in a way similar to conventional words, and found that nonce words functioned indistinguishably from conventional ones. Finally, a series of control experiments (C1–C3) showed the meanings people attribute to *foove* and *crelch* do not appear to derive from similarity to real words.

This work is motivated by a non-traditional view of word meaning according to which the 'meaning' of a word is the effect it has on the user's mental activity. Consistent with Elman (2004, 2009, 2011), we argue that the forms of words (whether phonological or orthographic) are cues to activate information in memory (Casasanto & Lupyan, in press). On this view, word forms do not need to be paired with meanings by convention in order to be 'meaningful'. Even novel word forms can guide the activation of information in memory systematically, on the basis of the contexts in which they are encountered, and in some cases on the basis of aspects of the novel words' forms, per se. Word forms, therefore, function much like other kinds of sensory stimuli (e.g., objects, gestures, pictures); their effects on our mental activity are not mediated by a process of looking up stored entries in a mental lexicon.

Both conventional words and nonsense words can activate mental representations. To the extent that certain novel stimuli like *foove* and *crelch* reliably activate representations of sensory features like roundness or angularity – whatever the origin of such links may be – these 'nonsense' words are meaningful. Some researchers have called non-conventional words for which people construct ad hoc meanings 'nonce words' – words that have meanings FOR THE NONCE (Clark, 1983). In this paper we sought to test the hypothesis that, under some circumstances, nonce words and 'real' words can guide learning and behavior in similar ways.

2. General description of materials

2.1. NOVEL WORD STIMULI

The two nonce words used in the present studies – *foove* and *crelch* – were selected on the basis of the first author's intuitions from a larger list of nonce words (*shonk*, *whelph*, *scaif*, *crelch*, *foove*, and *streil*) used originally by Lupyan and Thompson-Schill (2012). These six words were selected from a published database of monosyllabic nonwords with phonotactically legal syllables (ARC database: Rastle, Harrington, & Coltheart, 2002). The two words used in this study have similar orthographic and phonetic bigram frequencies and very limited immediate orthographic neighborhoods (*foove* has one immediate orthographic neighbor and *crelch* has none). According to the ARC database, *foove* has five immediate phonological neighbors (e.g., *move*) and *crelch* has none.[1] The words were presented visually in the Preliminary Experiments 1–2 and Control Experiments C1–C3, and auditorily in the main experiment.

2.2. NOVEL VISUAL STIMULI

Participants were asked to perform judgments and learn to categorize two species of ‘aliens’ from the YUFO stimulus set originally designed as a set of organic-looking novel stimuli for categorization studies (Gauthier, James, Curby, & Tarr, 2003). The tasks here used two YUFO ‘species’. All stimuli are shown in Figure 1. The primary difference between the two species, as reported by almost all the participants in post-study questionnaires in both the present and a previous study (Lupyan et al., 2007), concern differences in head shapes. The aliens from one species (Figures 1 a1–a12) have a subtle groove or ridge on their ‘heads’ and have narrower heads, while the aliens from the other species (Figure 1 b1–b12) have smoother and more rounded heads.

3. Preliminary experiments

Preliminary Experiments 1 and 2 were designed to quantify the extent to which the nonce words *foove* and *crelch* are matched by naive participants to various adjectives (Preliminary Experiment 1) and the degree to which the aliens can be grouped into the predetermined species simply by asking participants to assign them to the ‘foove’ and ‘crelch’ species (Preliminary Experiment 2).

3.1. PRELIMINARY EXPERIMENT 1: LABEL-CUED ATTRIBUTE SELECTIONS

3.1.1. *Participants, materials, and procedure*

Thirty-two participants were recruited via Amazon Mechanical Turk. The only restriction was that the participants were from the United States and, to

---

[1] In general, immediate neighbors are words that can be formed by adding, subtracting, or substituting a single letter or phoneme. Unfortunately, the authors of the ARC database did not provide their operational definitions of immediate neighbors.

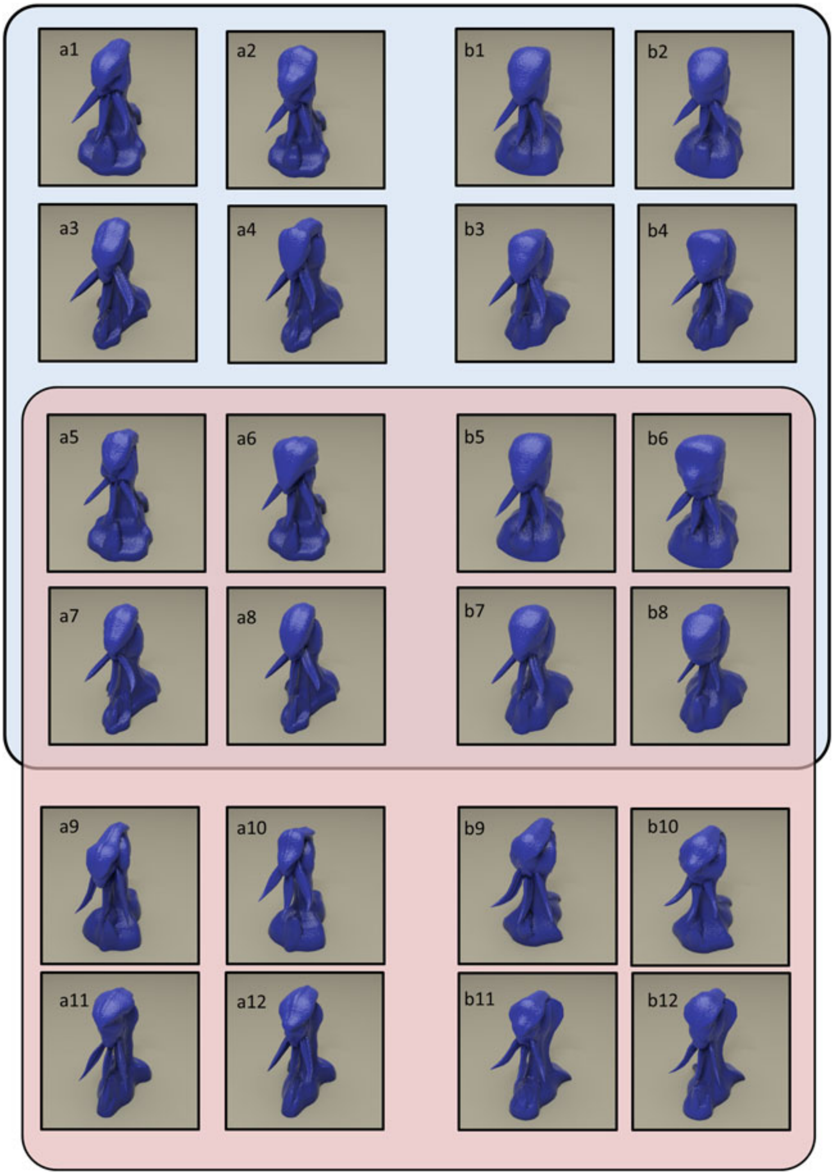

Fig. 1. The 'alien' stimuli used in Preliminary Study 2 and Experiment 1. See Section 4.1 for further details on the categorization task. Preliminary Study 2 used a1–a8 and b1–b8. The middle items (a5–a8 and b5–b8) were used in for the entire 18 blocks of the categorization task of Experiment 1. Stimuli a1–a4 and b1–b4 were replaced by stimuli a9–a12 and b9–b12 during block 6 of training.

ensure quality, participants had to have an overall task failure rate of less than 5% (a Mechanical Turk default setting).[2] Each participant took part in only one task. Participants were asked to imagine a shape called a *foove* and a *crelch*, and select appropriate attributes for each shape. The exact query was: "Imagine a shape called a 'foove' ['crelch'] Would such a shape be … (check all that apply)." Below the prompt was a list of attributes shown in Figure 2 with a checkbox next to each one.[3] All participants responded to each prompt, with the order of the nonce words and choices randomized.

The attributes we used were compiled from post-study questionnaires from a previous study using the same 'alien' stimuli used here (Lupyan et al., 2007). In this study, people learned – in much the same way as they did in Experiment 1, detailed below – to distinguish between two types of 'alien' and were then asked to describe what features distinguished the two groups. From these answers we extracted fifteen modal descriptors that referenced shape attributes for each alien species (Figure 2). In addition, four non-perceptual attributes ('safe', 'dangerous', 'friendly', 'unfriendly') were included because they were frequently mentioned in the questionnaires. However, because our focus is on PERCEPTUAL information cued by the nonce words, we omit these non-perceptual properties from the statistical analyses presented below.

3.1.2. *Results and discussion*

The basic results are shown in Figure 2. Participants selected an average of 4.3 attributes for *foove* and 4.6 for *crelch*. To quantify participants' biases for matching the nonce words to specific shape adjectives, we designated a word-attribute mapping as 'congruent' if it was provided by more than 50% of participants. For example, mapping 'pointy' to *crelch* was scored as a congruent mapping and to *foove* as incongruent.

Of all chosen attributes, 78% were congruent with the prompt. Of all possible choices, a congruent attribute was 36% likely to be selected while an incongruent attribute was only 10% likely to be selected – a highly reliable difference as revealed by a within-subject logistic regression ($z = 5.66$, $p \ll .0001$). There were no differences in congruency between *foove* and *crelch* ($t < 1$). The results show that, when choices are constrained to shape

---

[2] Although we did not require participants to be native English speakers, the great majority (96% across all the studies using Mechanical Turk) reported English to be their native language. We did not have any theoretical reason to exclude non-native English speakers from the analyses and did not have sufficient power to examine if the responses of non-native speakers were meaningfully different.

[3] Participants were actually shown twenty choices, but due to an error, 'Narrow' appeared twice, making for nineteen unique options.

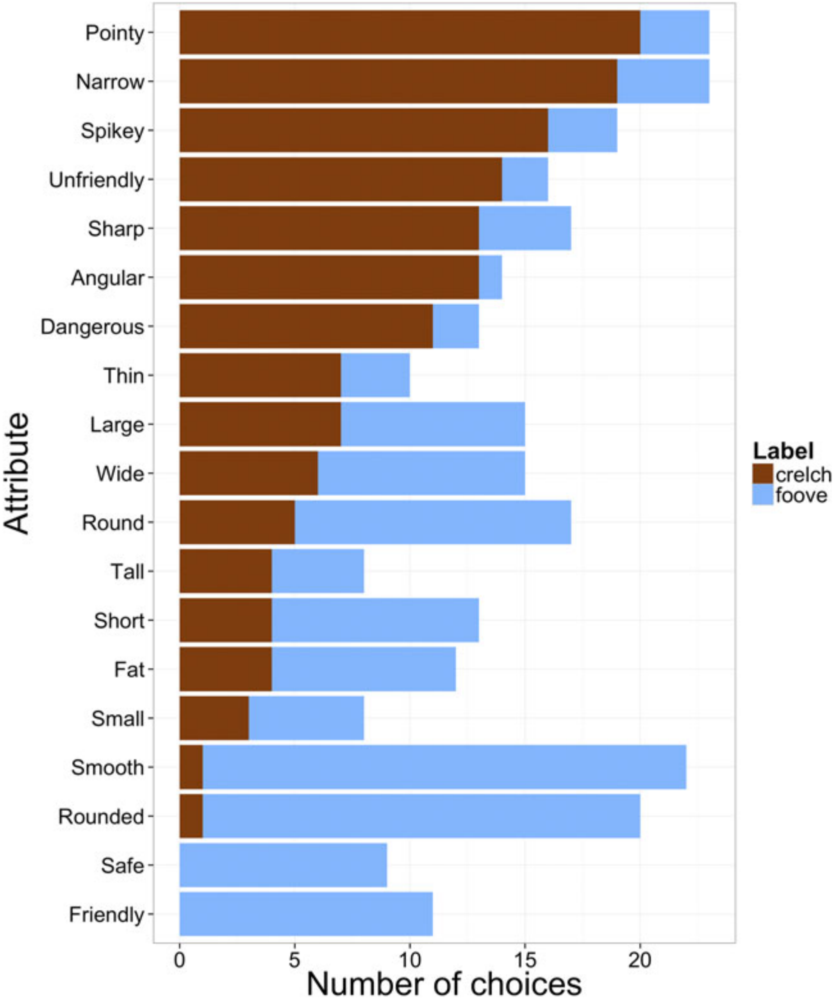

Fig. 2. Results from Preliminary Experiment 1 showing the proportion of participants selected each attribute when asked to imagine a shape called a 'foove' or 'crelch'.

and affective attributes, the words *foove* and *crelch* activate reliable meanings. Might the consistent responses to the nonce words be mediated by conventional words? Control Experiments C1–C3 were designed to answer this question. As described below, the results offered no evidence of mediation by conventional words (at least no conscious mediation), suggesting instead that the responses rely on implicit links between certain sounds and visual dimensions.

3.2. PRELIMINARY EXPERIMENT 2: LABEL-CUED ITEM SELECTIONS

Preliminary Experiment 1 showed that the nonce words *foove* and *crelch* activate reliable meanings in the context of selecting adjectives. In Preliminary Experiment 2, we asked whether nonce words can also guide selection of complex visual items: the 'alien' creatures shown in Figure 1.

3.2.1. *Participants, materials, and procedure*

Sixty participants were recruited via Amazon Mechanical Turk and assigned to a conventional-label (n = 30) or nonce-label (n = 30) condition. Each participant received two questions showing sixteen randomly ordered aliens (a1–a8 and b1–b8 from Figure 1) arranged in a four-by-four grid. Participants in the conventional-label condition were prompted to select between four and twelve 'aliens' (this range was picked arbitrarily) that were members of the 'round' and 'pointy' species (order counterbalanced). Participants in the nonce-label condition performed the identical task but were prompted to select aliens that were members of the *foove* and *crelch* species.

For the conventional-label condition, participants were asked:

> "Which of these aliens do you think belong to the 'round' ['pointy'] species? Please examine all the choices before making your selections and choose between 4 and 12 aliens you think are the 'rounds' ['pointies']."

For the nonce-label condition, participants were asked:

> "Which of these aliens do you think belong to the 'crelch' ['foove'] species? Please examine all the choices before making your selections and choose between 4 and 12 aliens you think are the 'crelches' ['fooves']."

Each participant made selections for both species with order randomized between participants.

3.2.2. *Results and discussion*

To test the hypothesis that the labels affected which aliens participants chose, we calculated the likelihood of clicking on each item as a function of its species and the congruency between the label prompt and alien species. Clicks on the round-headed aliens when prompted by 'round' or 'foove' and clicks on the pointy-headed aliens when prompted by 'crelch' or 'pointy' were classified as congruent; the rest were incongruent. Of the selected aliens, 67% were congruent with the label. Overall, the likelihood of clicking on aliens congruent or incongruent with the label prompt was 47% and 23%, respectively ($z = 10.96$, $p \ll .0001$) (mixed-effect model using congruency as a predictor). The congruency effect was much stronger in the conventional-label condition (75% of selected items were congruent) compared to the nonce-label condition (59%). This congruency-by-label-type interaction was highly reliable ($z = 5.37$, $p \ll .0005$). Nevertheless, participants were still significantly more likely to choose round-headed aliens when asked to choose 'fooves' and pointy-headed aliens when asked to choose 'crelches' ($z = 4.10$, $p \ll .0005$). The results of these studies show that nonce labels guide not only selections of verbally described visual attributes (Preliminary Study 1), but also guide selections of complex visual stimuli that embody those visual properties (Preliminary Study 2).

A question left open by Preliminary Studies 1–2 is whether the consistency of participants' responses to our nonce words is due to mediation via conventional words, e.g., *crelch* priming *crinkle*. We attempt to rule out this explanation in Experiments C1–C3 which follow the main study, to which we now turn.

4. Experiment 1: effects of 'nonsense' labels on category learning

Having established that 'fooves' are more likely to be matched to the round-headed aliens and 'crelches' to the pointy-headed aliens, we sought to determine whether these labels would guide category learning itself. Although it is not difficult to find the attributes that distinguish the two alien species when the aliens are viewed all together (allowing parallel comparisons), learning to categorize the aliens into their respective 'species' is quite challenging when items are viewed individually and when participants have to learn what the distinguishing features are. In prior work, we have shown that learning to categorize stimuli like these is vastly improved when, in addition to performing the supervised categorization task, participants learn NAMES for the categories (Lupyan et al., 2007). In contrast to claims that we learn names to label our concepts (e.g., Bloom & Keil, 2001; Pullum, 1989; Snedeker & Gleitman, 2004), this finding shows that names augment the process by which we learn the concepts in the first place (see Lupyan, 2012, for review). Here, we examine whether nonsense labels that are meaningful, insomuch as they cue activation of category-distinguishing features, can guide the learning process in the same way that conventional labels like 'round' and 'pointy' guide category learning.

4.1. PARTICIPANTS, MATERIALS, AND PROCEDURE

4.1.1. *Participants*

A total of 113 participants from UW-Madison participated in exchange for course credit. Participants were randomly assigned to one of several labeling conditions: a NO-LABEL (n = 33) condition omitted labels entirely. A NONCE-LABEL condition (n = 29) used the labels ‘foove’ and ‘crelch’. For some participants (n = 14) the labels were mapped to the aliens in a congruent way, pairing the label ‘foove’ with the round-headed aliens and ‘crelch’ with the pointier / more grooved-headed aliens. The incongruent-label condition (n = 14) reversed the mapping. An additional group of participants (n = 23) was assigned to a conventional-label condition and heard the label ‘round’ associated with the round-headed aliens (Figure 1 b1–b12) and either the word *grooved* (n = 12) or *pointy* (n = 11) mapped to the pointy-headed aliens (Figure 1 a1–a12). Another conventional-label condition (n = 14) from the same UW undergraduate student pool was added subsequently. This condition reversed the assignment of the conventional labels such that ‘round’ was associated with the more pointy aliens and ‘pointy’/‘grooved’ with the round ones.

4.1.2. *Materials*

We used the same aliens as in Preliminary Experiment 2. After five blocks of training, eight (a1–a4, b1–b4) of the original sixteen aliens were removed and replaced by new ones (a9–12, b9–b12) for reasons detailed below. The species of these new aliens could also be distinguished based on the relative roundness of the head-shape, but the instantiation of roundness was somewhat different, as can be seen in Figure 1.

4.2.3. *Procedure*

The category learning procedure was similar to Lupyan et al. (2007). Participants were told to imagine that they were explorers on another planet, learning which aliens they should approach and which they should move away from. They were told that they would be guessing initially, but the correct/incorrect feedback sounds would help them learn to distinguish the two types of alien. Those assigned to the label conditions were told that previous visitors to the planet had found it useful to name the two kinds of alien, and that they should pay careful attention to the labels. On each training trial, one of the sixteen aliens appeared in the center of the screen. After 500 ms, an ‘explorer’ – a person in a space suit – appeared in one of four positions: to the left of, to the right of, above, or below the alien. Participants used a gamepad controller to move the explorer toward or away from the alien. After the explorer character approached or moved away from the alien, participants heard accuracy feedback (a buzz or bleep sounded after the alien moved toward/away from the explorer, about 500 ms after the response). In the labeling conditions, participants then heard an auditory label. In the no-label condition, no label was played. To ensure that participants in the label conditions paid attention to the labels and to have a measure of label learning, two verification trials were included after each training block. On these trials, one of the aliens would appear together with a written label, and participants had to respond ‘yes’ or ‘no’ depending on whether the alien and its label matched. No feedback was provided for verification trials. The pairing of the labels with the categories (move away vs. move toward) and with the alien species (more rounded / more pointy / grooved) was counterbalanced between participants.

As mentioned above, part way through training, half of the aliens were removed from the training set and replaced by new ones. This substitution had two goals: first, it allowed us to determine the specificity of learning. A larger decrease in performance following the switch would indicate greater reliance on memory for specific items rather than more general regularities. Second, the rate at which the new items were learned indicated how quickly participants in the various labeling conditions could integrate new exemplars into the newly learned categories.

4.2. RESULTS AND DISCUSSION

We focus our initial analysis on the first four labeling conditions: no-labels, nonce-labels (congruent vs. incongruent), and congruent conventional labels. Categorization accuracy across block and conditions is shown in Figure 3. Category learning (overall percent correct) was reliably affected by the labeling condition ($F(3,81) = 5.88$, $p = .001$). Planned comparisons showed that participants assigned to the congruent nonce-label condition outperformed those in the incongruent nonce-label condition ($F(1,27) = 4.65$, $p = .04$), as well as those in the no-label condition ($F(1,46) = 12.05$, $p = .001$).

Categorization accuracy in the two conventional-label conditions (round/pointy and round/grooved) was comparable ($F < 1$), and we collapsed the ‘pointy’ and ‘grooved’ conditions into a single conventional-label group. These participants outperformed those in the no-label condition ($F(1,54) = 10.53$, $p = .002$), but performed at a comparable level to the congruent-label condition ($F < 1$). The no-label condition and incongruent nonce-label condition were likewise not reliably different from one-another ($F(1,45) = 1.62$, $p > .2$). The performance advantage in the congruent nonce-label and conventional-label conditions over the no-label condition was also present for the new stimuli introduced on block 6 (Figure 1 a9–a12, b9–b12). However, the congruent-label condition no longer showed a reliable advantage over the incongruent-label condition ($t(26) = 0.9$, $p = .3$), indicating that the overall advantage for the congruent over the incongruent labels did not extend to these new stimuli, although the congruent-label advantage was numerically present in both cases.

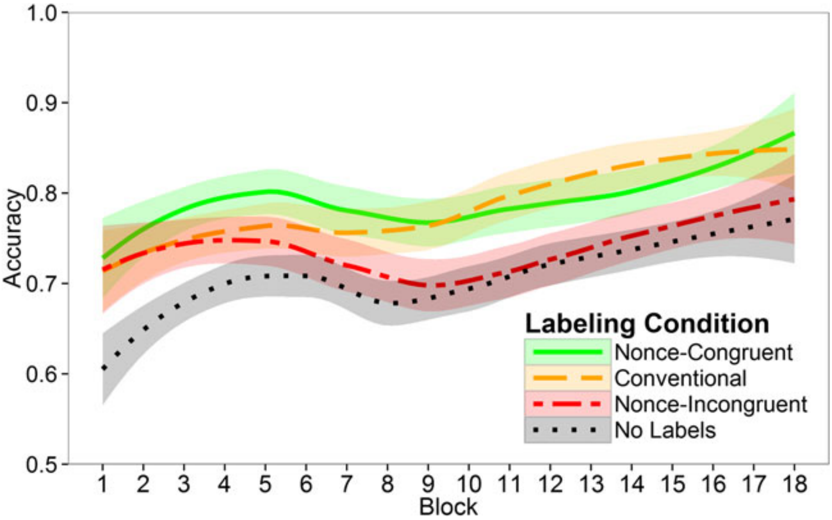

Fig. 3. Categorization performance for each block for the four main labeling conditions. New stimuli were introduced on block 6. The lines are smoothed using loess. Confidence bands show the 95% CI for each condition.

For the analysis described above we used the standard general linear model (GLM) approach. However, mixed-effects models offer considerable advantages to standard GLMs in their power and flexibility (Baayen, Davidson, & Bates, 2008), particularly when analyzing dichotomous variables. In these cases, the use of mixed-effects models with logistic regression allows for modeling of individual responses rather than collapsing each subject to a single estimate, as well as modeling performance over time without collapsing each block to a single point (Jaeger, 2008; Winter, 2013).

Mixed-effects model analysis of the basic contrasts above produced comparable results to the conventional GLM analysis described above: Accuracy analysis using logistic regression showed that congruent-labels led to better performance than incongruent-labels ($z = 2.191$, $p = .03$). The difference remained reliable with stimuli instead of subjects as a random effect ($z = 6.24$, $p < .0001$), and with both stimuli and subjects as random effects ($z = 2.16$, $p = .031$). In the more complex analyses below we utilize mixed-effects analysis and report the z-statistic with the corresponding p-value, and the chi-squared (χ^2) statistic for comparing models with/without additional predictors, as necessary.

Our next question was how labels affected the rate at which participants learned the alien categories. The response (correct/incorrect) for each trial was entered into a mixed-effects logistic regression with block and labeling-condition as predictors. Performance on the original and novel stimuli was analyzed separately. Performance for the original stimuli of course improved over time ($z = 5.97$, $p < .0005$), but the speed of learning was affected by the labeling condition as shown by a highly reliable label-condition × block interaction ($\chi^2(3) = 15.78$, $p = .001$; model comparisons). This interaction was driven by participants in the incongruent-label trials whose initial performance paralleled that of the other labeling conditions, but then stagnated: learning in the incongruent condition was slower than in all others (ps < .01 in all cases). Performance on the new stimuli likewise improved over time ($z = 3.49$, $p = .001$), but there were no reliable differences in the rate of learning for the new stimuli as a function of the labeling condition ($p > .1$).

Recall that in block 6, half of the original stimuli were removed and replaced by new ones. We reasoned that the introduction of new stimuli would disrupt performance on the original stimuli. This is exactly what happened. For example, average performance on the original stimuli increased by 3.3% from block 3 to block 5 ($t(84) = 2.23$, $p = .029$), but fell by 3.5% from block 5 to block 7 ($t(84) = 2.12$, $p = .037$). A linear model showed that introducing novel items significantly reduced the rate of learning ($z = 3.22$, $p = .0013$). We next examine whether this disruption was affected by the labeling condition. Insofar as labels improve category learning, they may also produce category representations that are more resistant to interference from novel items. To test this hypothesis, we compared a series of logistic mixed-effect models using accuracy as the dependent variable. In comparison to a base model that included block, labeling condition, and a variable indicating whether the new stimuli had been introduced, including an interaction between block and labeling condition improved the model fit ($\chi^2(3) = 16.12$, $p = .001$).

The introduction of novel items affected learning similarly for the conventional-label and congruent-label conditions, as revealed by a significant main effect of introducing novel items on learning ($z = 2.31$, $p = .02$), but no interaction ($z < 1$). Introduction of novel items impacted learning more negatively for the incongruent-label than either conventional-label ($z = 3.51$, $p = .001$) or the congruent-label condition ($z = 2.50$, $p = .01$). Learning in the incongruent-label condition was also impacted more negatively by the introduction of novel items compared to the no-label condition ($z = 2.88$, $p = .004$).

4.2.1. *Effects of labels on inter-item variability*

Not surprisingly, some aliens were more difficult to categorize than others. For example, overall accuracy for item a5 (see Figure 1) was 85%; for item a8 it was 77%. We were interested in examining whether labels not only affected overall accuracy and sensitivity to the introduction of novel items into the category, but also robustness of learning as measured by the degree of inter-item variability. Consider a child with a nascent concept of a dog.

This concept may be sufficient for categorizing the family dog and dogs with similar surface properties, but may not be robust enough to generalize to more dissimilar breeds. One way to measure categorization robustness is through inter-item variability. Just as a more robust representation of the dog category should lead to correct classification over a larger range of different dogs (i.e., lower-inter-item classification variability) we expect that more robust category representations of the YUFO aliens should lead to measurably lower inter-item variability for the items within a given species. A comparison of models with/without a labeling-condition-by-item interaction showed that labeling condition interacted very strongly with items, i.e., the performance profile across items differed as a function of labeling condition ($\chi^2(45) = 138$, $p < .0001$). For conciseness, we focus on only a simple comparison of variability quantified by the coefficient of variation. A comparison of coefficients of variation across the four labeling conditions revealed a main effect of labeling type ($F(3,81) = 3.74$, $p = .013$). Planned comparisons showed that the coefficient of variation in the meaningful-label condition ($M = 27.2$) was reliably lower than in the no-label condition ($M = 33.3$) ($t(37) = 2.24$, $p = .03$). The coefficient of variation was marginally lower in the congruent-label group ($M = 25.5$) than the incongruent label group ($M = 30.26$) ($t(24) = 1.88$, $p = .07$). Qualitatively, the pattern of variability differences was very similar to the pattern of overall accuracy, showing that the 'easy' items were easy for everyone, but others were disproportionately more difficult for the no-label and incongruent-label groups. This result is, admittedly, exploratory and can benefit from replication in future work.

4.2.2. *Reaction times*

An analysis of reaction times (RTs) (correct trials only; RTs over 2.5 sec – 2.4 SDs – above the mean, comprising 2.3% of the data, were removed) revealed a main effect of labeling condition ($F(3,81) = 3.61$, $p = .017$). Planned comparisons showed that, somewhat unexpectedly, the RTs of people assigned to the no-label group ($M = 680$ ms) were significantly faster than those assigned to the meaningful-label group, ($M = 814$ ms) ($t(41) = 3.15$, $p = .003$). The label-congruent ($M = 735$ ms) and label-incongruent groups ($M = 724$ ms) had RTs that were in between the meaningful and no-label groups, not differing reliably from them. There was no evidence of a speed–accuracy trade-offs in that, at the subject level, response speed was uncorrelated with accuracy ($p > .9$).

4.2.3. *Verification performance*

We included verification trials to encourage participants in the labeling conditions to attend to the labels (which, after all, were entirely redundant to

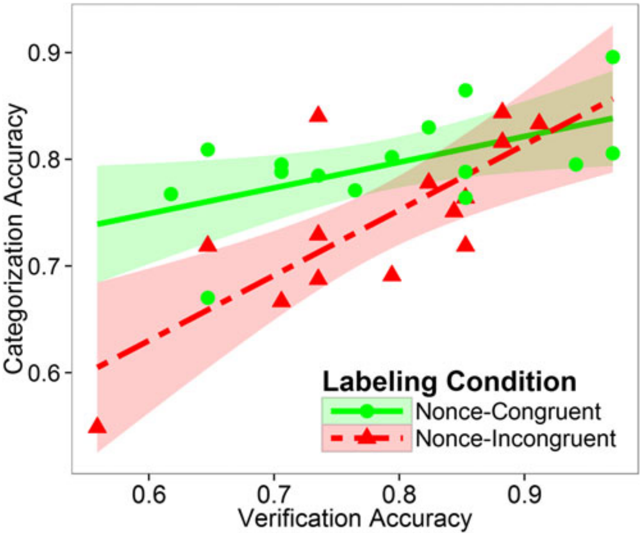

Fig. 4. Correlation plot showing categorization accuracy as a function of verification performance. Lines indicate linear fits surrounded by a 95% CI band.

the task) and to assess the learning of the alien-to-label association. Performance on the verification trials did not differ between the three labeling conditions ($M_{\text{incongruent}} = .78$, $M_{\text{congruent}} = .79$, $M_{\text{conventional}} = .83$) ($F < 1$). Not surprisingly, verification accuracy correlated with categorization accuracy ($r = 0.69$, $b = 0.47$, $p < .0005$). Interestingly, the correlation in the congruent-label condition ($r = 0.56$, $p. = .03$) was somewhat weaker than the correlation in the incongruent-label ($r = 0.76$, $p = .002$) and conventional-label ($r = 0.74$, $p < .0005$) conditions. We focus on the two nonce-label conditions because they both involved learning novel words. There was a significant difference in the relationship between verification accuracy and categorization between the two nonce-label conditions ($F(1,25) = 4.51$, $p = .04$). As can be seen in Figure 4, participants with the highest verification accuracy performed well on the categorization task regardless of the condition to which they were assigned. For the congruent-label group, categorization performance remained high even for participants who were poor learners of the labels. In contrast, slight decreases in label learning predicted much poorer categorization performance in the incongruent-label condition.

4.2.4. *Comparing congruency effects for conventional and nonce labels*

In addition to the four conditions described above, we also ran a conventional label condition in which the labels were incongruently mapped. Participants in this condition heard the same labels as in the conventional-label condition (‘round’/‘pointy’ or ‘round’/‘grooved’), but the mapping was reversed such that ‘round’ was associated with the pointy-headed aliens and vice versa. This condition allowed us to compare the effect of congruency for conventional labels (‘round’ = round-headed vs. ‘round’ = pointy-headed) to congruency effects for nonce labels (‘foove’ = round-headed vs. ‘foove’ = pointy-headed).

Participants’ overall performance in this condition (M = 73.4%) was indistinguishable from those in the nonce-incongruent condition (M = 74.2%) ($t < 1$). The difference between congruent mappings (M = 79.0%) and incongruent mappings (M = 73.4) for conventional labels was highly reliable, as shown by a mixed-effect logistic regression ($z = 6.8$, $p = .008$). There was no overall effect of label-type (conventional vs. nonce) ($z = 0.001$, $p = .99$), and congruency and label-type did not interact ($z = 0.22$, $p = .83$).

Although there were no differences in overall performance between the nonce-incongruent and conventional-incongruent conditions, the effects of congruency for nonce and conventional labels were not quite identical. Compared to incongruent nonce labels, incongruent conventional labels were harder to learn, as shown by significantly lower verification performance in this condition (M = 72.9%) compared to the congruent-conventional label condition (M = 82.6%) ($F(1,35) = 4.18$, $p = .048$). More importantly, we found a highly reliable three-way interaction between block, label-type, and congruency ($z = 3.79$, $p < .001$). This interaction can be unpacked in the following way: When conventional words like *round* and *pointy* are used incongruently, they severely impact INITIAL performance – much more so than nonce labels used incongruently. This is not surprising considering that the meanings of ‘round’ and ‘pointy’ are much more entrenched and well-specified than the meanings of ‘foove’ and ‘crelch’. Over the course of training, performance in the incongruent-conventional label condition catches up with the congruent-conventional label and congruent-nonce label conditions. Explaining the different dynamics of these real- and nonce-word congruity effects requires further research.

5. Control experiments determining whether ‘foove’ and ‘crelch’ activated shape information via mediation through conventional words

We have argued that our results provide evidence that nonce words like *foove* and *crelch* are ‘cues-in-context’ (Casasanto & Lupyan, in press) that activate shape information on the basis of their sounds. An alternative is that the effects arise solely through mediation by existing words. For instance, if people are asked the meaning of *turple* they might respond with ‘purple turtle’ – a clear indication of phonological mediation. To test this alternative explanation, we ran three additional experiments. In Experiment C1, people were asked to generate phonological neighbors of *foove* and *crelch*. The results showed that none of the phonological neighbors generated for *foove* and *crelch* denote shapes that could explain attribute choices in Preliminary Experiment 1 and effects of the nonce words on categorization in Experiment 1. In Experiment C2, people were asked to describe the differences between the two alien species displayed in Figure 1 that either included the nonce labels or omitted the labels to check whether inclusion of the labels led to phonologically related descriptors. The results showed that people's descriptions were not affected by the nonce labels. Finally, in Experiment C3 people were asked to generate definitions of 'foove' and 'crelch'. The results showed that in this relatively unconstrained context most people DID rely on phonological neighbors, e.g., the most frequently mentioned phonological neighbor of *crelch* was *belch*, and many of the provided definitions of 'crelch' relate to bodily functions (also mediated by *retch*). Thus, although people often rely on phonological neighbors when generating definitions of nonce words, the evidence presented here offers no support for the possibility that such verbal mediation is responsible for the effects reported in the main experiment.

5.1. EXPERIMENT C1: PHONOLOGICAL NEIGHBOR GENERATION TASK

If the nonce words *foove* and *crelch* guide behavior solely through the activation of conventional words via, for instance, phonological priming, we would expect these words to have at least some phonological neighbors having meanings similar to the attributes used in Preliminary Study 1.

5.1.1. *Participants, materials, and procedure*

Eighteen participants were recruited via Amazon Mechanical Turk and asked to provide real words that sounded like *foove* and *crelch* (i.e., a neighbor generation task). Participants were instructed to: "Consider the sound of the nonsense word 'foove' ['crelch']. What *real* words does it sound like? Please enter at least 5 words." Participants were not instructed about what counts as a phonological neighbor.

5.1.2. *Results and discussion*

Table 1 lists all the responses that were provided by more than one participant. Notice that NONE of these words describe shape or even visual features. These words represent those that were mentioned at least twice. We also examined the remainder of the 185 unique responses. Of these, only two

TABLE 1. *Results from the neighbor generation task of Experiment C*1

Foove	n	Crelch	n
move	13	belch	15
groove	11	welch	8
hoove	11	mulch	7
love	8	crutch	6
prove	7	fetch	6
dove	4	couch	5
food	4	crouch	5
foot	4	wretch	5
shove	4	felch	4
hoof	3	crawl	3
soothe	3	crotch	3
tube	3	crunch	3
you've	3	grouch	3
behoove	2	squelch	3
booth	2	welsh	3
few	2	belt	2
fool	2	bets	2
froze	2	felt	2
grove	2	fence	2
hooves	2	filch	2
improve	2	filtch	2
jews	2	flech	2
jove	2	melts	2
phew	2	well	2
rove	2		
shoes	2		

were remotely germane to the shape distinction in question – one person listed *smooth* as one of the neighbors of *foove*, and one person listed *crevice* as one neighbor of *crelch*. These results show that when asked to generate phonological neighbors, only very rarely do people come up with any conventional words that are relevant to the perceptual distinction between the two species of alien used in the categorization task.

5.2. EXPERIMENT C2: LABEL-CUED ALIEN DESCRIPTIONS

An additional way to test the possibility that the nonce words guided performance in Preliminary Experiment 2 and Experiment 1 through mediation by conventional words is by asking people to generate conventional words describing the differences between the two species and to compare responses in a condition where the alien species are accompanied by the novel labels ‘foove’ and ‘crelch’ to a condition where they are not. If the reason people in Preliminary Study 2 thought that ‘fooves’ were the round and smooth aliens and ‘crelches’ were narrow and ridged aliens was that the nonce words phonologically primed conventional words with related meanings, then we should find a greater proportion of phonological neighbors in the description of relevant shape distinctions of the alien species that are labeled with the nonce words.

5.2.1. *Participants, materials, and procedure*

Fifty participants were recruited from Amazon Mechanical Turk. Each person was shown the sixteen original aliens (Figure 1 a1–a8, b1–b8) in two vertically arranged grids and asked to describe what distinguished the two species. Participants were encouraged to examine the two species closely to notice the features that best distinguished them.

Participants were randomly assigned to one of two conditions (n = 25 in each). In the NO-LABEL condition, the two species were referred in a general way: "Here is the first species … Here is the second species." In the LABEL condition, the descriptions included the labels 'foove' and 'crelch'. The query instructed participants to "List between 2 and 5 ways in which the top species of aliens [fooves] is different from the bottom species [crelches], e.g., if you think the aliens in the top group are larger, you would put 'larger' for the Top species and 'smaller' for the Bottom species. Scroll the page back and forth as necessary to compare the two groups."

5.2.2. *Results and discussion*

Altogether, participants generated 237 unique responses, of which 44% were single-word adjectives such as *smaller*, *angular*, *meaner*, *narrower*. The rest were multiword descriptors such as *thicker heads*, *spread out*, *more dynamic in head size*, and *long narrow head*.

Each word was checked against the phonological neighbor list generated in Experiment C1 (substituting base forms of words as necessary to maximize matching, e.g., *narrow* and *narrower* were coded as the same response). In all the responses, only a single word was a phonological neighbor: one person wrote "less platform under foot" for one of their descriptions (foot being one of the neighbors listed for *foove*).

Next, we sought to check whether ANY words were mentioned at different rates in the two conditions. To do this, we matched the words that were used multiple times in both the label and no-label conditions and compared their relative frequencies. For example, the word *fat* was used 3 times in the no-label condition and 2 times in the label condition. For each word pair, we computed a difference score (label – no-label). The frequencies for the two conditions had modal and median values of 0 (M = .87, SD = 3.91). However, one word stood out. The word *head* was mentioned 32 times in the label condition, but only 11 times in the no-label condition, a difference that is reliable by a 2-proportion test ($z = 2.95$, $p = .003$). It is the head that contains the most distinguishing information, and it was explicitly mentioned more often in the label condition. We did not predict this result and it should thus be treated with caution. If it turns out that the mere presence of incidental labels is sufficient to push people toward more category-relevant properties, such a result would be consistent with previous demonstrations that more prolonged exposure to category labels leads to better abstraction over idiosyncratic properties and more robust representation of category-typical/category-diagnostic properties (Loewenstein & Gentner, 2005; Perry & Lupyan, in press; Lupyan, 2008, and see 2012 for review and a computational model).

When allowed to examine all the category items simultaneously, participants mentioned all the same visual features. Using the labels 'foove' and 'crelch' did not lead people to mention any specific features phonologically related to *foove* or *crelch*. Once again, these results fail to find evidence for the verbal mediation hypothesis.

5.3. EXPERIMENT C3: MEANING GENERATION TASK

Experiments C1–C2 do not find support for the hypothesis that the shape information cued by *foove* and *crelch* in our main studies can be explained through explicit mediation via conventional words. A potential critique is that the meaning of *foove* and *crelch* may indeed derive from the nonce words activating conventional words, but our methods are simply not sensitive to discover such effects. In Experiment C3, we show that, under some circumstances, the responses participants provide to our nonce words ARE verbally mediated in a highly transparent way. Critically, these verbally mediated responses are not related to the visual properties that distinguished the two alien species.

5.3.1. *Participants, materials, and procedure*

Eighteen participants were recruited from Amazon Mechanical Turk and asked to provide a definition for the words *foove* and *crelch*. The exact text of the query was: "If the word 'foove' ['crelch'] were a real word, what do you think it would mean? Please provide a definition below. Just type the first definition that comes to mind."

5.3.2. *Results and discussion*

Responses are shown in Table 2, with each row representing an individual participant. Notice that only 2 of 36 definitions (starred) explicitly mentioned

TABLE 2. *Results of Experiment C3 (each row shows the definitions generated by a separate participant)*

Foove	Crelch
fake, a copy of	crabby, anti-social
It would be something to do with either animals or food. Maybe something made with animal parts.	It would be something medical, maybe intestinal or some kind of stomach illness?
*a very fat foot	a bug
a type of pottery	a type of underwear
a funky groove	a deep sea crustacean
someone who acts like a moron	a pedal that makes a car stop
the bottom of a horse's foot	a hole dug in the ground
to have a favorite	to crouch down low
a rude name meant to insult the person it is being aimed at	an unpleasant noise made when one is sick
a fake love; i.e., those teenagers are in foove with each other but really they just like the way the other looks	to vomit, i.e., he drank so much that he crelched the next day
a type of soil sediment	a small crustacean-like animal
to get punched in the gut so hard that the wind gets knocked out of you	to throw up after having cramps
Foove would mean the edge of something where an object can fall off.	It would mean the sound of metal screeching and squeaking against another piece of metal.
I think that it would be the description of food in general but in the plural sense.	I think it would be a description of a person from the historical times of ancient England
a split in a horse's hoof	*a type of pottery with colorful glazes and patterns
It would refer to a horse hoove that is furry.	It would mean when one coughs and then belches.
A nonsensical word by a toddler that is trying to annoy their parents. They say it over and over and over.	a crunchy belch where you find food in your mouth
odd name for a bathroom	weird belch

NOTE: Asterisks highlight definitions that make explicit mention of perceptual properties.

ANY shape information or visual features. Notice also that many of these explicit definitions have clear phonological mediators. For example, the definitions of *crelch* as 'some kind of stomach illness', 'an unpleasant noise made when one is sick', and 'to vomit' are likely to be mediated by *retch* and, to a lesser degree, *belch*. One person provided a definition that is apparently mediated simultaneously by *retch* and *cramps* ('to throw up after having cramps'). Two people mentioned crustaceans, likely mediated by *krill*, *shrimp*, and the word *crustacean* itself. The meanings of *foove* that mention hooves, fakeness, and a groove also have clear phonological mediators. We do not know how aware participants are that the meanings they generate are phonologically mediated. There are likely substantial individual differences and we think the definitions listed in Table 2 reflect a combination of strategic inferences and more automatic responding.

These results show that in some contexts, people's responses to *foove* and *crelch* ARE transparently mediated by phonological neighbors. As we observed, the neighbors people list in Experiment C1 are unrelated to visual features used to distinguish the two species of alien. The verbally mediated definitions that people generated here appear to be similarly unrelated.

6. General discussion

Successful category learning requires the learner to selectively represent the features or attributes that are shared within a category and contrasted between categories. The particular category structure used here required learners to attend to shapes characteristic of the 'aliens'. Not surprisingly, cues that guide learners toward distinguishing features – words like *round* and *pointy* – facilitated learning. But strikingly, the novel and putatively meaningless words *foove* and *crelch* not only helped learning, but helped as much as conventionally meaningful words (see Figure 3). These results suggest that word forms do not need to correspond to memorized senses or referents to be meaningful. As such, they call into question the necessity of the 'mental lexicon' as a construct (Casasanto & Lupyan, in press; Dilkina, McClelland, & Plaut, 2010; Elman, 2004, 2009, 2011).

One of the main ways in which *Jabberwocky* words are made meaningful is through syntactic support (Johnson & Goldberg, 2013). Syntax does not contribute to the meanings of our nonce words because we presented the words in isolation. Our effects also cannot be explained by participants consciously activating conventional words that are phonologically similar to the nonce words we used. When asked to provide phonological neighbors of *foove* and *crelch* participants showed no tendency to generate neighbors that described rounded or pointy shapes, respectively (see Experiments C1–C3).

Foove and *crelch* thus appear to activate roundness and pointiness via implicit links between their constituent sounds and mental representations of smooth and pointy shapes, links that may reflect direct cross-modal sound-to-shape correspondences. Our findings are consistent with work showing that linguistic mappings that take advantage of such links, i.e., Japanese mimetics, lead to more robust generalization of newly learned words both for Japanese- (Imai et al., 2008), and English-speaking children (Kantartzis et al., 2011), and that adults can use them productively in forming nonce words (Thompson & Estes, 2011).

The origin of these cue-to-meaning links is still unknown. One possibility is that they arise from correlations between certain speech sounds and dynamics of human articulators, e.g., the association between *bouba* and roundness may stem from the mouth shape required to articulate the word (Ramachandran & Hubbard, 2001). Such associations, which have been demonstrated in early infancy (Ozturk, Krehm, & Vouloumanos, 2013), may be learned by infants, or may reflect innate correspondences. Some of these synesthetic-type correspondences may reflect common neural coding of certain auditory and visual dimensions. For example, the behavioral correspondence between low pitch and dark shapes has been found in non-human primates (Ludwig, Adachi, & Matsuzawa, 2011; see Spence, 2011, for review of attested audio–visual correspondences), and may reflect deep similarities between coding of higher-energy sensory events (more light, greater auditory frequency) in the nervous system. Another intriguing possibility is that sound symbolic cue-to-meaning links arise from learning non-arbitrary associations in our environment. For example, people are remarkably good at telling shape from sound (e.g., Kunkler-Peck & Turvey, 2000), perhaps as a result of frequent opportunities to associate various visual forms with sounds, e.g., sounds made by touching certain textures, impact sounds made by differently shaped objects, correlations between movement speed and sound variation (cf. Shintel & Nusbaum, 2007), and so on. It is conceivable that the meaningfulness of nonce words like *foove* and *crelch* may reflect knowledge that is generalized from learning such mappings, so that the reason *crelch* leads to activation of narrowness and spikiness is that its acoustics overlap in some way the acoustics of auditory events that index narrow/spiky entities. The mechanisms summarized above are not mutually exclusive, and may operate simultaneously.

In the present case, the cue-to-meaning links may also be mediated by spreading activation from the sounds contained in the nonce words to similar sounds contained in English words – some of which have smooth- and pointy-shaped referents. Even if this is the case, our data suggest that people are unaware of activating these shape-relevant phonological neighbors; *foove* may sound like *smooth* upon reflection, and *crelch* may sound like *crevice*, but participants do not generate these words on their own, as shown by Experiments C1–C3. If such activation occurs, it is implicit rather than strategic.

It is important to note that non-arbitrary word-to-meaning mappings are not just a lab-based phenomenon. Although rarer in Indo-European languages, non-arbitrary mappings are common throughout the world (e.g., Dingemanse, 2011; Nuckolls, 2010), often described under the names of ‘expressives’, ‘ideaphones’, and ‘mimetics’ (see also Monaghan, Christiansen, & Fitneva, 2011; Monaghan, Mattock, & Walker, 2012 for a discussion of non-arbitrariness at the word-class level). An often overlooked property of such systems is RELATIVE ICONICITY. For example, in Siwu, increased protrusion (e.g., of a stomach) can be signaled in a graded way by backing and lowering a vowel: *pimbilii* → *pumbuluu* → *pɔmbɔlɔɔ* (/i/ is smallest, /ɔ/

is largest protrusion) (Dingemanse, 2011). Such gradations may reflect experience with auditory mappings in the non-verbal domain (environmental sounds, impact sounds, etc.), which, unlike most linguistic associations, are highly motivated (Edmiston & Lupyan, 2013).

Our findings extend the literature on sound symbolism by showing, for the first time, that the sound properties of nonce words can guide the learning of novel categories in a way that is comparable to using conventional words which cue participants to activate the features relevant to the category distinction. These results also challenge traditional theories of word meaning, according to which word forms yield meanings by accessing memorized form–meaning mappings in a mental lexicon (Jackendoff, 2002). On such accounts, nonsense words like *foove*, lacking lexical entries, cannot be meaningful. Our results are, however, compatible with theories on which words cue activation of a context-appropriate network of information, operating "in the same way [as] other kinds of sensory stimuli: [by acting] directly on mental states" (Elman, 2004, p. 301). Especially given a constraining context, both novel and conventional word forms can prompt readers to construct mental representations that are sufficiently similar across time and individuals to predictably guide behavior.

We have argued that *foove* and *crelch* have meanings insofar as they guide behavior in predictable ways. But are the meanings cued by *foove* and *crelch* of the same sort that are cued by conventional words? We propose that processes by which both nonce and conventional words cue readers or listeners to construct mental representations are the same; what differs may be the conventionality of the word forms, the depth of the network of information that is activated, and the amount of contextual support needed for word forms to guide the construction of meanings reliably across instances and across people. For example, in our alien categorization experiment, *foove* and *crelch* reliably activated shape information in part because participants were processing these labels in the context of categorizing the aliens based on their shapes.

An obvious difference between nonce and conventional words is familiarity. The meanings of conventional words are constrained by our history of using the same word repeatedly, in different contexts. Yet the meanings that a given word form cues the user to construct may differ from one instance to the next depending on the linguistic and extralinguistic context, sometimes subtly and other times dramatically (Casasanto & Lupyan, in press; Clark, 1983), leading scholars like William James to conclude that the notion of a context-invariant concept or word meaning is a "mythological entity" (1890, p. 230). If word meanings are neurocognitive responses to cues (i.e., the word forms), and these cues operate in an ever-changing context, then perhaps the meanings of real words, like those of nonce words, are always constructed FOR THE NONCE.

REFERENCES

Aveyard, M. (2012). Some consonants sound curvy: effects of sound symbolism on object recognition. *Memory & Cognition*, **40** (1), 83–92.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59** (4), 390–412.

Bloom, P., & Keil, F. C. (2001). Thinking through language. *Mind & Language*, **16** (4), 351–367.

Cabrera, A., & Billman, D. (1996). Language-driven concept learning: deciphering Jabberwocky. *Journal of Experimental Psychology—Learning Memory and Cognition*, **22** (2), 539–555.

Carroll, L. (2010 [1871]). *Through the looking glass and what Alice found there*. London: Penguin Books Limited.

Casasanto, D., & Lupyan, G. (in press). All concepts are ad hoc concepts. In Margolis, E. & Laurence, S. (Eds.), *Concepts: new directions*. Cambridge, MA: MIT Press.

Clark, H. H. (1983). Making sense of nonce sense. In G. B. F. d'Arcais & R. J. Jarvella (Eds.), *The process of language understanding* (pp. 297–331). New York: Wiley-Blackwell.

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2010). Are there mental lexicons? The role of semantics in lexical decision. *Brain Research*, **1365**, 66–81.

Dingemanse, M. (2011). Ezra Pound among the Mawu: ideophones and iconicity in Siwu. In P. Michelucci, O. Fischer, & C. Ljungberg (Eds.), *Semblance and signification* (pp. 39–54). Amsterdam: John Benjamins Publishing Company.

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.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Sciences*, **8** (7), 301–306.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science*, **33** (4), 547–582.

Elman, J. L. (2011). Lexical knowledge without a lexicon? *The Mental Lexicon*, **6** (1), 1–33.

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

Evans, V. (forthcoming). *The language myth: uncovering the true nature of language, mind and being human*. Cambridge: Cambridge University Press.

Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, **20** (3-6), 507–523.

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, **109** (1), 54–65.

Jackendoff, R. S. (2002). *Foundations of language: brain, meaning, grammar, and evolution*. Oxford: Oxford University Press.

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, **59** (4), 434–446.

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

Johnson, M. A., & Goldberg, A. E. (2013). Evidence for automatic accessing of constructional meaning: Jabberwocky sentences prime associated verbs. *Language and Cognitive Processes*, **28** (10), 1439–1452.

Kantartzis, K., Imai, M., & Kita, S. (2011). Japanese sound–symbolism facilitates word learning in English–speaking children. *Cognitive Science*, **35** (3), 575–586.

Kohler, W. (1947). *Gestalt psychology: an introduction to new concepts in modern psychology*. New York: Liveright Publishing Corporation.

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, **114** (1), 19–28.

Kunkler-Peck, A. J., & Turvey, M. T. (2000). Hearing shape. *Journal of Experimental Psychology. Human Perception and Performance*, **26** (1), 279–294.

Loewenstein, J., & Gentner, D. (2005). Relational language and the development of relational mapping. *Cognitive Psychology*, **50** (4), 315–353.

Ludwig, V. U., Adachi, I., & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings of the National Academy of Sciences*, **108** (51), 20661–20665.

Lupyan, G. (2008). 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. (2012). What do words do? Towards a theory of language-augmented thought. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 255–297). Academic Press. Online: <http://www.sciencedirect.com/science/article/pii/B9780123942937000078>.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: labels facilitate learning of novel categories. *Psychological Science*, **18** (12), 1077–1082.

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.

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: sound−shape correspondences in toddlers and adults. *Developmental Science*, **9** (3), 316–322.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology. General*, **140** (3), 325–347.

Monaghan, P., Mattock, K., & Walker, P. (2012). The role of sound symbolism in language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **38** (5), 1152–1164.

Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology*, **28**, 225–252.

Nuckolls, J. B. (2010). The sound-symbolic expression of animacy in Amazonian Ecuador. *Diversity*, **2** (3), 353–369.

Nygaard, L. C., Cook, A. E., & Namy, L. L. (2009). Sound to meaning correspondences facilitate word learning. *Cognition*, **112** (1), 181–186.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: evidence for sound−shape cross-modal correspondences in 4-month-olds. *Journal of Experimental Child Psychology*, **114** (2), 173–186.

Parise, C., & Spence, C. (2012). Audiovisual crossmodal correspondences and sound symbolism: a study using the implicit association test. *Experimental Brain Research*, **220** (3), 319–333.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in Psychology*, **1**. doi:10.3389/fpsyg.2010.00227.

Perry, L. K., & Lupyan, G. (in press). The role of language in multi-dimensional categorization: evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*.

Pullum, G. K. (1989). The great Eskimo vocabulary hoax. *Natural Language & Linguistic Theory*, **7** (2), 275–281.

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia–a window into perception, thought and language. *Journal of Consciousness Studies*, **8** (12), 3–34.

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: the ARC - nonword database. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, **55** (4), 1339–1362.

Shintel, H., & Nusbaum, H. C. (2007). The sound of motion in spoken language: visual information conveyed by acoustic properties of speech. *Cognition*, **105** (3), 681–690.

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

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception & Psychophysics*, **73** (4), 971–995.

Taylor, J. R. (2012). *The mental corpus: how language is represented in the mind*. Oxford and New York: Oxford University Press.

Thompson, P. D., & Estes, Z. (2011). Sound symbolic naming of novel objects is a graded function. *Quarterly Journal of Experimental Psychology*, **64** (12), 2392–2404.

Westbury, C. (2005). Implicit sound symbolism in lexical access: evidence from an interference task. *Brain and Language*, **93** (1), 10–19.

Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. *arXiv:1308.5499 [cs]*. Online: <http://arxiv.org/abs/1308.5499>.