Language and Cognition (2020), Page 1 of 22. doi:10.1017/langcog.2019.49
© UK Cognitive Linguistics Association, 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Playful iconicity: structural markedness underlies the relation between funniness and iconicity*

MARK DINGEMANSE

Centre for Language Studies, Radboud University, Nijmegen

AND

BILL THOMPSON

University of California, Berkeley

(*Received* 30 *September* 2019 – *Revised* 02 *December* 2019 – *Accepted* 03 *December* 2019)

ABSTRACT

Words like 'waddle', 'flop', and 'zigzag' combine playful connotations with iconic form–meaning resemblances. Here we propose that structural markedness may be a common factor underlying perceptions of playfulness and iconicity. Using collected and estimated lexical ratings covering a total of over 70,000 English words, we assess the robustness of this association. We identify cues of phonotactic complexity that covary with funniness and iconicity ratings and that, we propose, serve as metacommunicative signals to draw attention to words as playful and performative. To assess the generalisability of the findings we develop a method to estimate lexical ratings from distributional semantics and apply it to a dataset 20 times the size of the original set of human ratings. The method can be used more generally to extend coverage of lexical ratings. We find that it reliably reproduces correlations between funniness and iconicity as well as cues of structural markedness, though it also amplifies biases present in the human ratings. Our study shows that the playful and the poetic are part of the very texture of the lexicon.

KEYWORDS: playfulness, iconicity, lexical ratings, markedness, semiotics.

[*] Address for correspondence: m.dingemanse@let.ru.nl

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

"This is play."
(Bateson, 1955)

1. Introduction

Iconic words are widespread in natural languages (Nuckolls, 1999; Perniss, Thompson, & Vigliocco, 2010), and scholars working on them have long drawn attention to their expressive and playful nature (Samarin, 1970; Jakobson & Waugh, 1979; Klamer, 2002). However, empirical studies of when and why some words appear more playful and performative than others are rare. Here we study the intersection of iconicity and playfulness using new data on funniness and iconicity for thousands of English words. We propose that structural markedness underlies both funniness and iconicity, and test this theory by combining linguistic analysis with quantitative evidence from human lexical ratings. We also introduce and benchmark a method for estimating lexical ratings on the basis of distributional semantics, allowing us to test the generalizability of our proposals. The method is applicable more generally to the task of substantially increasing the intersection between sets of lexical ratings.

Substantial numbers of iconic words are found in many of the world's languages, often in the form of an open lexical class of ideophones, but also scattered across the lexicon as sensory words that show phonaesthetic form–meaning associations (Nuckolls, 1999; Dingemanse, 2019). The marked phonology of iconic words has been connected to playful and expressive functions of language (Samarin, 1970; Zwicky & Pullum, 1987; Kunene, 2001; Haiman, 2014), and ideophones have been defined – only partly tongue-in-cheek – as "those words which are such fun to use" (Welmers, 1973). In an independent strand of research, people have recently started to investigate the perceived funniness of word forms (Westbury, Shaoul, Moroschan, & Ramscar, 2016; Engelthaler & Hills, 2018). One aim of this paper is to make these worlds meet. Playfulness and iconicity are pervasive features of language. In investigating them together, this paper seeks to contribute to a recentering of linguistics, which has focused mostly on the referential function of language to the neglect of its poetic, expressive, and other functions (Jakobson, 1960).

1.1. Research questions and theoretical background

What makes people think of words as iconic? What makes people think of words as funny? And is there a relation between the two? These questions are motivated by prior work on the link between playfulness and performativity in language and communication (Fortune, 1962; Samarin, 1970; Dingemanse, 2011). For instance, ideophones and other forms of expressive language often show elements of phonetic and linguistic play, drawing attention to themselves for purposes of dramatisation and entertainment (Samarin 1970). Likewise, puns

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

and word plays are characterised by the use of linguistic material for aesthetic purposes (Jakobson & Waugh, 1979). Recent work suggests that words are rated as funnier when they have improbable orthographic or phonological structure (Westbury & Hollis, 2019). We propose that perceptions of words as iconic and/or funny may be underpinned by a shared semiotic mechanism: foregrounding by means of structural markedness.

In linguistics, FOREGROUNDING has been defined as the use of linguistic signs "in such a way that this use itself attracts attention" (Havránek, 1964, p. 10). Foregrounding in this sense can be achieved in several ways, including lexical choice, prosody, and most importantly for present purposes, by STRUCTURAL MARKEDNESS: formal properties of lexical roots that make them stand out from other words. Work on iconicity has shown that iconic words often show such structural markedness in the form of phonotactic patterns and structures that deviate from other segments of vocabulary (Samarin, 1970; Klamer, 2002; Nuckolls, Nielsen, Stanley, & Hopper, 2016). These special formal characteristics help signal their special status as depictions (Nuckolls, 1999; Dingemanse, 2019). In semiotic terms, structural markedness can serve as a meta-communicative signal that draws attention to the word *qua* word and thereby invites language users to treat it as playful, poetic, and performative.

Behind the linguistic sense of foregrounding lies theoretical work in human ethology and sociology, according to which metacommunicative signals can frame strips of behaviour as "play" versus "not play" (Bateson, 1955) or as "nonserious" versus "serious" (Goffman, 1974). Bateson suggested that this metacommunicative distinction marks a major transition in the evolution of communication. Goffman showed its relevance in everyday social interaction, where we regularly combine serious actions with acting, playing, and pretending. This brings into view a deeper conceptual connection between playfulness and iconicity: both belong to a world of make-believe where words are valued for their performative character as much as their informative content.

While funniness and iconicity have been connected conceptually, their relation has not been studied empirically in a large dataset. This is what we do here using lexical ratings for thousands of words. Databases of lexical norms have long been used to achieve experimental control and model psycholinguistic processes. The growing number of properties and dimensions for which norms are available makes such resources increasingly important in quantitative studies of many fundamental questions in the language sciences (Winter, 2019). For instance, cross-linguistic collections of iconicity ratings can be used to better understand modality-specific affordances for iconicity (Perlman, Little, Thompson, & Thompson, 2018); and ratings of affective meaning can be investigated for their relation to phonetic and sublexical measures of affect (Aryani, Conrad, Schmidtke, & Jacobs, 2018).

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

With many sets of lexical ratings within easy reach, it is important to understand their affordances and limitations (Motamedi, Little, Nielsen, & Sulik, 2019). In sufficiently large datasets, almost any combination of lexical ratings will show some correlation. This makes it important to constrain analytical degrees of freedom by means of theory. The theory-driven proposal of this paper is that foregrounding, achieved through structural markedness, unites playfulness and iconicity. This implies two predictions for the kind of lexical data we study: (i) high iconicity ratings and high funniness ratings should go hand in hand; and (ii) words rated high in funniness and iconicity should show relatively larger degrees of structural markedness. Although we test these predictions using lexical data from English, given the generality of the account, we expect the findings to hold across a wide range of languages. We see this study therefore as improving our theoretical and empirical grasp of the relation between playfulness and iconicity.

2. Methods and materials

Our starting point is the intersection of recently published funniness ratings (Engelthaler & Hills, 2018) and iconicity ratings (Perry, Perlman, Winter, Massaro, & Lupyan, 2017), illustrated in Figure 1. Both sets of HUMAN RATINGS have been collected by asking people to rate words on continuous scales, with every word rated by at least 10 people. For the iconicity ratings, people were asked to rate words on a scale that runs from –5 (anti-iconic or "words that sound like the opposite of what they mean") via 0 (arbitrary or "words that do not sound like what they mean or the opposite") to 5 (iconic or "words that sound like what they mean"). As Figure 1, panel B shows, the negative end of the scale was underused; subsequent analysis suggests that it was also used less consistently (Motamedi et al., 2019). The positive end of the scale successfully picked out words that show iconicity, defined (for spoken languages) as perceptual resemblances between aspects of word sound and meaning (Svantesson, 2017).

Among the items rated high in iconicity in this study are also quite a few morphologically complex words with analysable compositional structure, like 'dishwasher', 'skateboard', 'downpour', 'seaweed', 'corkscrew', 'airplane', and 'bedroom'. Morphological analysability is quite distinct from perceptual resemblances between sound and meaning (for instance, it is only accessible to those who already know the meaning of the compound elements), so such words are not actually iconic in the sense used in the rating study (Perry, Perlman, & Lupyan, 2015). However, it is easy to see why naive participants would treat them as words that "sound like what they mean". We will later see that these analysable compounds may introduce a bias that is amplified in imputed ratings.

For the funniness ratings, people were asked to rate words on a scale from 1 to 5 in terms of funniness (Figure 1, panel C). As the instructions

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

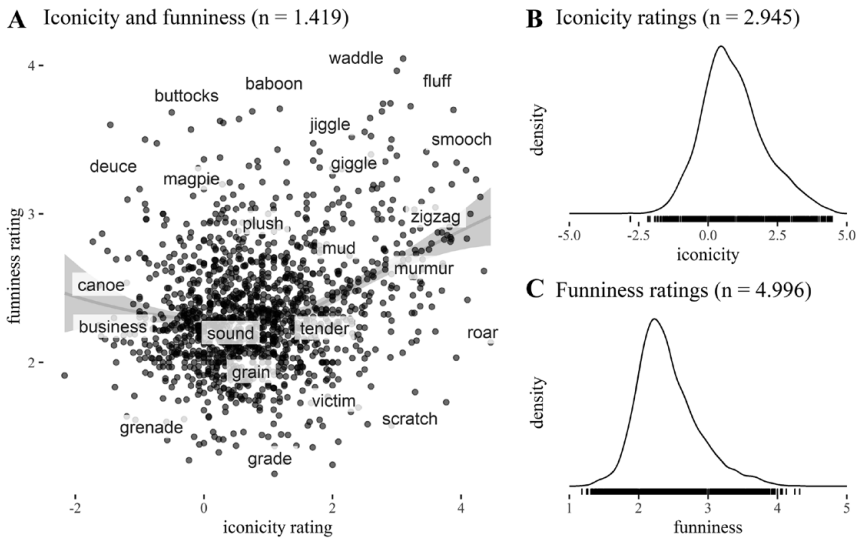

Fig. 1. The intersection of iconicity and funniness ratings for 1419 words. **A**: Scatterplot of iconicity and funniness ratings in which each dot corresponds to a word. A loess function generates the smoothed conditional mean with 0.95 confidence interval. Panels **B** and **C** show the distribution of iconicity and funniness ratings in this dataset.

mentioned, "The rating scale ranges from 1 (humourless = not funny at all) to 5 (humorous = most funny)". Because participants were instructed to interpret the scale in terms of funniness, we think the ratings are best described as "funniness ratings" rather than "humour norms" (which is what Engelthaler and Hills call them). Humour is a broad field of study: the perceived funniness of words is only one aspect of a phenomenon that ranges from the fine details of prosody and phonology (Menninghaus, Bohrn, Altmann, Lubrich, & Jacobs, 2014; Westbury & Hollis, 2019) to discourse and ethnopragmatics (Glenn, 2003; Levisen, 2018), and whose stylistic realisations include puns, allusions, jokes, and anecdotes (Dynel, 2009; Attardo, 2018).

To test the generalisability of our findings, we developed a meaning-based algorithm to estimate funniness and iconicity for any English word. The algorithm works in two steps. First, it is trained on a large corpus of natural language text. Using the lexical co-occurence statistics in this corpus, it learns semantic relationships between millions of English words (words that appear in similar contexts are treated as similar in meaning). Second, it is trained to predict the iconicity (or funniness) of words that have already been rated by experimental participants. Once it can accurately predict known ratings, it is asked to predict iconicity (or funniness) for new words. It is able to do this for virtually any new word by using the semantic relationships it learned in step one.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

For example, say the new word is 'waggle'. In step one, the algorithm learned that 'waggle' occurs in similar contexts to 'wiggle' and 'wobble'. In step two, it learned that 'wiggle' and 'wobble' were rated as highly iconic by participants. As a result, it predicts that 'waggle' will be highly iconic too. Technically, our algorithm is based on a linear regression model that predicts experimental ratings from word vectors trained on Wikipedia (Bojanowski, Grave, Joulin, & Mikolov, 2017). Similar methods have been studied elsewhere (see, e.g., Mandera, Keuleers, & Brysbaert, 2015; Hollis, Westbury, & Lefsrud, 2017; Thompson & Lupyan, 2018). By combining lexical co-occurrence statistics with funniness ratings for 4996 English words and with iconicity ratings for 2945 English words, we estimated funniness and iconicity ratings for a total of 70202 words. We call these the IMPUTED RATINGS to distinguish them from the human ratings.

The following subsets of the data will feature most prominently in the analyses below (Figure 2): set A, 1419 words that people have rated for both funniness and iconicity; set B, 3577 words for which we compare human funniness ratings with imputed iconicity ratings; and set C, 63680 words for which only imputed ratings are available. Set A allows us to establish the ground truth about the relation between iconicity and funniness ratings and about the occurrence of cues of structural markedness. Set B allows us to test whether our imputation method makes sense. Set C allows us see whether the iconicity–funniness relation holds even in words for which we have only imputed ratings, and whether the formal cues of structural markedness also show up in these words.

We supplement the data with SUBTLEX-US frequency norms (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014), removing 62 words for which no frequency data is available (1 from the funniness ratings and 61 from the iconicity ratings). We further add lexical decision times (Keuleers, Lacey, Rastle, & Brysbaert, 2012), phonotactic data from the Irvine Phonotactic Online Dictionary (Vaden, Halpin, & Hickok, 2009), and data on number of morphemes from the British Lexicon Project (Balota et al., 2007).

2.1. ANALYSIS

We conduct all analyses using R version 3.6.1 (R Core Team, 2019). The most important packages in our analysis pipeline are tidyverse (Wickham, 2017), ggplot2 (Wickham, 2016), car (Fox & Weisberg, 2019) and ppcor (Kim, 2015). For all linear models reported below, variance inflation factors are below 2, indicating no problems with (multi)collinearity, and visual inspection of Q-Q plots and residuals plotted against fitted values revealed no deviations from normality or homoscedasticity. All data and analyses are available through the online materials at <https://osf.io/7s6xc/>.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

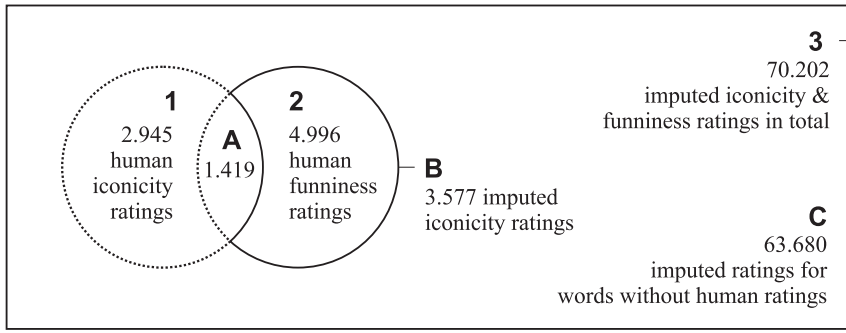

Fig. 2. Venn diagram of lexical data used in the study. Sets **1** and **2** represent human word-level ratings for iconicity (n = 2945) and funniness (n = 4996). These are also the training data for the imputed ratings in set **3**, the full set of 70202 words for which we imputed values for funniness and iconicity. The main datasets used in the analyses are set **A**, the 1419 words for which both human iconicity and human funniness ratings are available; set **B**, the 3577 words for which we have human funniness ratings but only imputed iconicity ratings; and set **C**, the 63680 words for which only imputed ratings are available.

The analysis comes in four parts. First, using human ratings, we examine the relation between funniness ratings and three other variables: iconicity ratings (our main focus), word frequency (a known covariate of both funniness and iconicity), and lexical decision time (reported by Engelthaler & Hills (2018) as the most important correlate of funniness ratings after frequency). Second, we go beyond known iconicity ratings to test the relation between funniness ratings and imputed iconicity. This is a first benchmark of the imputation method and serves to test whether the relation identified for human ratings also holds for imputed iconicity ratings. Third, we investigate the relation between imputed funniness and imputed iconicity ratings as a further test of the generalisability of the imputation method. In all these analyses, we control for frequency and lexical decision time. Finally, we investigate the structural properties of the highest rated words and inductively identify cues of structural markedness to explain the relation between funniness ratings and iconicity ratings.

3. Results

3.1. FUNNINESS AND ICONICITY

We first consider the relation of funniness ratings to frequency and lexical decision time, the two measures identified by Engelthaler and Hills (2018) as the strongest correlates for perceived funniness. Like them, we find that uncorrected correlations in the full dataset hover around 28%, with log frequency negatively correlating with funniness (less frequent words are rated as more funny) and lexical decision time positively (words with longer

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

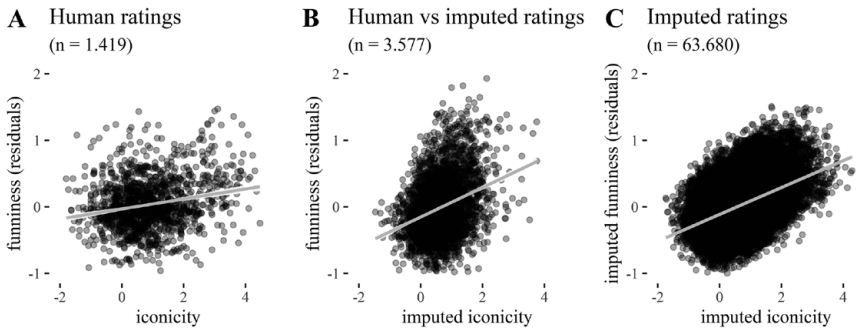

Fig. 3. Relations between funniness and iconicity after controlling for word frequency, in: **A** words with human ratings; **B** words with human funniness ratings and imputed iconicity ratings; **C** words for which we only have imputed ratings. Funniness is residualised to control for frequency, so scales on the *y*-axis are not directly relatable to the original 1–5 rating scale.

reaction times are rated as more funny). A linear model with funniness as dependent variable and frequency and lexical decision time as predictors shows a role for both, though a larger portion of the variance is accounted for by frequency ($F = 454.1$, $p < .0001$, partial $\eta^2 = 8.3\%$) than by lexical decision time ($F = 100.4$, $p < .0001$, partial $\eta^2 = 2\%$).

To assess the role of iconicity we carry out this analysis for the subset of 1419 words for which we have both iconicity and funniness ratings, and compare linear models with and without iconicity as an additional predictor. We find that, in this subset, as expected, funniness ratings are partially predicted by frequency and lexical decision time. Model comparison shows that a model including iconicity as a predictor provides a significantly better fit ($F = 63.7$, $p < .0001$) and explains a larger portion of the variance (adjusted $R^2 = 0.188$ versus 0.152). In this fuller model, while frequency remains the strongest (negative) correlate of funniness ratings ($F = 258.8$, $p < .0001$, partial $\eta^2 = 15.5\%$), iconicity is the second strongest predictor ($F = 63.7$, $p < .0001$, partial $\eta^2 = 4.3\%$), followed at some distance by lexical decision time ($F = 8.9$, $p = .003$, partial $\eta^2 = 0.6\%$).

Since iconicity is also known to bear a weak relation to word frequency (Winter, Perlman, Perry, & Lupyan, 2017), we test whether the relation between iconicity and funniness ratings is reducible to the effect of frequency using partial correlations (Kim, 2015). In set A, we find that there is 20.6% of covariance between iconicity and funniness that is not explained by word frequency: words rated higher in iconicity are still rated higher in funniness, controlling for frequency ($r = 0.206$, $p < .0001$). The relation between iconicity and funniness ratings, controlling for frequency, is depicted in Figure 3, panel A.

Table 1 shows example words from the four quadrants of the funniness and iconicity ratings space. Many highly iconic words are rated as highly funny, and many words rated as not iconic are rated as not funny. Areas where the ratings

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

TABLE 1. *Sample words from the extremes of each quadrant of funniness and iconicity ratings (total n = 1419)*

	HIGH ICONICITY	LOW ICONICITY
HIGH FUNNINESS	zigzag, squeak, chirp, pop, clunk, moo, clang, oink, zoom, smooch	belly, buttocks, beaver, chipmunk, turkey, bra, hippo, chimp, blonde, penis
LOW FUNNINESS	click, roar, crash, chime, scratch, swift, sunshine, low, break, clash	silent, statement, poor, cellar, incest, window, lie, coffin, platform, address

deviate bring to light other mediating factors. For instance, 'buttocks', 'chimp', and 'blonde' are rated as highly funny but not iconic; their funniness rating is likely derived from co-occurrence relations (e.g., appearance in joke genres) rather than from any phonological characteristics. On the other hand, highly iconic words like 'roar, 'crash', and 'scratch' are low in funniness ratings, likely because they are associated with negative events. The word 'sunshine' is an example of a non-iconic word that is likely rated as highly iconic because of its transparent compositional structure; about 10% of the top 150 nouns with high iconicity ratings are of this type.

3.2. FUNNINESS AND IMPUTED ICONICITY (KNOWN UNKNOWNS)

As a first test of the imputation method we look at the intersection of funniness ratings and imputed iconicity ratings for the 3577 words that have been human-rated for funniness but not iconicity (Figure 3, panel B). We formulate a linear model with funniness rating as the dependent variable. Model comparison shows that a model including imputed iconicity as predictor provides a significantly better fit ($F = 451.8$, $p < .0001$) and explains more than double the amount of variance (adjusted $R^2 = 0.187$ versus 0.084) than a model with just log frequency and lexical decision time. In the fuller model, imputed iconicity rises to be the strongest predictor ($F = 451.8$, $p < .0001$, partial $\eta^2 = 11.2\%$), followed by frequency ($F = 245.7$, $p < .0001$, partial $\eta^2 = 6.4\%$) and lexical decision time ($F = 127.4$, $p < .0001$, partial $\eta^2 = 3.4\%$). A partial correlations analysis shows that imputed iconicity values correlate with funniness ratings at at least the same level as actual iconicity ratings, controlling for frequency ($r = 0.32$, $p < .0001$).

Many of the words identified as high in iconicity by our imputation method (Table 2) are clearly imitative in origin, as seen for example in OED definitions like 'swish' "to make the sound expressed by 'swish'", 'chug' "a plunging, muffled, or explosive sound", 'oomph' "the quality of being exciting, energetic, or sexually attractive (imitative in origin)". Words high in funniness and low in imputed iconicity include animals ('heifer', 'sheepdog') and taboo words ('nudist', 'harlot'), replicating the patterns seen above and confirming the

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

TABLE 2. *Sample words from the extremes of each quadrant of funniness and imputed iconicity ratings (total n = 3577)*

	HIGH IMPUTED ICONICITY	LOW IMPUTED ICONICITY
HIGH FUNNINESS	swish, chug, bop, gobble, smack, blip, whack, oomph, poke, wallop	heifer, dinghy, cuckold, nudist, sheepdog, oddball, spam, harlot, getup, rickshaw
LOW FUNNINESS	shudder, scrape, taps, fright, heartbeat, puncture, choke, tremor, biceps, glimpse	subject, ransom, libel, bible, siege, hospice, conduct, arsenic, clothing, negro

generalizability of our imputation method. However, as above, about 10% of the top 200 nouns with high imputed iconicity are compound nouns with transparent but non-iconic structure (e.g., 'heartbeat', 'mouthful', 'handshake', 'bellboy', 'comeback', 'catchphrase'), suggesting the imputation method is sensitive to the presence of such words in the training set.

Although not our focus here, in the online materials we report a further quality check of the imputation method on the inverse set of data (testing how human iconicity ratings covary with imputed funniness for 1526 words), which is consistent with our results.

3.3. IMPUTED ICONICITY AND IMPUTED FUNNINESS (UNKNOWN UNKNOWNS)

With the imputation technique validated against human funniness ratings, we can move on to the next step: the relation between imputed funniness and imputed iconicity in the set of 63680 words for which we have no human ratings (Figure 3, panel C). We formulate a linear model with imputed funniness as the dependent variable. Model comparison shows that a model including imputed iconicity as a predictor provides a significantly better fit ($F = 4536.3$, $p < .0001$) and explains a much larger portion of the variance (adjusted $R^2 = 0.237$ versus 0.057) than a model with just log frequency and lexical decision time. In the fuller model, imputed iconicity rises to be the strongest predictor ($F = 4552.9$, $p < .0001$, partial $\eta^2 = 19.1\%$), followed by frequency ($F = 1241.8$, $p < .0001$, partial $\eta^2 = 6.1\%$) and lexical decision time ($F = 182.4$, $p < .0001$, partial $\eta^2 = 0.9\%$). A partial correlations analysis shows that imputed iconicity values show 43% covariance with imputed funniness ratings, controlling for word frequency ($r = 0.43$, $p < .0001$).

As above, many of the words identified as high in iconicity by our imputation method are clearly imitative in origin: 'whoosh', 'whirr', 'chomp', etc. (Table 3). Words low in imputed iconicity and high in funniness include animals ('pigs', 'monkeys', 'penguins') but also words from other languages ('herr', 'beau', 'raja'), consistent with co-occurrence relations in the discursive context of jokes.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

TABLE 3. *Sample words from the extremes of each quadrant of imputed funniness and imputed iconicity ratings (n = 63680)*

	HIGH IMPUTED ICONICITY	LOW IMPUTED ICONICITY
HIGH IMPUTED FUNNINESS	whoosh, whirr, whooshing, brr, argh, chomp, whir, swoosh, brrr, zaps	pigs, monkeys, herr, raja, franz, lulu, von, beau, caviar, penguins
LOW IMPUTED FUNNINESS	slashes, gunshots, footstep, cries, fade, froze, swelter, crushing, piercing	apr, dei, covenants, palestinians, covenant, clothier, variant, mitochondria, israelis

For high imputed iconicity and low imputed funniness we find negatively valenced words like 'slashes', 'gunshots', 'swelter', and 'cries', though the iconic quality of some of these words is less clear, a sign of limitations of the semantically based imputation method. About 15% of a random sample of 200 out of the top 3560 nouns with high imputed iconicity (a sample size chosen to be proportionate to the other datasets) are analysable compounds like 'fireworm', 'uppercut', 'woodwork', 'biotech', suggesting that the imputation method may be amplifying the bias toward non-iconic analysable compounds introduced in the training set. The extreme of the opposite quadrant of low imputed iconicity and low imputed funniness seems to pick up mostly rare words.

3.4. STRUCTURAL PROPERTIES

With the relation between funniness and iconicity established in human as well as imputed ratings, we turn to the structural properties of words rated high in funniness and iconicity. The prediction is that they should show signs of structural markedness. Our analyses in this section are part confirmatory, part exploratory. The confirmatory part investigates the role of phonological improbability as a proxy for structural markedness, in line with our hypothesis that markedness, as a form of foregrounding, makes it more likely for words to be seen as playful and iconic. The exploratory part examines the set of words rated highest for iconicity and funniness to inductively characterize cues of structural markedness in these words, and then traces these cues across other segments of the dataset to examine the generalisability of the findings.

3.4.1. *Log letter frequency*

Prior work has shown that phonemic and orthographic improbability may help to explain funniness ratings; in particular, log letter frequency (a measure of orthographic unexpectedness) emerges as a strong correlate of perceived word

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

funniness (Westbury & Hollis, 2019). We reproduce this result in the human-rated subset of words, finding that a model including log letter frequency provides a significantly better fit ($F = 93.899$, $p < .0001$) and explains a larger portion of the variance (adjusted $R^2 = 0.208$ vs. 0.188) than the second model in §3.1 above with just word frequency, iconicity, and lexical decision time as predictors.

Our theory of structural foregrounding predicts that log letter frequency (insofar as it is a proxy of markedness) will show a relation to both funniness and iconicity ratings. Partial correlations indeed show that funniness rating and log letter frequency have a covariance of –15.7% controlling for iconicity, and that iconicity and log letter frequency have a covariance of –16.3% controlling for funniness ratings (all $p < .0001$ correcting for multiple comparisons). In other words, log letter frequency relates as strongly to iconicity as to funniness.

We construct a linear model predicting the combined funniness and iconicity ranking of words (standardized to z-scores and summed). Model comparison shows that a model including log letter frequency provides a significantly better fit ($F = 96.41$, $p < .0001$) and explains a larger portion of the variance (adjusted $R^2 = 0.18$ vs. 0.13) than a model with just word frequency and lexical decision time as predictors. In this model, word frequency is the most important predictor ($F = 219.96$, $p < .0001$, partial $\eta^2 = 13.5\%$), followed by log letter frequency ($F = 96.41$, $p < .0001$, partial $\eta^2 = 6.4\%$), while the influence of lexical decision time is dwarfed ($F = 2.89$, $p = .09$, partial $\eta^2 = 0.2\%$), perhaps because words with lower log letter frequency have higher lexical decision times in general.

Somewhat to our surprise, the relatively coarse measure of log letter frequency is more informative than more subtle phonological and phonotactic measures from the Irvine Phonotactic Online Dictionary (Vaden et al., 2009). For the current dataset, the measures of phonological density, biphone probability, and triphone probability do not seem to offer additional explanatory power beyond log letter frequency, as reported in the online materials. Perhaps this reflects the written origin of the iconicity and funniness ratings.

3.4.2. *Structural analysis*

To better understand the structural properties of words rated high in iconicity and funniness, we carried out a linguistic analysis of the combined upper ten percentiles of iconicity and funniness ratings, representing 80 words. We catalogued the phonotactic complexity of these words and found three recurring cues of structural markedness. Of these words, 38% had complex onsets, as in 'flap', 'sniff', 'drizzle'; 20% had complex codas, as in 'oink', 'whirl', 'clunk'; and 11% had the expressive verbal diminutive suffix '-le' as in 'tingle', 'wobble', 'wiggle' (Table 4). These cues do not exhaust the structurally

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

TABLE 4. *Cues of structural markedness identified in the highest-rated words; their relative prevalence in the top 80 versus the remaining 1339 words of set A*

CUE TYPE	ATTESTED FORMS	EXAMPLES	IN TOP 80	IN REST
onsets	*bl, cl, cr, dr, fl, sc, sl, sn, sp, spl, sw, tr, pr, sq*	bleep, crunch, flap, flick, prick, sniff, slick, slime, splash, squeeze	38%	15%
codas	*nch, mp, nk, rt, rl, rr, sh, wk*	dump, splash, limp, crunch, clamp, mush, snort, whirl, swirl, squawk	20%	5%
verbal suffix	*-le*	tickle, wiggle, giggle, babble, wobble, jiggle, ramble, scuttle, waddle	11%	0.6%

marked properties of the individual words, but they are the most readily recognisable.

Each of these inductively identified cues turns out to be connected to playfulness and sound symbolism. The complex onsets and codas are examples of phonaesthemes: submorphemic elements often showing non-arbitrary form–meaning associations (Kwon & Round, 2014). The verbal suffix '-le' is connected to iterative and diminutive meanings that often have a ludic or non-serious character (Dressler & Merlini Barbaresi, 1994; Audring, Booij, & Jackendoff, 2017); in many of the higher-rated words it is connected to a sense of movement and plurality. These same cues of structural markedness are much rarer in the remaining 1339 words in set A: complex onsets occur in 15%, complex codas in only 5%, and the verbal suffix '-le' in only 0.6% of words (Figure 4, panels A–C).

As the cues can co-occur in words, we sum them to form a cumulative measure of structural markedness (so 'cat' and 'ape' score 0, 'flap' and 'dump' score 1 for their complex onset or coda, and 'clunk' and 'drizzle' both score 2 for their combinations of onset, coda, and/or verbal diminutive suffix). Operationalised in this way, the average cumulative structural markedness of the set of 80 high-iconicity high-funniness words is much higher than expected if they resembled a randomly drawn sample from the larger dataset (0.69 versus 0.20, $t(82.7) = 6.23$, $p < .0001$, Cohen's $d = 0.9$). Revisiting the linear model predicting the combined funniness and iconicity ranking of words, model comparison shows that a model including this new measure of cumulative markedness as predictor provides a significantly better fit ($F = 52.78$, $p < .0001$) and explains a larger portion of the variance (adjusted $R^2 = 0.21$ vs. 0.18) than a model with word frequency, lexical decision time, and log letter frequency. Figure 4 shows the patterning of cumulative structural markedness along with the individual cues for funniness rating percentiles, iconicity rating percentiles, and combined percentiles.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

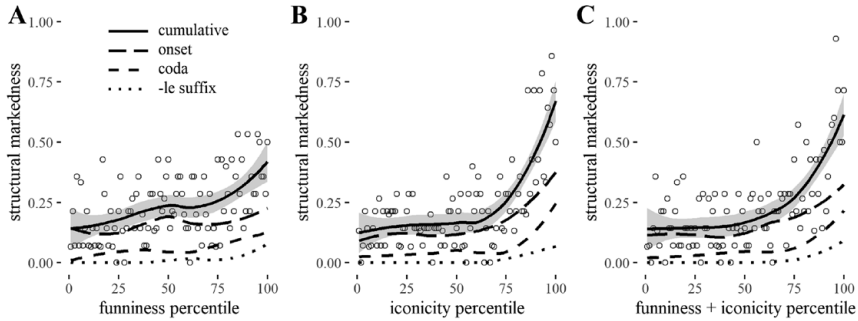

Fig. 4. The relation between structural markedness and **A** funniness ratings, **B** iconicity ratings, and **C** funniness and iconicity together, all in set A (1419 human-rated words). Ratings are rescaled to 0–100 percentiles for comparability. Each dot represents 14 or 15 words. Solid lines and shading represent a loess function of cumulative markedness with 95% confidence intervals. Other lines show relative prevalence of complex onsets, codas, and verbal diminutives.

As a final test of the utility of our imputation method we trace the inductively identified structural properties of high-iconicity high-funniness words in the subset of data for which we have only imputed ratings. We find a similarly skewed distribution of structural markedness: in the upper ten percent of imputed iconicity ratings, 23% of 6368 words contain one or more cues of structural markedness (examples are 'swoosh', 'squish', 'crush', 'dribble', 'crackles', 'flickered'), whereas this level is only 9% in the remaining 57312 words (examples are 'snowman', 'drank', 'spaceport', 'trench', 'swedish', 'schubert'). Comparison of models with combined imputed funniness and iconicity as a dependent variable shows that a linear model including cumulative markedness as predictor provides a significantly better fit ($F = 337.3$, $p < .0001$) and explains a little bit more of the variance (adjusted $R^2 = 0.124$ vs. 0.109) than a model with just word frequency, lexical decision time, and log letter frequency (see figures in the online materials). In other words, the inductively identified structural correlates of human iconicity and funniness ratings also show up in words for which we have only imputed ratings.

4. Discussion

We have found that human ratings for funniness and iconicity show a tendency to converge, especially at the higher end: words like 'zigzag', 'squeak', and 'waddle' are rated as highly iconic and highly funny. This underlines the special relation between playfulness and performativity and makes it relevant to examine underlying factors. We found that a measure of phonological unexpectedness, previously shown to correlate with funniness ratings (Westbury & Hollis, 2019), correlates at least as strongly with iconicity

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

ratings. While prior work has ascribed the phonological unexpectedness of funny words to a theory of humour based on incongruity (Westbury et al., 2016), the finding that it applies just as strongly to iconic words strengthens the case for the more general theoretical account we propose here, according to which structural markedness unites playful and iconic words. A linguistic analysis of high-iconicity high-funniness words helped identify three reliable cues of structural markedness in English: complex onsets, complex codas, and the verbal suffix '-le'. These structural properties, we propose, exemplify the metacommunicative cues that help foreground words and invite us to experience them as playful, poetic, and performative. The strongly skewed distribution of these cues across the vocabulary provides further supporting evidence for this role.

Our theoretical account does not lead us to expect that iconicity and funniness ratings are uniformly consonant across the board, and indeed discrepancies bring to light other contributing factors. Words rated high in iconicity but low in funniness tend to present vivid depictions of negatively valenced events like 'crash' or 'roar', reproducing a familiar relation between word funniness and valence that is independent of iconicity (Westbury & Hollis, 2019). Words rated high in funniness but low in iconicity like 'buttocks' or 'blonde' tend to be associated with taboos and socio-semantic categories that figure in some genres of Anglo jokes. This is a contributor to ratings that is more likely to be culturally variable than structural markedness cues (Low, 2011), which has implications for the cross-linguistic generalisability of funniness ratings.

Imputed iconicity ratings correlate well with human funniness ratings and show the same general patterns we find in the training datasets. Remarkably, the correlation is amplified in successively larger datasets: it is 20.6% in the core set of human ratings, goes up to 32.3% when comparing imputed iconicity ratings to human funniness ratings (n = 3577), and up again to 42.8% in the two sets of imputed ratings (n = 63680). That at least some of the same broad patterns show up in a dataset at least twenty times as large as the training set suggests that imputation can be a useful pursuit.

The structural markedness cues inductively discovered in the training set – complex onsets, codas, and evaluative morphology – also show up in words for which no human ratings are available. This is notable because the vector-based imputation method is primarily based on distributional semantics and not on explicit word-level form–meaning associations. It means that the imputation method is relatively reliable and can be used to increase the coverage of lexical ratings beyond small sets of seed words, generating new data for follow-up research. For instance, high imputed iconicity words can be put to the test in experimental or corpus-based investigations of iconicity, and words with high imputed funniness can be used in research on verbal humour, substantially extending the existing funniness ratings.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

4.1. GENERALISATIONS AND PREDICTIONS

We have found that words perceived as highly funny and highly iconic are united in showing signs of structural markedness, consistent with the theory that structural markedness can function as a metacommunicative cue inviting playful and performative interpretations (Bateson, 1955). Our account generates predictions in the areas of comparative linguistics, cultural evolutionary modelling, and corpus studies of multimodal language use.

In the domain of comparative linguistics, our account provides an explanatory framework for qualitative observations reported for languages around the world, from the playful connotations of *ts*-initial words in Greek (Joseph, 1994) and the "attitude of playfulness" detected in imitative words in Spanish and Basque (Pharies, 1990, p. 107) to the mirth associated with ideophones in Alto-Perene (Arawak, Peru; Mihas, 2012), Hamar (Omotic, Ethiopia; Lydall, 2000), Kalam (Trans New Guinea, Papua New Guinea; Pawley, 2010), and Shona (Bantu, Zimbabwe; Fortune, 1962). Such observations, along with the quantitative evidence from English presented here, make us confident that the predictions of our account – that high iconicity and high funniness go together, and that they are underpinned by signs of structural markedness – should hold across a wide variety of languages.

To the extent that structural markedness serves as a metacommunicative signal of playfulness and performativity, it also has consequences for the cultural evolution of lexical structure. Our prediction is that structural markedness confers a selective advantage on words intended to be iconic and/or funny, as their recognisability would make them more fit to survive processes of cultural transmission in which the recognition of such intentions is functionally important. This prediction is ripe for testing in laboratory experiments or computational models of cultural evolution.

Metacommunicative cues that say "this is play" are of course also found beyond the phonotactic structures studied here in written words. As the Prague school linguist Havránek wrote, "conventional conversational devices are automatized, but to liven up the conversation and to achieve surprise (wonderment) foregrounded units are used" (Havránek, 1964, p. 10). Our account predicts that, in everyday language use, words framed as special by means of performative foregrounding – from expressive prosody to playful morphology – are more likely to be perceived as both playful and iconic. Again, qualitative observations from across languages support this view, for instance in the form of work on ideophones as playful multimodal depictions (Dingemanse, 2011; Ibarretxe-Antuñano, 2017) and on reduplication as a sign of playfulness (Rastall, 2004; Haiman, 2014). Here as elsewhere our predictions are not deterministic but probabilistic: not all reduplicated words are funny or iconic, but given the possible role of reduplication as a

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

metacommunicative sign of play, it is more likely for such words to be used and perceived that way.

Most generally, the kind of metacommunicative framing studied here in lexical items is associated with depiction as a mode of communication (Clark, 2016). Depiction often lends itself to playful connotations, for at least two reasons: (i) the sensory imagery offered by depictions give us a palpable sense of presence by enabling us to experience what it is like to perceive the scene depicted (Lydall, 2000); and (ii) the selectivity of depictions foregrounds salient sensory features and backgrounds others much like cartoons or caricatures can do, and to similar playful effect (Samarin, 1969). Indeed, both vivid sensory imagery (Graesser, Long, & Mio, 1989) and selectivity and exaggeration (Kris & Gombrich, 1938) are connected to humour and playfulness. So 'whiff', 'waddle', and 'zigzag' may be perceived as funny not just because of their marked phonology, but also because of their depictive semiotics. To the extent that words prone to be used depictively occur in similar distributional contexts (from vivid stories to entertaining dialogues), this may also help to explain the performance of our imputation method, which relies primarily on distributional semantics.

We arrive, therefore, at a more precise characterisation of the link between playfulness and iconicity. Summing up the lessons learned:

I. While not all funny words are iconic, and not all iconic words are funny, many highly iconic words are perceived as funny.
II. Words perceived as iconic and funny feature cues of structural markedness that serve to foreground them and invite perceptions of playfulness and performativity.
III. The link between playfulness and iconicity is further reinforced by the depictive semiotics of iconic words, in particular their vivid sensory imagery and selective depictive properties.

To the best of our knowledge, our study is the first large-scale investigation of English vocabulary (and perhaps of vocabulary in any language) to firmly establish points I–II both in human-rated words and in a much larger set of words with imputed ratings. Point III has not been the main target of this study and represents an important area for future research.

4.2. LIMITATIONS AND RECOMMENDATIONS

Norm imputation can distort rating scales and can amplify rating artefacts (Mandera et al., 2015), as we saw for analysable compound nouns like 'footstep', 'catchphrase', and 'biotech', which received high imputed iconicity ratings probably because of artefacts introduced in the original ratings data.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

The relative proportion of such words went up from 10% in the training set to an estimated 15% in the larger set of imputed ratings. So, while a large majority of words in the higher end of the combined imputed ratings are clear and uncontroversial examples of funny and iconic words, there is some reason to be cautious. One way to mitigate the consequences of the bias introduced by analysable compounds is to focus on monomorphemic words, which do not allow the conflation of iconicity with analysability. The online materials show that the patterns reported above emerge even more clearly in monomorphemic words, and all quantitative findings are at least as strong in the 8642 monomorphemic words for which we have human and imputed ratings.

We inductively identified three simple structural cues of markedness that occurred in up to 38% of the highest-rated words and that helped explain the relation between iconicity and funniness over and above other known factors. No doubt there are many more contributors to perceived funniness and iconicity, ranging from phonetic features to distributional and semantic properties (Westbury, Hollis, Sidhu, & Pexman, 2017). For instance, German words with voiceless consonants tend to be perceived as more arousing and negative (Aryani et al., 2018), and English auditory and tactile words tend to be more iconic (Winter et al., 2017). It is also likely that bottom-up data-driven approaches could identify more cues correlated with non-arbitrary structure in the lexicon (Nuckolls et al., 2016; Pimentel, McCarthy, Blasi, Roark, & Cotterell, 2019), and could be used to further boost the performance of methods for imputing lexical ratings.

The combination of quantitative and qualitative analysis employed here brings out some of the strengths and weaknesses of lexical ratings, both collected and imputed. Ratings can reveal robust correlations which can be made sense of using linguistic analysis. However, potential ambiguities in instructions can introduce artefacts and imputation methods can amplify them. Our recommendation is to never take ratings at face value and to always triangulate robustness and validity using other methods or data. With that caveat in mind, however, imputed ratings can serve to increase data coverage and allow confirmatory and exploratory analyses in large-scale datasets that will remain, for some time at least, out of reach of human-collected ratings.

5. Conclusions

The use of structurally marked words to evoke the playful and poetic is probably as old as the use of language itself. Here we have examined the theory that the structural markedness of words can serve as a metacommunicative signal (Bateson, 1955), allowing words to break frame and attract attention to themselves as playful and performative. Our

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

investigation has put the playfulness of iconic words on a firm empirical footing. We have found formal cues of structural markedness whose distribution strongly correlates with people's perceptions of words as funny and iconic. We have introduced and benchmarked a method for imputing lexical ratings of funniness and iconicity, with reason for cautious optimism about the generalisability of the results. And we examined some of the strengths and limitations of lexical ratings by combining qualitative and quantitative analysis.

Approaching iconicity using quantitative methods may seem to take away the magic of make-believe these words thrive on (Dingemanse, 2014). Likewise, explaining humour has been compared to dissecting an animal: you understand it better, but it dies in the process (White, 1941). If, as our study suggests, structural markedness helps to explain the relation between funniness and iconicity, at least we have killed two birds with one stone ¯_(ツ)_/¯.

Linguistics has long focused on the referential function of language to the exclusion of its expressive and poetic potential (Crystal, 1996; Jakobson, 1960). Studying ludic aspects of the lexicon is valuable if linguistics is to be a truly comprehensive science of language. But there is more to it than that. As Bateson (1955) noted, the metacommunicative abstraction involved in the ability to distinguish "play" from "not play" may well hold one of the keys to the origins of communication and therefore the evolution of language. Here we have seen that some of the metacommunicative signals to tell the playful from the prosaic may well be built into the very texture of the lexicon.

Online materials

An Rmarkdown code notebook of all analyses in this paper, along with Python code for the rating imputation method, all data files, and a set of supplementary analyses can be found in the OSF repository at <https://osf.io/7s6xc/>.

Acknowledgements

This work has benefited from audience feedback at the 12th Iconicity in Language & Literature Symposium in Lund, May 2019. For helpful and incisive comments, we are grateful to Marieke Woensdregt (to whom we owe the point about the cultural evolutionary import of structural markedness) and to two anonymous reviewers. Thanks also to Luca Bischetti for feedback on the preprint and to Bodo Winter for once tweeting that "iconicity is just plain fun". MD is funded by the Dutch Research Council (grant 016.vidi.185.205). BT acknowledges generous support from a Levinson scholarship through the Language and Cognition Department at the MPI for Psycholinguistics.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

Contributions: MD designed the research, conducted the quantitative and linguistic analyses, and wrote the first draft. BT designed and described the imputation method and contributed imputed ratings and letter frequency measures. Both authors contributed to revisions of the paper and approve of the final version.

REFERENCES

Aryani, A., Conrad, M., Schmidtke, D. & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PLOS ONE* **13**(6), e0198430.

Attardo, S. (2018). Universals in puns and humorous wordplay. In E. Winter-Froemel & V. Thaler (eds.), *Cultures and traditions of wordplay and wordplay research* (pp. 89–110). Berlin/Boston: Walter de Gruyter.

Audring, J., Booij, G. & Jackendoff, R. (2017). Menscheln, kibbelen, sparkle. *Linguistics in the Netherlands* **34**(1), 1–15.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B. … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods* **39**(3), 445–459.

Bateson, G. (1955). A theory of play and fantasy. *Psychiatric Research Reports* **2**(39), 39–51.

Bojanowski, P., Grave, E., Joulin, A. & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* **5**, 135–146.

Clark, H. H. (2016). Depicting as a method of communication. *Psychological Review* **123**(3), 324–347.

Crystal, D. (1996). Playing with linguistic problems: from Orwell to Plato and back again. In *Georgetown University Round Table on Languages and Linguistics (GURT) 1996: Linguistics, Language Acquisition, and Language Variation: Current Trends and Future Prospects* (pp. 5–29).

Dingemanse, M. (2011). Ideophones and the aesthetics of everyday language in a West-African society. *The Senses and Society* **6**(1), 77–85.

Dingemanse, M. (2014). Making new ideophones in Siwu: creative depiction in conversation. *Pragmatics and Society* **5**(3), 384–405.

Dingemanse, M. (2019). 'Ideophone' as a comparative concept. In K. Akita & P. Pardeshi (eds), *Ideophones, mimetics, expressives* (pp. 13–33). Amsterdam/Philadelphia: John Benjamins.

Dressler, W. U. & Merlini Barbaresi, L. (1994). *Morphopragmatics: diminutives and intensifiers in Italian, German, and other languages*. Berlin/New York: M. de Gruyter.

Dynel, M. (2009). Beyond a joke: types of conversational humour. *Language and Linguistics Compass* **3**(5), 1284–1299.

Engelthaler, T. & Hills, T. T. (2018). Humor norms for 4,997 English words. *Behavior Research Methods* **50**(3), 1116–1124.

Fortune, G. (1962). *Ideophones in Shona: an inaugural lecture given in the University College of Rhodesia and Nyasaland on 28 April 1961*. London/New York: Oxford University Press.

Fox, J. & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Retrieved from <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.

Glenn, P. J. (2003). *Laughter in interaction*. New York: Cambridge University Press.

Goffman, E. (1974). *Frame analysis: an essay on the organization of experience*. Cambridge, MA: Harvard University Press.

Graesser, A. C., Long, D. L. & Mio, J. S. (1989). What are the cognitive and conceptual components of humorous text? *Poetics* **18**(1), 143–163.

Haiman, J. (2014). Six competing motives for repetition. In B. MacWhinney, A. Malchukov & E. Moravcsik (eds), *Competing motivations in grammar and usage* (pp. 246–260). New York: Oxford University Press.

Havránek, B. (1964). The functional differentiation of the standard language. In P. L. Garvin (ed.), *A Prague School reader on esthetics, literary structure, and style* (pp. 3–16). Washington, DC: Georgetown University Press.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

Hollis, G., Westbury, C. & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning. *Quarterly Journal of Experimental Psychology* **70**(8), 1603–1619.
Ibarretxe-Antuñano, I. (2017). Basque ideophones from a typological perspective. *Canadian Journal of Linguistics / La Revue Canadienne de Linguistique* **62**(2), 196–220.
Jakobson, R. (1960). Linguistics and poetics. In T. A. Sebeok (ed.), *Style in language* (pp. 350–377). Cambridge, MA: MIT Press.
Jakobson, R. & Waugh, L. R. (1979). *The sound shape of language*. Bloomington, IN: Indiana University Press.
Joseph, B. D. (1994). Modern Greek ts: beyond sound symbolism. In L. Hinton, J. Nichols & J. J. Ohala (eds), *Sound symbolism* (pp. 222–236). Cambridge: Cambridge University Press.
Keuleers, E., Lacey, P., Rastle, K. & Brysbaert, M. (2012). The British Lexicon Project: lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods* **44**(1), 287–304.
Kim, S. (2015). Ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Communications for Statistical Applications and Methods* **22**(6), 665–674.
Klamer, M. (2002). Semantically motivated lexical patterns: a study of Dutch and Kambera expressives. *Language* **78**(2), 258–286.
Kris, E. & Gombrich, E. (1938). The principles of caricature. *British Journal of Medical Psychology* **17**(3/4), 319–342.
Kunene, D. P. (2001). Speaking the act: the ideophone as a linguistic rebel. In F. K. E. Voeltz & C. Kilian-Hatz (eds), *Ideophones* (pp. 183–191). Amsterdam: John Benjamins.
Kwon, N. & Round, E. R. (2014). Phonaesthemes in morphological theory. *Morphology* **25**(1), 1–27.
Levisen, C. (2018). Dark, but Danish: ethnopragmatic perspectives on black humor. *Intercultural Pragmatics* **15**(4), 515–531.
Low, P. A. (2011). Translating jokes and puns. *Perspectives* **19**(1), 59–70.
Lydall, J. (2000). Having fun with ideophones: a socio-linguistic look at ideophones in Hamar, Southern Ethiopia. In B. Yimam, R. Pankhurst, D. Chapple, Y. Admassu, A. Pankhurst & B. Teferra (eds), *Proceedings of the XIV International Conference of Ethiopian Studies* (pp. 886–891). Addis Ababa: Addis Ababa University.
Mandera, P., Keuleers, E. & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *Quarterly Journal of Experimental Psychology* **68**(8), 1623–1642.
Menninghaus, W., Bohrn, I. C., Altmann, U., Lubrich, O. & Jacobs, A. M. (2014). Sounds funny? Humor effects of phonological and prosodic figures of speech. *Psychology of Aesthetics, Creativity, and the Arts* **8**(1), 71–76.
Mihas, E. (2012). Ideophones in Alto Perene (Arawak) from Eastern Peru. *Studies in Language* **36**(2), 300–344.
Motamedi, Y., Little, H., Nielsen, A. & Sulik, J. (2019). The iconicity toolbox: empirical approaches to measuring iconicity. *Language and Cognition* **11**(2), 188–207.
Nuckolls, J. B. (1999). The case for sound symbolism. *Annual Review of Anthropology* **28**, 225–252.
Nuckolls, J. B., Nielsen, E., Stanley, J. A. & Hopper, R. (2016). The systematic stretching and contracting of ideophonic phonology in Pastaza Quichua. *International Journal of American Linguistics* **82**(1), 95–116.
Pawley, A. (2010). Helter skelter and ñugl ñagl: English and Kalam rhyming jingles and the psychic unity of mankind. In K. A. McElhanon & G. Reesink (eds), *A mosaic of languages and cultures: studies celebrating the career of Karl J. Franklin* (pp. 273–293). Dallas, TX: SIL e-Books.
Perlman, M., Little, H., Thompson, B. & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: a comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology* **9**, e01433.
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**, e00227.
Perry, L. K., Perlman, M. & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE* **10**(9), e0137147.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W. & Lupyan, G. (2017). Iconicity in the speech of children and adults. *Developmental Science*, **21**(3), e12572.
Pharies, D. A. (1990). A Structural correspondence in the lexicons of Basque and Spanish. *Neuphilologische Mitteilungen* **91**(1), 107–121.
Pimentel, T., McCarthy, A. D., Blasi, D. E., Roark, B. & Cotterell, R. (2019). Meaning to form: measuring systematicity as information. *ArXiv:1906.05906 [Cs]*. Retrieved from <http://arxiv.org/abs/1906.05906>.
R Core Team. (2019). *R: a language and environment for statistical computing*. Retrieved from <https://www.R-project.org/>.
Rastall, P. (2004). Playful English: kinds of reduplication. *English Today* **20**(4), 38–41.
Samarin, W. J. (1969). The art of Gbeya insults. *International Journal of American Linguistics* **35**(4), 323–329.
Samarin, W. J. (1970). Inventory and choice in expressive language. *Word* **26**, 153–169.
Svantesson, J.-O. (2017). Sound symbolism: the role of word sound in meaning. *Wiley Interdisciplinary Reviews: Cognitive Science* **8**(5), e1441.
Thompson, B. & Lupyan, G. (2018). Automatic estimation of lexical concreteness in 77 languages. In C. Kalish, M. Rau, J. Zhu & T. T. Rogers (eds), *Proceedings of the 40th Annual Conference of the Cognitive Science Society (CogSci 2018)* (pp. 1122–1127). Retrieved from <http://mindmodeling.org/cogsci2018/papers/0222/0222.pdf>.
Vaden, K. L., Halpin, H. R. & Hickok, G. S. (2009). *IphOD: Irvine phonotactic online dictionary, version 2.0*. Retrieved from <http://www.iphod.com>.
Van Heuven, W. J. B., Mandera, P., Keuleers, E. & Brysbaert, M. (2014). SUBTLEX-UK: a new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology* **67**(6), 1176–1190.
Welmers, W. E. (1973). *African language structures*. Berkeley, CA: University of California Press.
Westbury, C. & Hollis, G. (2019). Wriggly, squiffy, lummox, and boobs: What makes some words funny? *Journal of Experimental Psychology: General* **148**(1), 97–123.
Westbury, C., Hollis, G., Sidhu, D. M. & Pexman, P. M. (2017). Weighing up the evidence for sound symbolism: distributional properties predict cue strength. *Journal of Memory and Language* **99**, 122–150.
Westbury, C., Shaoul, C., Moroschan, G. & Ramscar, M. (2016). Telling the world's least funny jokes: on the quantification of humor as entropy. *Journal of Memory and Language* **86**, 141–156.
White, E. B. (1941). Preface. In E. B. White & K. S. White (eds), *A subtreasury of American humor* (pp. xi–xxii). New York: Coward-McCann.
Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Retrieved from <https://ggplot2.tidyverse.org/>.
Wickham, H. (2017). *tidyverse: easily install and load the 'Tidyverse'*. Retrieved from <https://CRAN.R-project.org/package=tidyverse>.
Winter, B. (2019). *Sensory linguistics*. Amsterdam: John Benjamins.
Winter, B., Perlman, M., Perry, L. & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies* **18**(3), 432–453.
Zwicky, A. M. & Pullum, G. K. (1987). Plain morphology and expressive morphology. In J. Aske, N. Beery, L. Michaelis & H. Filip (eds), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society: Vol. VII* (pp. 330–340). Berkeley, CA: Berkeley Linguistics Society.

Downloaded from https://www.cambridge.org/core. IP address: 71.59.67.45, on 28 Feb 2020 at 12:39:31, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/langcog.2019.49