nature human behaviour

Article

https://doi.org/10.1038/s41562-022-01483-8

Valence-dependent mutation in lexical evolution

Received: 10 May 2022

Accepted: 14 October 2022

Published online: 28 November 2022

Check for updates

Joshua Conrad Jackson[1], **Kristen Lindquist**[2], **Ryan Drabble**[2], **Quentin Atkinson**[3,4] & **Joseph Watts**[4,5]

A central goal of linguistics is to understand how words evolve. Past research has found that macro-level factors such as frequency of word usage and population size explain the pace of lexical evolution. Here we focus on cognitive and affective factors, testing whether valence (positivity–negativity) explains lexical evolution rates. Using estimates of cognate replacement rates for 200 concepts on an Indo-European language tree spanning six to ten millennia, we find that negative valence correlates with faster cognate replacement. This association holds when controlling for frequency of use, and follow-up analyses show that it is most robust for adjectives ('dirty' versus 'clean'; 'bad' versus 'good'); it does not consistently reach statistical significance for verbs, and never reaches significance for nouns. We also present experiments showing that individuals are more likely to replace words for negative versus positive concepts. Our findings suggest that emotional valence affects micro-level guided variation, which drives macro-level valence-dependent mutation in adjectives.

There are more than 7,000 living languages across the globe today[1]. This linguistic diversity has emerged through a process that resembles Darwinian evolution[2,3]. Much like how humans and chimpanzees share a common ancestor from millions of years ago, languages such as Hindi and German shared a common Indo-European ancestor several thousand years ago[4,5]. And much like species share traits such as hooves and lactation due to common ancestry, many languages share cognate words that show systematic sound correspondences indicating that they diverged from a common origin. For example, the English word 'good' is translated as 'gut' in German, and 'góður' in Faroese, suggesting a common origin. Despite the commonalities between lexical and genetic evolution, there are also key differences between these processes. One important difference is that cultural and linguistic mutations are significantly less random than genetic mutations[6,7]. Models of guided variation[8] and semantically biased transmission[9] claim that humans often non-randomly transform stories, beliefs and practices based on psychological biases and preferences[10–13]. In this Article, we explore how these micro-scale processes of non-random transformation may shape the macro-scale evolution of languages over millennia.

Cognition and emotion clearly impact how humans use language[14], yet few studies in linguistics have tested whether people's cognitive and affective appraisals of word meanings explain long-term trends in the evolution of language. Past studies of lexical evolution have largely focused on how demographic (for example, population size[15] and inter-group contact[16]) and linguistic (for example, frequency of word use[17], length of word[18], average age of acquisition[19] and number of synonyms[20]) properties influence lexical evolution. On the other hand, many psychological studies have analysed how semantic factors such as valence (positivity versus negativity), dominance and arousal shape how humans process language[10,21–23], but do not consider the implications of these factors for long-term language change. Only rarely have studies considered how psychological properties scale up to produce population-level changes in lexical evolution[20]. Studying the psychological underpinnings of long-term lexical evolution therefore represents a significant step forward for both linguistics and psychology.

[1]Kellogg School of Management, Northwestern University, Evanston, IL, USA. [2]Department of Psychology and Neuroscience, University of North Carolina, Chapel Hill, NC, USA. [3]School of Psychology, University of Auckland, Auckland, New Zealand. [4]Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [5]Religion Programme, University of Otago, Dunedin, New Zealand. e-mail: joshua.jackson@kellogg.northwestern.edu

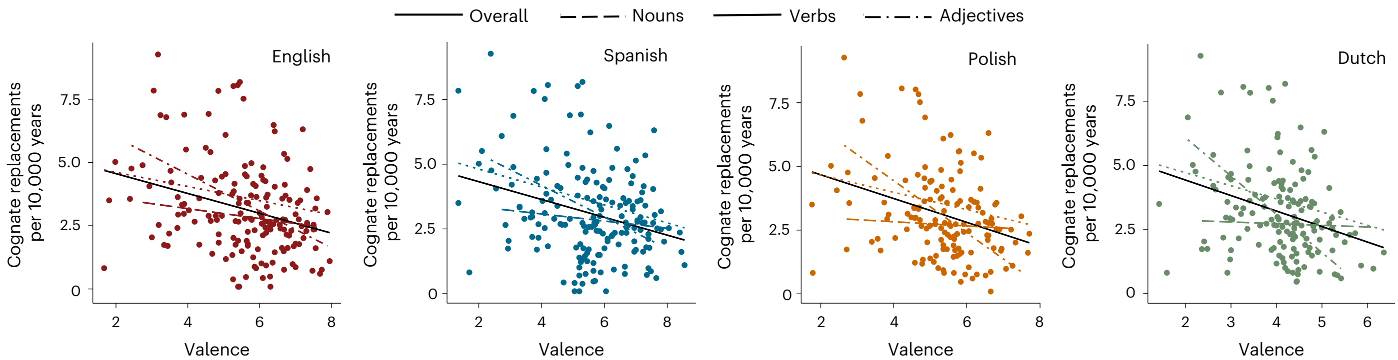

Fig. 1 | The relationship between semantic valence and cognate replacement rate in four Indo-European languages. Higher values of valence represent more positive words and negative values represent more negative words. Negative valence is correlated with faster rates of cognate replacement for all four languages. The overall effect is displayed in black, and effects by verbs, adjectives and nouns are displayed in dashed coloured lines. The English-language panel contains 168 words; the Spanish-language panel contains 175 words; the Polish-language panel contains 144 words; the Dutch-language panel contains 151 words. Supplementary Fig. 1 reproduces this figure while adjusting for control variables.

Drawing from past research in cognitive and affective science, we examine whether valence explains variation in rates of lexical evolution. We focus on valence because it appears to be a culturally and developmentally universal semantic dimension. Valence can predict how languages around the world encode word meaning better than dominance, certainty and physiological arousal[24,25]. The human brain is highly attuned to valence, and people rapidly process valence when they perceive faces, objects and words[26–28]. Humans appear to gain tools for understanding and communicating valence early in life with minimal learning. 'Bad' and 'good' are among the first words learned by children[21,29,30]. Valence is also an important dimension in the study of cultural evolution, which has documented changes to the positivity of music[31], song lyrics[32] and literature[33] over time. If valence is a psychologically meaningful semantic dimension for all humans, it may influence which words mutate and which words remain stable at both the micro- and macro-levels of language evolution.

Past research raises competing hypotheses about how valence may affect lexical evolution. Some evidence suggests that negative words may evolve faster than positive words. Negative information draws more attention[34–36], and people often perceive negative information as more self-relevant than positive information[37], which leads them to evaluate negative information with more analytic and elaborative processing than positive information[38–40]. Research on the Pollyanna principle also finds that people have a tendency to remember and share positive information when they think about the past[41], leading people to change negative information at a greater rate than positive information. A preference for positive information and a tendency to scrutinize and change negative information may lead to negative words mutating more than positive words over time. In contrast, other studies suggest that cultural transmission biases could result in words for positive concepts mutating more rapidly than words for negative concepts. Humans learn negative concepts faster than positive concepts[42]. People are also more likely to accurately communicate negative concepts in transmission chain studies that assess how a message changes as people relay it to one another (like the classic game of 'Telephone')[12]. If people learn and communicate negative information with high fidelity, this could lead to negative words mutating less than positive words over time.

We use two approaches to test whether valence can explain patterns of lexical evolution. Our first series of analyses draws from large cross-cultural semantic norming studies that measured the semantic properties of thousands of concepts[43–46]. We combine these data with data on rates of cognate replacement—the rate at which languages replace words for concepts with a non-cognate form—for 200 basic concepts across 87 Indo-European languages over 6,000–10,000 years (ref. [17]). Indo-European is one of the largest and most diverse language families in the world, covering languages in South Asia, Central Asia and Europe. This language family is also well studied in linguistics research, making it possible to analyse data on word frequencies, borrowing, age of acquisition and other semantic properties (for example, physiological arousal and dominance) as control variables (for more information on all control variables in our analysis, see Methods). The 200 'Swadesh list' concepts that we use (Supplementary Table 1) are often used in comparative linguistics research because they have relatively comparable meanings across cultures and low rates of borrowing[47]. This macro-level analysis allows us to test whether valence can explain variation in cognate replacement, and to test whether these effects are independent of frequency of use, and other potentially relevant factors. We can also test whether a potential relationship between valence and cognate replacement varies across different parts of speech.

Our second series of analyses examines individual-level decision-making to test whether the semantic preferences of contemporary English speakers can account for an association between valence and lexical evolution. We develop a new paradigm in which we can estimate the rate at which participants select words for replacement when creating a new dialect of the English language. This paradigm models key features of lexical evolution, in so far as our participants represent 'interactors' and the words ('utterances') they use represent units of cultural information that are subject to selection[48]. While this paradigm does not model inter-group contact and social interaction, which play important roles in language change[48], it allows us to directly model the selection of positive versus negative words under highly controlled conditions. We compare whether words that are frequently replaced by English-speaking individuals also show high rates of population-level cognate replacement over millennia. In our Supplementary Methods and Results, we replicate the findings from this individual-level study with a slightly different procedure that serves as a robustness check.

Results

Valence and cognate replacement at the population level

Did positive and negative concepts evolve at different rates in our population-level data? Initial correlations revealed consistent evidence such that negative concepts evolved faster than positive concepts. We found a similar dynamic across English ($r(166) = -0.29$, $P < 0.001$, 95% confidence interval (CI) −0.43 to −0.15), Spanish ($r(173) = -0.29$, $P < 0.001$, 95% CI −0.42 to −0.15), Polish ($r(142) = -0.32$, $P < 0.001$, 95% CI −0.46 to −0.16) and Dutch ($r(149) = -0.30$, $P < 0.001$, 95% CI −0.44 to −0.15). These are not independent findings because cognate replacement rate is estimated across the whole Indo-European language phylogeny, meaning that each language in our analysis had the same cognate replacement data. Our language-specific analyses therefore

https://doi.org/10.1038/s41562-022-01483-8

Table 1 | Multilevel regression models predicting cognate replacement per 10,000 years

Predictor	*n*	*b* (standard error)	*β*	*t*	*P*	95% CI
English: model 1	168					
Valence		−0.38 (0.10)	−0.27	−3.80	<0.001	−0.57 to −0.18
Arousal		−0.19 (0.14)	−0.09	−1.31	0.19	−0.47 to 0.09
English: model 2	166					
Valence		−0.29 (0.10)	−0.21	−2.81	0.006	−0.49 to −0.09
Arousal		−0.17 (0.14)	−0.09	−1.22	0.23	−0.44 to 0.10
Frequency		−0.19 (0.10)	−0.14	−1.90	0.06	−0.37 to 0.006
Borrowing		−0.26 (0.09)	−0.20	−2.91	0.004	−0.43 to −0.09
Spanish: model 1	175					
Valence		−0.36 (0.11)	−0.30	−3.29	0.001	−0.57 to −0.15
Arousal		−0.11 (0.16)	−0.06	−0.66	0.51	−0.41 to 0.21
Spanish: model 2	175					
Valence		−0.27 (0.12)	−0.22	−2.31	0.02	−0.50 to −0.04
Arousal		−0.05 (0.16)	−0.03	−0.32	0.75	−0.36 to 0.26
Frequency		−0.16 (0.08)	−0.15	−1.94	0.05	−0.32 to 0.003
Polish: model 1	144					
Valence		−0.43 (0.14)	−0.28	−3.17	0.002	−0.70 to −0.16
Arousal		−0.17 (0.22)	−0.07	−0.77	0.44	−0.60 to 0.26
Polish: model 2	143					
Valence		−0.18 (0.15)	−0.11	−1.18	0.24	−0.46 to 0.11
Arousal		−0.25 (0.22)	−0.09	−1.11	0.27	−0.67 to 0.18
Frequency		−0.09 (0.09)	−0.10	−1.03	0.31	−0.26 to 0.09
Age of acquisition		0.49 (0.13)	0.33	3.67	<0.001	0.23 to 0.75
Dutch: model 1	151					
Valence		−0.51 (0.16)	−0.24	−3.09	0.002	−0.83 to −0.19
Arousal		0.38 (0.25)	0.13	1.50	0.14	−0.11 to 0.86
Dominance		−0.27 (0.29)	−0.08	−0.92	0.36	−0.83 to 0.30
Dutch: model 2	150					
Valence		−0.39 (0.18)	−0.18	−2.15	0.03	−0.74 to −0.04
Arousal		0.41 (0.25)	0.15	1.65	0.10	−0.07 to 0.90
Dominance		−0.31 (0.30)	−0.10	−1.05	0.30	−0.88 to 0.26
Frequency		−0.04 (0.10)	−0.04	−0.41	0.68	−0.22 to 0.15
Borrowing		−0.20 (0.18)	-.0.09	−1.14	0.25	−0.54 to 0.14
Age of acquisition		0.21 (0.17)	0.12	1.27	0.21	−0.11 to 0.53

Note. Two-tailed regression models with fixed effects mean-centred within part of speech. Sample size (*n*) denotes number of unique concepts in each analysis.

serve as robustness checks (that is, reproducing the association for each language means that our results are not confined to how English speakers alone view the positivity of words) rather than independent replications. Figure 1 shows the relationship between valence and cognate replacement in each language (for a reproduction of Fig. 1 residualized on key variables, see Supplementary Fig. 1).

We next evaluated the robustness of these associations using multilevel regressions containing covariates. Our first model in each language (Table 1, model 1) focused on semantic variables, and our second model (Table 1, model 2) included the full range of control variables. In these models, we nested words within their part of speech using a random-effects approach and centred the

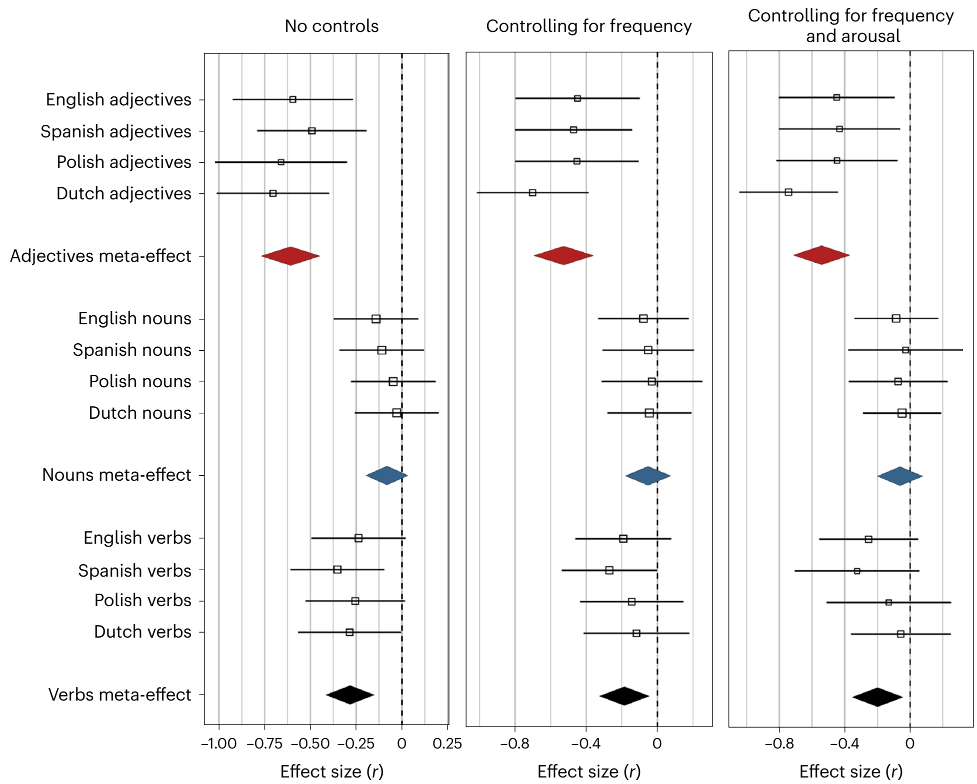

Fig. 2 | The meta-analytic relationship between valence and cognate replacement rate for different parts of speech. Node size in the plot is scaled to sample size (number of words within a part of speech within a language), and error bars represent 95% CI. Rate of cognate replacement is the outcome in all models. For adjectives, the relationship is consistently significant. For verbs, the meta-effect is consistently significant, but the relationship is significant only for specific languages without control variables. For nouns, the relationship is never significant. These models contain 168 English words, 175 Spanish words, 144 Polish words and 151 Dutch words.

Table 2 | Coefficients from part-of-speech random-effects meta-analysis models

Part of speech	Model type	Effect	95% CI
Adjectives	Zero-order	−0.71	−0.91 to −0.51
Adjectives	Controlling for frequency	−0.59	−0.79 to −0.39
Adjectives	Controlling for frequency and arousal	−0.59	−0.83 to −0.36
Nouns	Zero-order	−0.08	−0.21 to 0.04
Nouns	Controlling for frequency	−0.05	−0.18 to 0.07
Nouns	Controlling for frequency and arousal	−0.06	−0.18 to 0.06
Verbs	Zero-order	−0.29	−0.44 to −0.15
Verbs	Controlling for frequency	−0.19	−0.33 to −0.04
Verbs	Controlling for frequency and arousal	−0.20	−0.35 to −0.06

Note. Effect sizes represent *r* coefficients. CIs are two-tailed.

values of all predictors within part of speech to ensure the relationships we identify are not simply an artefact of differences in rates of cognate replacement between parts of speech (following ref. [49]; for results using a fixed-effects approach, see Supplementary Table 2). Valence was negatively associated with cognate replacement rate in all models, and this relationship reached statistical significance in all models except for one Polish model where age of acquisition was the dominant predictor (although age of acquisition was not significant in the Dutch model). The effect size of valence was larger than that of frequency of use in all models.

We also examined how the association between valence and cognate replacement rate varied within different parts of speech. Our full sample of concepts contained eight parts of speech (nouns, verbs, adjectives, adverbs, number words, conjunctions, prepositions and pronouns). However, 86% of the sample were nouns, verbs or adjectives. We focused on variation between these three major parts of speech because we had so few cases of other parts of speech. We fit a random-effects meta-analysis that generated meta-analytic estimates—across the four languages—of the magnitude of the association between valence and cognate replacement rate for each part of speech. We evaluated this model under three conditions: (1) with no control variables, (2) controlling for frequency of use and (3) controlling for frequency of use and semantic arousal, the covariates available for all four languages. Figure 2 summarizes the model estimates, and we describe and report the models in depth within the Supplementary Results (for zero-order effects for each language and part of speech, see also Supplementary Table 3).

In each model, the meta-effect of valence was largest for adjectives, followed by verbs, and was smallest for nouns. For nouns, the effect did not reach statistical significance in any model. For verbs, the meta-effect of valence was statistically significant, but valence was not significantly associated with cognate replacement in any individual language when controlling for covariates. This demonstrates that valence is most robustly associated with cognate replacement rate at the population level for adjectives, although does not rule out a weaker effect for verbs and nouns. The full statistics for the meta-effects in each model are presented in Table 2. The effect for nouns may be weak—and consistently non-significant—because of two restrictions of range.

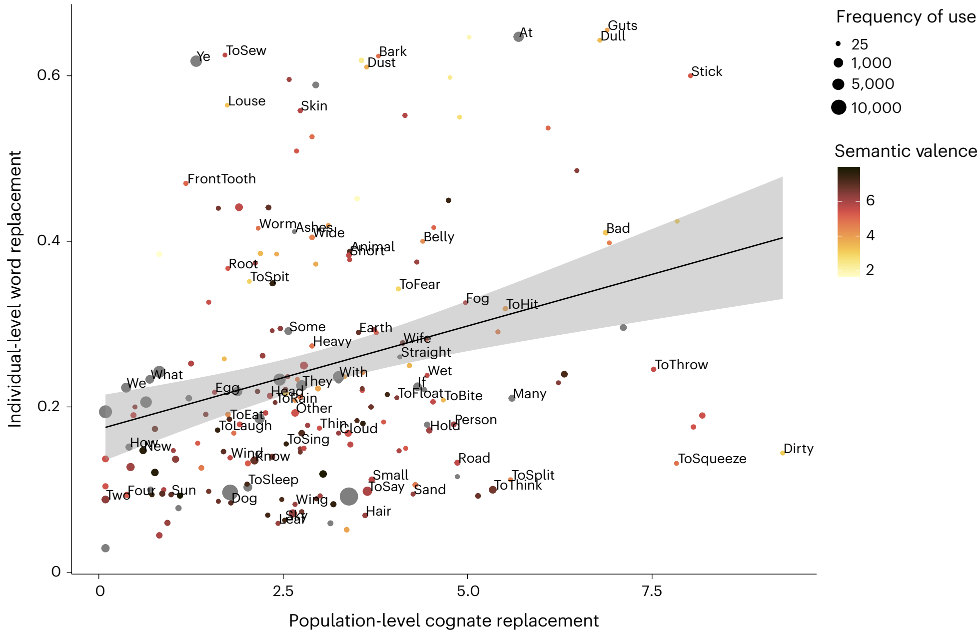

Fig. 3 | The relationship between population-level cognate replacement and individual-level word replacement. Words with high levels of population-level cognate replacement also have high rates of individual-level word replacement, on average. Node colour indicates semantic valence. Node size indicates frequency of use (usage per million words). Shading represents s.e.m. One-third of concepts (chosen randomly) are labelled. Grey nodes represent concepts that did not have semantic valence data available. This plot contains data on 168 English-language words.

Table 3 | Multilevel regression models predicting cognate replacement per 10,000 years

Predictor	*n*	*b* (s.e.m.)	*β*	*t*	*P*	95% CI
Model 1	165					
Word replacement		2.33 (0.94)	0.20	2.47	0.02	0.49 to 4.17
Frequency		−0.16 (0.11)	−0.13	−1.57	0.12	−0.37 to 0.04
Borrowing		−0.26 (0.09)	−0.21	−2.96	0.004	−0.43 to −0.09
Arousal		−0.09 (0.14)	−0.05	−0.66	0.51	−0.36 to 0.18
Model 2	165					
Word replacement		1.47 (1.04)	0.12	1.42	0.16	−0.54 to 3.48
Frequency		−0.13 (0.11)	−0.10	−1.26	0.21	−0.34 to 0.07
Borrowing		−0.26 (0.09)	−0.20	−2.95	0.004	−0.43 to −0.09
Valence		−0.22 (0.12)	−0.16	−1.94	0.05	−0.45 to −0.0001
Arousal		−0.16 (0.14)	−0.08	−1.14	0.26	−0.43 to 0.01

Note. Two-tailed regression models with fixed effects mean-centred within part of speech. Word replacement refers to individual-level word replacement rates. Frequency, semantic features and borrowing come from the English data. Sample size (*n*) denotes number of unique concepts in each analysis.

First, nouns have a significantly slower mean rate of cognate replacement (mean 2.81) than adjectives (mean 3.44) and verbs (mean 3.67) ($b = -0.77$, $P = 0.004$, 95% CI −1.28 to −0.25). Second, the nouns in our sample have a narrower range of valence than the verbs and adjectives because no nouns in our sample are highly negative. The most negative noun is more positive (minimum 2.36) than the most negative verb (minimum 1.35) and most negative adjective (minimum 2.05). However, these statistics do not explain why the association between valence and cognate replacement rate was more robust for adjectives than for verbs.

In addition to these part-of-speech analyses, Supplementary Information provides a variety of other robustness checks. Our Supplementary Methods describe model diagnostics that show that our findings are not driven by problematic outliers or influential datapoints (Supplementary Fig. 2). Our Supplementary Results describe several additional analytic procedures. For example, we describe a latent variable approach in which we used separate principal components to represent the valence and frequency of use for all languages (Supplementary Tables 4 and 5). We also show that valence does not significantly interact with arousal in our models and describe additional data collection and analyses in which we show that the effect of valence in our models cannot be explained by taboo and difficulty of communication (Supplementary Table 6).

We next analysed our individual-level behavioural data to test if individual people were also more likely to select words connoting negative (versus positive) concepts for replacement, and whether individuals' word replacement selections correlate with population-level cognate replacement rates. For the sake of parsimony, we refer to participants' word replacement selections as 'word replacement', but it is important to note that participants were selecting words to replace but not choosing a replacement in our individual-level studies.

Valence and word replacement at the individual level

We tested the relationship between individual-level word replacement and valence to evaluate whether biased individual-level replacement decisions could serve as a mechanism for the population-level association between valence and cognate replacement. A zero-order correlation found that, as with population-level cognate replacement, individual-level word replacement was negatively correlated with semantic valence ($r(165) = -0.48$, $P < 0.001$, 95% CI −0.59 to −0.36). This association replicated in a model controlling for arousal, frequency of use and borrowing, and centring all terms within part of speech to control for potential part of speech effects. In this model, valence ($b = -0.05$, $\beta = -0.41$, standard error of the mean (s.e.m.) 0.008, $t = -2.83$, $P < 0.001$, 95% CI −0.06 to −0.03) and frequency of use ($b = -0.04$, $\beta = -0.34$, s.e.m. 0.007, $t = -5.03$, $P < 0.001$, 95% CI −0.05 to −0.02) were both significantly and negatively associated with word replacement rates. Arousal and borrowing were not significant predictors (for full set of coefficients, see Supplementary Table 7; for association between valence and cognate replacement rate in this study broken down by part of speech, see Supplementary Table 3).

We also performed a pre-registered follow-up study in which we no longer required participants to replace words, and we implemented several means of reducing researcher demand characteristics, including a direct probe for demand at the end of the study. This supplementary study also used a larger and more diverse sample of 1,200 concepts. Our Supplementary Methods describe the procedure and methods of this study in detail, and our Supplementary Results section describe the findings of the study (Supplementary Table 8). This follow-up replicated the strong negative association between valence and word replacement for each part of speech, which suggests that this is a robust association that generalizes beyond the Swadesh list sample of concepts.

Individual-level and population-level replacement

We next tested whether individual-level word replacement rates of Swadesh list concepts in our individual-level study were correlated with the population-level cognate replacement rates over 6–10 million years of linguistic evolution. Analyses showed a symmetry in replacement rates across levels of analysis, such that words that were most frequently replaced by individuals in our study also tended to have higher rates of cognate replacement in our population-level dataset (Fig. 3) ($r(197) = 0.29$, $P < 0.001$, 95% CI 0.16 to 0.42). Individual-level word replacement remained a significant predictor of population-level cognate replacement when controlling for borrowing rate, frequency of use and arousal (Table 3, model 1). When valence was added to the model, word replacement was no longer significant, showing that valence partly accounts for why concepts with high levels of cognate replacement at the population level also have a high rate of individual-level word replacement (Table 3, model 2).

In sum, we find robust evidence that negative adjectives mutate faster than positive adjectives over lexical evolution, and individual people tend to replace negative words more than positive words when they modify language. Moreover, current-day English speakers in an online survey tend to replace the same words that have had rapid rates of cognate replacement over the last 10,000 years of Indo-European lexical evolution.

Discussion

The word for 'good' in English has cognate forms in German and Faroese. However, the word for 'bad' is expressed as 'bad' in English, 'schlecht' in German and 'illur' in Faroese. Here we show that this difference signifies a broader pattern that we call 'valence-dependent mutation', in which negative valence correlates with a faster rate of cognate replacement in lexical evolution across a language family representing 6–10 millennia of language change. Valence-dependent mutation showed similar characteristics across all four Indo-European languages in this study: it consistently reached statistical significance for adjectives, it reached statistical significance for verbs without control variables but not when controlling for frequency of use and semantic arousal, and it did not reach statistical significance for nouns.

We replicate valence-dependent mutation at the individual level by showing that people are more likely to replace words for negative concepts than positive concepts—across parts of speech—when they create a new dialect of English. Our individual-level study observed a correlation between individual-level word replacement decisions and population-level cognate replacement rates: words that people frequently replaced were also the words with the highest rates of cognate replacement over thousands of years of language evolution. This dynamic suggests that people's preferences about language may drive valence-dependent mutation. A supplementary individual-level study found that this effect was robust to the subset of concepts sampled and to the potential for experiment-induced demand on participants' selections. Our results show how insights from cognitive and affective science may enrich our understanding of lexical evolution.

Valence-dependent mutation not only informs evolutionary linguistics, it may also explain why negative concepts are often more 'granular', meaning that they are more numerous and differentiated, and have more specific meanings than positive concepts[50–52]. For instance, there are approximately 66% more words that name negative emotions than words that name positive emotions in the English language[53–55], and 75% more words that describe negative than positive personality traits[53,56]. The origins of this asymmetry have long been a mystery in psychology[57], but we suggest that valence-dependent mutation may be one mechanism for this phenomenon, since newly replaced words often remain in lexicons and lead to larger lexicons. For example, the English adjective 'sleazy' was introduced in the seventeenth century to express the same concept as the older English word 'sordid'[58], yet 'sordid' remains in use today. There may be other mechanisms behind the higher relative granularity of negative compared with positive words, but we consider valence-dependent mutation to be a potential key to this puzzling phenomenon.

Our correlational data cannot conclusively identify the mechanism behind valence-dependent mutation, which is a limitation of this paper. However, we can partly rule out some potential mechanisms. For adjectives at least, the association between valence and lexical evolution was robust to controlling for frequency of use. Negative adjectives are therefore unlikely to evolve faster than positive adjectives simply because they are less frequent. Our Supplementary Results also describe additional norming data that we gathered from a set of online raters on difficulty of communication and level of taboo. These analyses showed that taboo and difficulty of communication could not account for valence-dependent mutation, suggesting that negative adjectives do not evolve rapidly simply because they become taboo, or because they are difficult to communicate.

We propose that the cultural evolutionary process of guided variation is an important mechanism underlying valance-dependent mutation. Guided variation refers to biased intra-individual transformations, which then spread throughout cultural groups[8]. There are several mechanisms of guided variation that could underlie our findings. One potential mechanism is that people may be more motivated to precisely describe negative phenomena than positive phenomena, leading them to develop new negative adjectives at a faster rate than positive adjectives. This 'descriptive precision' account could explain why population-level valence-dependent mutation was only consistently significant for adjectives, as adjectives are descriptors by definition. However, this mechanism does not explain why participants in our individual-level studies also replaced negative words more frequently than positive words for verbs and nouns. Potential general mechanisms of valance-dependent mutation are that negative words elicit more attention than positive words[34,35], or that people apply more analytic and effortful processing when they evaluate negative versus positive information, including negative language[39,40]. People may also have a bias towards retaining positive words because they view them more favourably, which would be consistent with the Pollyanna principle[41]. These more general mechanisms raise the possibility that valance-dependent mutation also occurs across different parts of speech at the population level, but our present findings cannot confirm or rule out this possibility. Therefore, we encourage future research that estimates valence-dependent mutation in large and heterogeneous samples of words and investigates the different mechanisms that we raise here.

An alternative possibility is that valence-dependent mutation arises through transmission biases. People may socially transmit positive adjectives with higher fidelity than negative adjectives. This possibility seems unlikely because transmission chain studies also show that people tend to transmit negative information with higher fidelity than positive information when they communicate information from one person to another[12]. People also recognize negative information more accurately than positive information[59]. Nevertheless, transmission effects would still be interesting to explore in future research because they can indicate how people's motivation to distort negative information in social interactions through innuendo or other audience-tuning strategies can accelerate the pace of valence-dependent mutation[60–62].

Our research also speaks to differences between lexical and biological evolution, and the nature of cultural evolution more broadly. There are examples of biased mutation in genetic evolution, such as contingency genes in *Hemophilus influenza* that allow the bacteria to mutate in adaptive ways in response to host antibodies during periods of stress[63]. However, there is broad consensus that cultural evolution includes a higher volume of non-random transformation than biological evolution, and that psychological preferences and biases influence the nature of this non-random transformation[6]. Our studies examine intra-individual processes that contribute to this non-random transformation, and we encourage future studies that focus on inter-individual mechanisms of biased mutation in language change, given the importance of sociality in human life[64]. Bridging micro- and macro-processes of cultural evolution promises to identify how individual people's motivations and cognitive biases may culminate in large-scale changes in language and culture across human history.

Methods

Population-level data

Lexical evolution properties. The first step in our population-level survey involved collecting data on cognate replacement rates—rates in which a cognate was replaced with a non-cognate—for word forms associated with a wide variety of concepts. We obtained these data from Pagel, Atkinson and Meade[17], based on their estimates of cognate replacement rates for 200 concepts that make up the Swadesh word list[47], a list of basic vocabulary terms thought to be relatively universal and relatively resistant to borrowing. Words for each concept were grouped into between 1 and 46 cognate sets (for example, words for the concept of big fell into 24 different cognate sets across the 87 Indo-European languages in the dataset). The paper then applied a statistical likelihood model of word evolution to the phylogenetic trees of Indo-European languages (for more details, see ref. [17]). For each of the 200 concepts, the authors used a likelihood model of word evolution and Bayesian inference framework to infer the rate of cognate replacement—the rate at which the word for a concept will be replaced with a non-cognate form—within an Indo-European language. Inferred rates varied by a factor of 10, from fewer than one cognate replacement per 10,000 years for concepts such as two and one, to as many as nine cognate replacements per 10,000 years for concepts such as dirty, stab and guts. While the age of the Indo-European language family is controversial, scaling the rates of cognate replacement to match different time depths for the family would change the absolute replacement rates but not the relative rates that were the focus of the present analysis.

Semantic properties. We collected data on concepts' semantic properties from four Indo-European languages: English, Spanish, Polish and Dutch. These languages span three of the major subclades of the Indo-European language family tree (Germanic, Italic and Balto-Slavic), to ensure our semantic data reflects general features of the Indo-European language family. We analysed the relationship between semantic valence and cognate replacement separately for each language to test whether our results were similar in magnitude across these diverse languages.

We obtained English-language norming data on valence and arousal from Warriner and colleagues[43] (n = 1,827 raters), Spanish-language data on valence and arousal from Stadthagen-Gonzalez and colleagues[44] (n = 512 raters), Polish-language data on valence and arousal from Imbir[45] (n = 400 raters) and Dutch-language data on valence, arousal and dominance from Moors and colleagues[46] (n = 224 raters). Not all concepts had semantic data available for all four languages. Of the 200 concepts for which we had cognate replacement data, 168 were available in our English norming data, 175 in Spanish, 144 in Polish and 151 in Dutch. These missing data meant that our analyses varied in degrees of freedom across languages (Table 1). Some concepts also had multiple words (that is, 'culebra' and 'serpiente' in Spanish for the concept of 'snake'). Native speakers of these language were consulted for each of these discrepancies so that we could choose the more appropriate lexical form. Supplementary Table 1 contains our concept list and the cognate replacement rate for each concept.

Frequency of use. In addition to data on semantic properties, we used frequency-of-use data on English from Warriner and colleagues[43], on Spanish from Davies[65], on Polish from Imbir[45] and on Dutch from Moors[46]. Consistent with past research, we log-transformed frequency of use before analyses. All significant results remain significant regardless of this transformation.

Borrowing. Words can be borrowed from one language to another (for example, the word 'schadenfreude' is borrowed in English from German). Borrowings between the Indo-European languages in our sample could potentially inflate the inferred age and stability of word forms within our phylogenetic analyses. While we have no reason to suspect that borrowing explains valance-dependent mutation, we nevertheless controlled for borrowing when data were available as a precaution. We obtained borrowing data on 196 of the 200 concepts from the World Loanword Database (WOLD)[66]. WOLD uses a 1–5 rating system where 1 represents words that are 'clearly borrowed', 2 represents words that are 'probably borrowed', 3 represents words that are 'perhaps borrowed', 4 represents words that are 'probably not borrowed' and 5 represents words that have 'no evidence of borrowing'. As the vast majority of concepts were classified as either 'clearly borrowed' or 'no evidence of borrowing', we collapsed the 5-point rating scale into a dummy-coded variable where words received a 0 if they had no evidence of borrowing, and a 1 if they had any evidence of borrowing (categories 1–4). A positive relationship between borrowing and cognate replacement would therefore denote that the concepts underlying borrowed words in a particular language had higher estimates of cognate replacement than the concepts underlying non-borrowed words. Results were substantively identical if we used the full rating scale or used a dummy-coded variable that collapsed categories 1–3 and 4–5. WOLD did not contain data on Spanish and Polish, so we used the English and Dutch data.

Part of speech. In their dataset, Pagel, Atkinson and Meade[17] classified words as adjectives (n = 41), adverbs (n = 7), conjunctions (n = 3), nouns (n = 75), numbers (n = 5), prepositions (n = 3), pronouns (n = 9) and verbs (n = 57). Our models nested words within these parts of speech using a random-effects approach and centred fixed effects within part of speech. For example, each word's semantic valence was scaled on the basis of the mean valence for the part of speech. This controlled for any confounding variance associated with part of speech effects. Results are the same if, instead of centring, we added dummy-coding fixed effects representing parts of speech to our regressions (Supplementary Table 2).

Age of acquisition. Age-of-acquisition data were available in our Dutch[46] and Polish[45] norming datasets.

Statistical test notes. All statistical tests reported in this paper are two-tailed. The Supplementary Methods section includes model diagnostics that show that our data meet the assumptions of our statistical tests (Supplementary Fig. 2). All analyses were performed in R Studio version 1.1.383.

Individual-level data

We pre-registered all properties of this study, including the sample sizes, exclusion criteria, procedures, designs and analytic plans. The pre-registration can be found at https://osf.io/f86wt/. We obtained approval from the University of North Carolina institutional review board.

Sample. We advertised for 500 participants from Amazon Mechanical Turk. We did not conduct a power analysis, but we deemed this sample size appropriately large to detect a meaningful effect. We advertised the study to workers who had an approval rate of 95% or higher. In total, 528 participants enroled in the survey, but we excluded 34 because they did not finish the study, or they failed our attention check (see 'Procedure'). The remaining 494 participants (age: mean 40.60 years, standard deviation 13.22 years; 248 men, 246 women) received 50 cents (USD) for their participation in the study.

Procedure. At the beginning of the study, we instructed participants to imagine that they were part of a new settlement on a foreign planet, and their job in this settlement was to establish a new language modelled after the English language. They would need to keep some existing English words but replace other English words with new words that convey the same meaning. Participants then viewed ten sets of four words, which we took from the 200 concepts in our macro-level study (all participants therefore saw all words). The word sets were sampled pseudo-randomly: we matched word sets on the basis of their part of speech (adjectives were presented with adjectives) and first letter, but word sets were otherwise randomized. This matching ensured that valence was not confounded with other aspects of language. Participants viewed each set in a multiple-choice format, and they were asked 'which of the following English words would you like to replace?' Participants then provided demographics, which included an attention check that asked participants for their favourite hobby in bold but specified in plain text that participants should select 'gardening' if they were paying attention.

Replacement rates. Replacement rates represented proportions, bounded by 0 and 1, such that a value of 0 meant that a word was never replaced and a value of 1 meant that a word was replaced every time it was viewed. Word replacement rate was normally distributed with a slight positive skew (1.02). We did not transform the variables in our analyses because residuals in our regression models appeared to be normally distributed, but results are virtually identical if we use log-transformation.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Our project page at https://osf.io/f86wt/ displays all data from this paper. Our analyses used external data from WOLD (https://wold.clld.org/) and from Pagel, Atkinson and Meade (https://www.nature.com/articles/nature06176?message=remove&pagewanted=all).

Code availability

Our project page at https://osf.io/f86wt/ displays all code from this paper.

References

1. Lewis, M. P. *Ethnologue: Languages of the World* (SIL International, 2009).
2. Croft, W. Evolutionary linguistics. *Annu. Rev. Anthropol.* **37**, 219–234 (2008).
3. Pagel, M. Human language as a culturally transmitted replicator. *Nat. Rev. Genet.* **10**, 405–415 (2009).
4. Gray, R. D. & Atkinson, Q. D. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* **426**, 435–439 (2003).
5. Jackson, J. C. et al. From text to thought: how analyzing language can advance psychological science. *Perspect. Psychol. Sci.* **17**, 805–826 (2021).
6. Dor, D. & Jablonka, E. From cultural selection to genetic selection: a framework for the evolution of language. *Selection* **1**, 33–56 (2001).
7. Jablonka, E. & Lamb, M. J. Précis of evolution in four dimensions. *Behav. Brain Sci.* **30**, 353 (2007).
8. Boyd, R. & Richerson, P. J. *Culture and the Evolutionary Process* (Univ. Chicago Press, 1988).
9. Sperber, D. *Explaining Culture: A Naturalistic Approach* (Blackwell, 1996).
10. Mesoudi, A. & Whiten, A. The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philos. Trans. R. Soc. B* **363**, 3489–3501 (2008).
11. Acerbi, A. & Alexander Bentley, R. Biases in cultural transmission shape the turnover of popular traits. *Evol. Hum. Behav.* **35**, 228–236 (2014).
12. Bebbington, K., MacLeod, C., Ellison, T. M. & Fay, N. The sky is falling: evidence of a negativity bias in the social transmission of information. *Evol. Hum. Behav.* **38**, 92–101 (2017).
13. Berger, J. Arousal increases social transmission of information. *Psychol. Sci.* **22**, 891–893 (2011).
14. Pennebaker, J. W., Mehl, M. R. & Niederhoffer, K. G. Psychological aspects of natural language use: our words, our selves. *Annu. Rev. Psychol.* **54**, 547–577 (2003).
15. Nettle, D. Is the rate of linguistic change constant? *Lingua* **108**, 119–136 (1999).
16. Thomason, S. G. Language change and language contact. *Encycl. Lang. Linguist.* **6**, 339–347 (2006).
17. Pagel, M., Atkinson, Q. D. & Meade, A. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history. *Nature* **449**, 717–720 (2007).
18. Wichmann, S. & Holman, E. W. in *Approaches to Measuring Linguistic Differences* (eds Borin, L. & Saxena, A.) (De Gruyter, 2013).
19. Monaghan, P. Age of acquisition predicts rate of lexical evolution. *Cognition* **133**, 530–534 (2014).
20. Vejdemo, S. & Hörberg, T. Semantic factors predict the rate of lexical replacement of content words. *PLoS ONE* **11**, e0147924 (2016).
21. Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. Bad is stronger than good. *Rev. Gen. Psychol.* **5**, 323–370 (2001).
22. Jefferies, L. N., Smilek, D., Eich, E. & Enns, J. T. Emotional valence and arousal interact in attentional control. *Psychol. Sci.* **19**, 290–295 (2008).
23. Kuperman, V., Estes, Z., Brysbaert, M. & Warriner, A. B. Emotion and language: valence and arousal affect word recognition. *J. Exp. Psychol. Gen.* **143**, 1065 (2014).
24. Jackson, J. C. et al. Emotion semantics show both cultural variation and universal structure. *Science* **366**, 1517–1522 (2019).
25. Di Natale, A., Pellert, M. & Garcia, D. Colexification networks encode affective meaning. *Affect. Sci.* **2**, 99–111 (2021).

26. Eimer, M. & Holmes, A. Event-related brain potential correlates of emotional face processing. *Neuropsychologia* **45**, 15–31 (2007).
27. Fischler, I. & Bradley, M. Event-related potential studies of language and emotion: words, phrases, and task effects. *Prog. Brain Res.* **156**, 185–203 (2006).
28. Schupp, H. T., Junghöfer, M., Weike, A. I. & Hamm, A. O. The selective processing of briefly presented affective pictures: an ERP analysis. *Psychophysiology* **41**, 441–449 (2004).
29. Osgood, C. E., & Tzeng, O. C. S. (eds). *Language, Meaning, and Culture: The Selected Papers of C. E. Osgood* (Praeger, 1990).
30. Bloom, P. *How Children Learn the Meanings of Words* (MIT Press, 2002).
31. Broze, Y. & Shanahan, D. Diachronic changes in jazz harmony: a cognitive perspective. *Music Percept. Interdiscip. J.* **31**, 32–45 (2013).
32. DeWall, C. N., Pond, R. S. Jr, Campbell, W. K. & Twenge, J. M. Tuning in to psychological change: linguistic markers of psychological traits and emotions over time in popular US song lyrics. *Psychol. Aesthet. Creat. Arts* **5**, 200 (2011).
33. Morin, O. & Acerbi, A. Birth of the cool: a two-centuries decline in emotional expression in Anglophone fiction. *Cogn. Emot.* **31**, 1663–1675 (2017).
34. Pratto, F. & John, O. P. Automatic vigilance: the attention-grabbing power of negative social information. *J. Pers. Soc. Psychol.* **61**, 380 (1991).
35. Fiske, S. T. Attention and weight in person perception: the impact of negative and extreme behavior. *J. Pers. Soc. Psychol.* **38**, 889 (1980).
36. Brady, W. J., Crockett, M. J. & Van Bavel, J. J. The MAD model of moral contagion: the role of motivation, attention, and design in the spread of moralized content online. *Perspect. Psychol. Sci.* **15**, 978–1010 (2020).
37. Taylor, S. E. Asymmetrical effects of positive and negative events: the mobilization-minimization hypothesis. *Psychol. Bull.* **110**, 67 (1991).
38. Bless, H. & Fiedler, K. Mood and the regulation of information processing and behavior. in *Affect in Social Thinking and Behavior* (ed Forgas, J. P.) 65–84 (Psychology Press, 2006).
39. Forgas, J. P. Feeling and doing: affective influences on interpersonal behavior. *Psychol. Inq.* **13**, 1–28 (2002).
40. Schwarz, N. & Bless, H. in *Emotion and Social Judgments* (ed Forgas, J. P.) 55–71 (Garland Science, 2020).
41. Matlin, M. W. & Stang, D. J. *The Pollyanna Principle: Selectivity in Language, Memory, and Thought* (Schenkman, 1978).
42. Forgas, J. P. Mood and the perception of unusual people: affective asymmetry in memory and social judgments. *Eur. J. Soc. Psychol.* **22**, 531–547 (1992).
43. Warriner, A. B., Kuperman, V. & Brysbaert, M. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods* **45**, 1191–1207 (2013).
44. Stadthagen-Gonzalez, H. et al. Norms of valence and arousal for 14,031 Spanish words. *Behav. Res. Methods* **49**, 111–123 (2017).
45. Imbir, K. K. Affective norms for 4900 Polish Words Reload (ANPW_R): assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and, age of acquisition. *Front. Psychol.* **7**, 1081 (2016).
46. Moors, A. et al. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behav. Res. Methods* **45**, 169–177 (2013).
47. Swadesh, M. Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. *Proc. Am. Philos. Soc.* **96**, 452–463 (1952).
48. Croft, W. *Explaining Language Change: An* Evolutionary Approach (Pearson Education, 2000).
49. Enders, C. K. & Tofighi, D. Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychol. Methods* **12**, 121 (2007).
50. Mulkar-Mehta, R., Hobbs, J. & Hovy, E. Granularity in natural language discourse. In *Proc. Ninth International Conference on Computational Semantics* (IWCS, 2011).
51. Smidt, K. E. & Suvak, M. K. A brief, but nuanced, review of emotional granularity and emotion differentiation research. *Curr. Opin. Psychol.* **3**, 48–51 (2015).
52. Tugade, M. M., Fredrickson, B. L. & Barrett, L. F. Psychological resilience and positive emotional granularity: examining the benefits of positive emotions on coping and health. *J. Pers.* **72**, 1161–1190 (2004).
53. Averill, J. R. *A Semantic Atlas of Emotional Concepts* (Am. Psychol. Assoc., 1975).
54. Clore, L. & Ortony, A. in *Cognitive Perspectives on Emotion and Motivation* (eds Hamilton, V., Bower, G. H. & Frijda, N. H.) 367–397 (Springer, 1988).
55. Russell, J. A., Fernández-Dols, J. M., Manstead, A. S. R. & Wellenkamp, J. *Everyday Conceptions of Emotion* (Springer, 1995).
56. Anderson, N. H. Likableness ratings of 555 personality-trait words. *J. Pers. Soc. Psychol.* **9**, 272–279 (1968).
57. Alves, H., Koch, A. & Unkelbach, C. Why good is more alike than bad: processing implications. *Trends Cogn. Sci.* **21**, 69–79 (2017).
58. Harper, D. Etymology, origin and meaning of sleazy by etymonline. *Online Etymology Dictionary* https://www.etymonline.com/word/sleazy (2000).
59. Ortony, A., Turner, T. J. & Antos, S. J. A puzzle about affect and recognition memory. *J. Exp. Psychol. Learn. Mem. Cogn.* **9**, 725 (1983).
60. Echterhoff, G., Higgins, E. T., Kopietz, R. & Groll, S. How communication goals determine when audience tuning biases memory. *J. Exp. Psychol. Gen.* **137**, 3 (2008).
61. Echterhoff, G., Higgins, E. T. & Groll, S. Audience-tuning effects on memory: the role of shared reality. *J. Pers. Soc. Psychol.* **89**, 257 (2005).
62. Ye, J., Zhao, L., Huang, Z. & Meng, F. The audience-tuning effect of negative stereotypes in communication. *Front. Psychol.* **12**, 663814 (2021).
63. Bayliss, C. D., Field, D. & Moxon, E. R. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J. Clin. Invest.* **107**, 657–666 (2001).
64. Lieberman, M. D. *Social: Why Our Brains Are Wired to Connect* (Crown, 2013).
65. Davies, M. Corpus del español: 10 billion words: dialects / genres / historical. https://www.corpusdelespanol.org/
66. Haspelmath, M. & Tadmor, U. *Loanwords in the World's Languages: A Comparative Handbook* (Walter de Gruyter, 2009).

Acknowledgements
J.W. thanks the Marsden Foundation of New Zealand (19-UOO-1932) for funding. The Marsden foundation played no role in the conceptualization, design, analysis or decision to publish this research. The authors thank I. Khismatova for research assistance.

Author contributions
J.C.J. conceptualized the study, analysed the data and co-wrote the manuscript. K.L. co-wrote the manuscript. R.D. collected the data. Q.A. co-wrote the manuscript. J.W. conceptualized the study and co-wrote the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-022-01483-8.

Correspondence and requests for materials should be addressed to Joshua Conrad Jackson.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature portfolio

Corresponding author(s): Joshua Conrad Jackson

Last updated by author(s): October 12, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used Qualtrics for data collection
Data analysis	All analyses were performed in R studio version 1.1.383. Our code is available at https://osf.io/f86wt/

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data that support the findings and figures of this study are available at https://osf.io/f86wt/. Our analyses used external data from the World Loanword Database (https://wold.clld.org/) and from Pagel, Atkinson, and Meade (https://www.nature.com/articles/nature06176?message=remove&pagewanted=all).

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	The 494 participants comprised 248 self-identified men and 246 women.
Population characteristics	Mage = 40.60, SDage = 13.22; 248 men, 246 women
Recruitment	Cloudresearch participant pool, which is well-established in behavioral science research. Participants self-selected into the study by clicking a link on Cloudresearch. The study (10-minute study) was named in such a way that participants would not self-select based on demographic characteristics or pre-existing attitudes.
Ethics oversight	University of North Carolina IRB Board. We obtained informed consent from all participants prior to data collection.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Our first study draws from large cross-cultural semantic norming studies that have measured the semantic properties of thousands of concepts in four languages. We combine these data with cognate replacement for 200 such concepts across 87 Indo-European languages over 6,000-10,000 years. Our second study uses individual-level human behavior to test whether the semantic preferences of contemporary English speakers can account for the association between valence and lexical evolution.
Research sample	Our first study sampled Indo-European languages. The Indo-European language family is one of the largest and most diverse language families in the world, covering languages in South Asia, Central Asia, and Europe, and dating back thousands of years. Indo-European languages also have been well studied by past linguistics research, making it possible to incorporate data on borrowing rates, word frequencies, and other semantic properties (e.g., physiological arousal, dominance) as control variables in our analysis. Our second sample was English-language speakers from the United States. We sought a sample of contemporary language speakers to test whether their word replacement decisions would mirror historical cognate replacement rates across Indo-European languages.
Sampling strategy	We focused on norming data in four Indo-European languages: English, Spanish, Polish, and Dutch. These languages span three of the major sub-clades of the Indo-European language family tree (Germanic, Italic and Balto-Slavic), so ensure our semantic data reflects general features of the Indo-European language family. These languages also had semantic norming data available. We obtained English-language norming data on Valence and Arousal from Warriner and colleagues (n = 1,827 raters), Spanish-language data on Valence and Arousal from Stadthagen-Gonzalez and colleagues (n = 512 raters), Polish-language data on Valence and Arousal from Imbir (n = 400 raters), and Dutch-language data on Valence, Arousal, and Dominance from Moors and colleagues (n = 224 raters). Not all concepts had semantic data available for all four languages. Of the 200 concepts for which we had cognate replacement data, 168 were available in our English norming data, 175 in Spanish, 144 in Polish, and 151 in Dutch. We obtained individual-level word replacement data from a sample of 494 participants. We estimated that a sample of 500 participants would give us adequate power to detect variation in replacement likelihood across our sample of concepts.
Data collection	We obtained these data from Pagel, Atkinson, and Meade1, based on their estimates of cognate replacement rates for 200 concepts that make up the Swadesh word list, a list of basic vocabulary terms thought to be relatively universal and relatively resistant to borrowing. Words for each concept were grouped into between 1 and 46 cognate sets (e.g., words for the concept of big fell into 24 different cognate sets across the 87 Indo-European languages in the dataset). The paper then applied a statistical likelihood model of word evolution to the phylogenetic trees of Indo-European languages. For each of the 200 concepts, the authors used a likelihood model of word evolution and Bayesian inference framework to infer the rate of cognate replacement across the Indo-European language family. Inferred rates varied by a factor of 10, from less than one cognate replacement per 10,000 years for concepts such as two and one, to as many as nine cognate replacements per 10,000 years for concepts such as dirty, stab, and guts. While the age of the Indo-European language family is controversial, scaling the rates of cognate replacement to match different time depths for the family would change the absolute replacement rates but not the relative rates that were the focus of the present analysis. Our data on semantic-norming and individual-level word replacement was collected using online surveys. Our data in Study 2 was hosted on Cloud Research and the survey was hosted on Qualtrics.
Timing and spatial scale	Estimates of cognate replacement rates spanned 6-10 millennia of language evolution.

Data exclusions	In our second study, we excluded 34 participants because they didn't finish the study, or they failed our attention check
Reproducibility	We reproduced our analyses across four languages. However, it is important to note that these are not independent findings because cognate replacement rate is estimated across the whole Indo-European language phylogeny, meaning that each language in our analysis had the same cognate replacement data. This means that our language-specific analyses should be viewed as robustness checks (i.e., reproducing the association for each language means that our results are not confined to how English speakers alone view the positivity of words) rather than independent replications. We have no evidence to suggest that our findings are not reproducible.
Randomization	No random assignment in our main text studies, but we did include multiple covariates. We controlled for frequency of use, borrowing, and other semantic properties. In our supplemental materials, we consider other covariates such as part of speech, age of acquisition, level of taboo, and difficulty of communication.
Blinding	The research assistant who gathered our semantic norming data was blind to our hypothesis.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

March 2021