Abstract concepts and emotion

Winter, Bodo

DOI:
https://doi.org/10.1098/rstb.2021.0368

Abstract concepts and emotion:

Cross-linguistic evidence and arguments against affective embodiment

Bodo Winter

Department of English Language and Linguistics

University of Birmingham

Abstract

How are abstract concepts represented in the mind? A prominent proposal suggests that concepts such as 'freedom' and 'democracy' are grounded in emotion. Supporting this 'affective embodiment' proposal, abstract concepts have been shown to have a statistical preponderance to be more strongly positive or more strongly negative. This paper demonstrates that this finding generalizes across languages by synthesizing rating data from Cantonese, Mandarin Chinese, Croatian, Dutch, French, German, Indonesian, Italian, Polish, and Spanish. However, a deeper look at the same data suggests that the idea of emotional grounding only characterizes a small subset of abstract concepts. Moreover, when the concreteness/abstractness dimension is not operationalized using concreteness ratings, it is actually found that concrete concepts are more emotional than abstract ones. Altogether, these results suggest strong limitations to the idea that emotion is a strong factor in the grounding of abstract concepts.

Keywords: abstract concepts; embodied cognition; semantic memory; emotion; concreteness; embodiment

1. Introduction

One of the most prominent distinctions in the study of concepts is that between abstract and concrete concepts. Words for abstract concepts are generally processed more slowly [1–10], memorized less accurately [11–14], and acquired later than words for concrete concepts [15–18]. Although there is controversy about what exactly characterizes the concreteness/abstractness divide [19–21], most researchers consider concepts to be abstract "if they do not apply to physical objects that we can touch, see, feel, hear, smell or taste" ([19], p. 1). Due to their lack of sensorimotor content, such concepts are generally seen as a challenge to grounded or embodied theories of cognition [22,23]. These theories posit that comprehending a concept involves performing mental simulations that engage sensorimotor systems in the brain [24–28], such as activating leg-related neurons when processing the verbal concept \`kick\` [29,30]. It seems that any theory that exclusively relies on sensorimotor simulation as a process of comprehension would have problems explaining how abstract concepts are comprehended [23,31]. After all, how could abstract concepts such as 'virtue', 'idea', or 'democracy' involve the activation of sensorimotor systems when the content of these concepts has so little to do with anything that can be directly perceived or acted upon in a physical manner?

Accepting "the challenge of abstract concepts" [22], researchers within the embodied tradition have proposed what some call "multiple representation" theories of abstract concepts [32], perhaps most prominently expressed in the 'Words as Social Tools' (WAT) account by Borghi and colleagues [33]. Based on this view, abstract concepts are supported by multiple distinct cognitive systems. The most dominant cognitive system supporting abstract concepts is generally thought to be language, with virtually any proposal about the nature of abstract concepts acknowledging the importance of linguistic or language-like representations [33–36]. Decades worth of research under the banner of Paivio's dual coding theory suggests that abstract concepts rely more exclusively on a verbal mode of representation than concrete concepts [37–41]. Neuro-imaging research finds enhanced hemodynamic activity for abstract as opposed to concrete concepts in left-hemisphere language networks [1,12,42–44]. In addition, language competence predicts improved processing of abstract words in children [45], and the richness of linguistic contexts facilitates abstract concepts more than concrete ones [46]. Thus, a large body of behavioral and neuroscientific evidence points to language as a key factor in the representation and processing of abstract concepts.

However, researchers have not been satisfied with the idea that abstract concepts are mentally represented merely in terms of language. Embodied cognition researchers generally follow the supposition that all concepts, including abstract ones, have to have some level of "grounding" in systems external to language. This has led to proposals that abstract concepts are supported by interoception [47], simulations of situated experiences [21,48], social metacognition [33,49], and conceptual metaphor [50]. Another prominent proposal that has attracted more attention recently argues that abstract concepts make up for their lack of sensorimotor content by virtue of being represented in terms of affective content [16,17,35,45,51–54]. Yao et al. [7] describe this account as one of "representational substitution," a term that encapsulates the idea that emotional experience may *substitute* for the sensorimotor experience that abstract concepts lack. This position has also been dubbed "emotional grounding" [4], "affective grounding" [55], or "affective embodiment" [33]. Henceforth, we will use "emotional grounding" as a cover term.

This paper will present evidence that at first sight seems to support the emotional grounding of abstract concepts by generalizing the account to data from new languages, adding a much-needed cross-linguistic dimension to a literature that is largely focused on abstract concepts *in English*. However, a closer look at the same data reveals that the idea of emotional grounding characterizes, if at all, only a very small minority of concepts. Much of the recent discussion on abstract concepts has argued that we need to recognize that abstract concepts as a group are characterized by heterogeneity [19,49,56]. Fully in line with this emerging recognition that there are varieties of distinct abstract concepts, this paper argues that emotional grounding has been erroneously predicated on *all* abstract concepts. Finally, along with many other critics of concreteness ratings [19,20,57–59], this paper empirically demonstrates that once we assess the idea of emotional grounding without concreteness ratings, opposing results are obtained.

2. Background: How does emotion relate to the concrete/abstract divide?

It is important to review the available evidence for emotional grounding, as well as existing empirical studies that already argue against this hypothesis. Papers on emotional grounding generally only discuss the idea that abstract concepts are higher in emotional content without explicitly discussing contriving evidence that has already been published. Because of this, this paper attempts to deliver a more comprehensive review of studies that look at the intersection of abstract concepts and emotionality.

Evidence for emotional grounding. A key piece of evidence for emotional grounding is the fact that emotional valence ratings are related to concreteness ratings following an inverted U shape pattern, with relatively more strongly positive as well as relatively more strongly negative concepts being relatively less concrete [17,54]. Newcombe et al. [52] furthermore show that a measure of 'emotional experience' — the relative ease with which a concept evokes emotional experiences — is negatively correlated with concreteness ratings (r = -0.26). Similarly, Villani and colleagues [49] demonstrated a positive correlation between abstractness ratings and emotionality ratings for Italian (r = 0.24).

A key reaction time study in this field of research was conducted by Kousta et al. [16], who unexpectedly found that abstract concepts are processed *faster* than concrete ones when important lexical variables are controlled for. They relate this observation to the fact that emotionally valenced words are processed faster than neutral words [7,60,61]. This interpretation is supported by an additional analysis which shows that the residual processing speed advantage of abstract concepts is accounted for by entering emotional valence as a covariate [16].

Newcombe et al. [52] show that for abstract words, emotional experience was associated with faster and more accurate semantic categorizations. Pauligk et al. [4] found that higher positive or negative emotional valence led to lower error rates specifically for abstract but not concrete words in a lexical decision task. Moffat et al. [51] show that emotional experience facilitated responses to abstract words in a verbal semantic categorization task, but only when blocking stimuli by emotional experience drew attention to the emotional dimension. On top of facilitation effects, Siakaluk et al. [53] were also able to demonstrate semantic interference effects of emotional experience for abstract concepts, but again only when stimuli were blocked by emotional experience to make this dimension more salient to participants.[1] A functional magnetic resonance imaging (fMRI) study conducted by Vigliocco and colleagues [54] furthermore found that only for abstract but not concrete concepts, emotional valence manipulated hemodynamic activity in rostral anterior cingulate cortex (rostral ACC), an area that has been implicated in emotion processing [62,63]. Subsequent studies however have failed to find an interaction between valence and concreteness in rostral ACC [64].

Ponari et al. [17] conducted a reaction time study with children aged 6 to 11, finding that the middle age group in this cohort shows an effect of emotionality for abstract but not concrete concepts, although this effect was only observed for positive valence. In a similar vein, Lund et al. [45] found that the middle out of three groups (5-7 year-olds) showed sensitivity to the emotional content of words, but again, an interaction between valence and concreteness was only obtained for positive valence. Finally, Kim et al. [65] showed that for recognition memory in 7-8

[1] As pointed out Borghi et al. ([22], p. 18), the fact that in these tasks emotion effects only emerge when the emotional dimension is made salient to participant could also be seen as a challenge to the idea of emotional grounding, since it shows that emotionality does *not* matter when no blocking occurs.

year old children, valence interacted with concreteness, but only for negative words. These acquisition studies are generally interpreted as supporting the idea of emotional grounding, although the fact that only partial effects are obtained (i.e., for either positive or negative valence, not both) could equally be seen as a disconfirmation of emotional grounding, given that the hypothesis was originally about *both* positive and negative valence [16].[2]

Evidence against emotional grounding. Yao et al. [7] found a facilitatory effect of emotional valence for concrete but *not* abstract words when controlling for similar types of variables as was done in the original study by Kousta et al. [16]. In addition, they showed that individual differences in alexithymia (the difficulty to identify and describe emotions) did not modulate the interaction between concreteness and valence, as would be expected if emotion understanding was a necessary component for the processing of abstract concepts. Additional evidence against emotional grounding comes from Kanske and Kotz [66], who found an emotion interaction on reaction times only for concrete words. Event-related potentials (ERPs) furthermore revealed that concreteness and valence interacted in the late positive component (LPC), an ERP signature that has been linked to mental imagery. Using recordings of facial muscle activity, Künecke et al. [67] found a valence effect in the *m. corrugator supercilii*, a muscle involved in frowning that is highly correlated with stimulus valence [68]. In opposition to the idea of emotional grounding, this valence effect was only observed in response to concrete but not abstract words. In an ERP study, Palazova et al. [69] found that emotion-related differences in an early posterior negativity (EPN) arose faster for concrete than abstract verbs.

This paper focuses on what Ponari et al. [70] have called the "starting point" for emotional grounding, which is that there is "a general statistical preponderance

[2] Without a convincing account for why sometimes only partial effects are found, the original theory needs to be changed (i.e., the emotional grounding hypothesis would not be predicated on both positive and negative valence anymore), or otherwise the partial effects actually speak against the original theory.

of affective information for abstract words" ([16], p. 25) [17,54]. For this paper, we take the inverted U-shape (both positively and negatively valenced words are more abstract) as the signature of the emotional grounding hypothesis and assess the extent to which this nonlinear pattern generalizes across languages, concepts, and rating scales.

3. Extending emotional grounding beyond English

3.1. Rationale

It is problematic to take English and other European majority languages as a vantage point when claims are actually predicated upon the conceptual system writ large [71,72]. A cross-linguistic test of the idea of emotional grounding is especially important because it is known that cultures differ with respect to emotion concepts [73,74], and corpus analyses show that the meanings of emotion-related concepts are not well aligned across cultures [75].

To assess the cross-linguistic generalizability of the emotional grounding hypothesis, the inverted U-shape relationship between emotional valence and concreteness reported for English by Vigliocco et al. [54] and Ponari et al. [17] will be assessed for the languages presented in Table 1. Although this dataset includes only three non-Indo-European languages (Indonesian, Mandarin Chinese, Cantonese), considering seven different languages is a considerable improvement vis-à-vis the existing literature, which almost exclusively focuses on English. Some studies have investigated the idea of an inverted U-shaped relationship between emotional valence and concreteness for particular languages, but this is the first study on this topic to synthesize results from across rating studies.

Language	***N* words**	**Source**
Cantonese	290	Yee [76]
Mandarin Chinese	1,100	Yao et al. [77]
Croatian	3,022	Ćoso et al. [78]

Dutch	valence: 4,300; concreteness: 30,000	Moors et al. [79]; Brysabert et al. [80]
French	valence: 1,000; concreteness: 1,660	Monnier & Syssau [81]; Bonin et al. [82]
German	1,000	Kanske & Kotz [83]
Indonesian	1,490	Sianipar et al. [84]
Italian	1,120	Montefinese et al. [85]
Polish	4,900	Imbir [86]
Spanish	1,400	Guasch et al. [87]

Table 1. Languages and rating datasets considered in this study

3.2. Methods

The `brms` package version 2.16.2 was used to fit Bayesian regression models [88,89], and the `tidyverse` package version 1.3.1 [90] was used for data processing. All analyses were conducted with R version 4.1.1 [91]. The `tidybayes` package version 3.0.1 [92] was used for plotting posterior distributions. The `patchwork` package version 1.1.1 [93] was used for creating multi-plot layouts. Throughout this paper, we use Bayesian regressions for the main analyses, but also report frequentist models when claims specifically relate to existing analyses for which *p*-values were the inference criterion. Data and analysis code is available under the following Open Science Framework repository: https://osf.io/8p2an/

In line with standard practice in the analysis of rating data, the individual word (averaged across ratings per participant) is the unit of analysis in the statistical models considered here. To calculate R^2 for the different languages, individual (frequentist) regression models were fitted, for which per-word concreteness ratings were regressed onto two predictors: valence, and valence-squared. To generalize across languages, the main Bayesian model considered all rating data, *z*-scored within languages to standardize the different scales, with a random effect for language. The model included by-language varying random slopes for both the linear and the quadratic valence effects, which is needed to support the claim that these effects generalize across languages [94,95]. The model we considered here does not consider a random effect for item because 1) datasets have different concepts, many of which do not overlap between the rating studies, and because 2) using a random effect for "item" amounts to assuming translational equivalence between the concepts across languages. As most concepts have non-overlapping glosses, the matching of concepts across languages is hard and laden with assumptions.

For more conservative inferences and to avoid overfitting [96,97], a weakly informative prior was set on all regression slopes for the main Bayesian model: $Normal(\mu = 0, \sigma = 0.2)$. Other than this, we followed the default priors automatically assigned by the `brms` package. Finally, when considering rating data, it is important to consider that for some concepts, participants disagree more in their ratings, as reflected in the corresponding standard deviations [59]. This was dealt with by adding standard deviations as regression weights to the linear mixed effects model, a method that has been shown to improve model fits for studies analyzing concreteness ratings [57]. These regression weights penalize high-SD words. Doing this improved the fit (from $R^2 = 0.09$ to $R^2 = 0.13$).

Markov Chain Monte Carlo estimation was executed with 4 chains and for 8,500 iterations (6,500 warmup samples discarded). There were no divergent transitions and all chains mixed well (Rhat = 1.0 for all parameters).

3.3. Results

Figure 1 shows scatter plots of concreteness ratings across the emotional valence predictor, with superimposed linear regression fits (maximum likelihood point estimate) of the corresponding polynomial regression models. Table 2 summarizes the quadratic coefficients and partial R^2 of the quadratic effect. As can be seen from both the figure and the table, there are negative quadratic effects for almost all languages except for Mandarin Chinese and Cantonese. It is noteworthy that we failed to reproduce the quadratic effect reported in Yao et al. [77], i.e., the same data does not yield the inverted U-shaped reported in the original study.

Language	Quadratic effect	Partial R^2 of quadratic effect
Cantonese	-0.04, *SE* = 0.07	0.0008
Mandarin Chinese	+0.0095, *SE* = 0.01	0.0007
Croatian	-0.06, *SE* = 0.007	0.02
Dutch	-0.30, *SE* = 0.03	0.10
French	-0.17, *SE* = 0.02	0.13
German	-1.06, *SE* = 0.56	0.27
Indonesian	-0.22, *SE* = 0.02	0.09
Italian	-0.14, *SE* = 0.01	0.03
Polish	-0.33, *SE* = 0.01	0.13
Spanish	-0.09, *SE* = 0.01	0.04

Table 2. Quadratic coefficients and standard errors extracted from the corresponding polynomial regressions; these models include regression weights penalizing high-SD words (cf. [57])

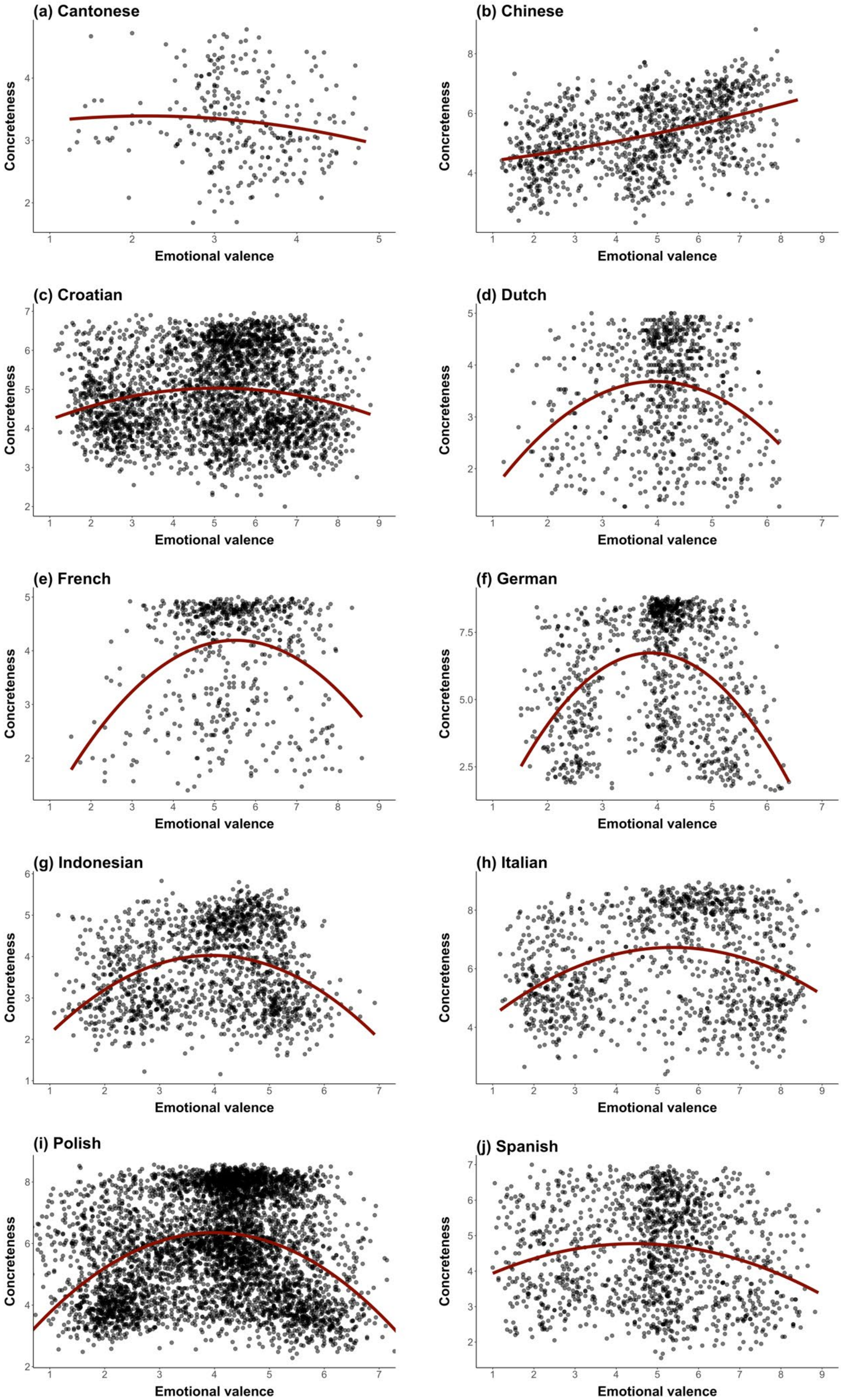
(a) Cantonese
(b) Chinese
(c) Croatian
(d) Dutch
(e) French
(f) German
(g) Indonesian
(h) Italian
(i) Polish
(j) Spanish
Concreteness
Emotional valence

Figure 1. Scatter plots of all words in concreteness x emotional valence space for the respective languages, with superimposed linear regression fits

The next analysis considers all languages together in a single linear mixed effects model with random effects (see Methods section 3.2) allowing us to generalize across this set of languages. Figure 2 shows the posterior distributions of the linear and quadratic coefficient from this conjoined analysis. As can be seen, the posterior distribution of the quadratic coefficient (posterior mean: -0.22, SE = 0.06) is far away from zero, with a 95% credible interval excluding zero, [-0.33, -0.10]. The posterior probability of this effect being of the same sign is very high, $p(\beta < 0) =$ 0.99. In contrast to the quadratic effect, the posterior distribution of the linear coefficient (posterior mean: +0.06, SE = 0.07) firmly included zero; 95% credible interval: [-0.07, +0.19]. The posterior probability of this effect being of the same sign was $p(\beta > 0) = 0.84$. The Bayesian mixed model described R^2 = 0.13 of the variance in concreteness ratings.

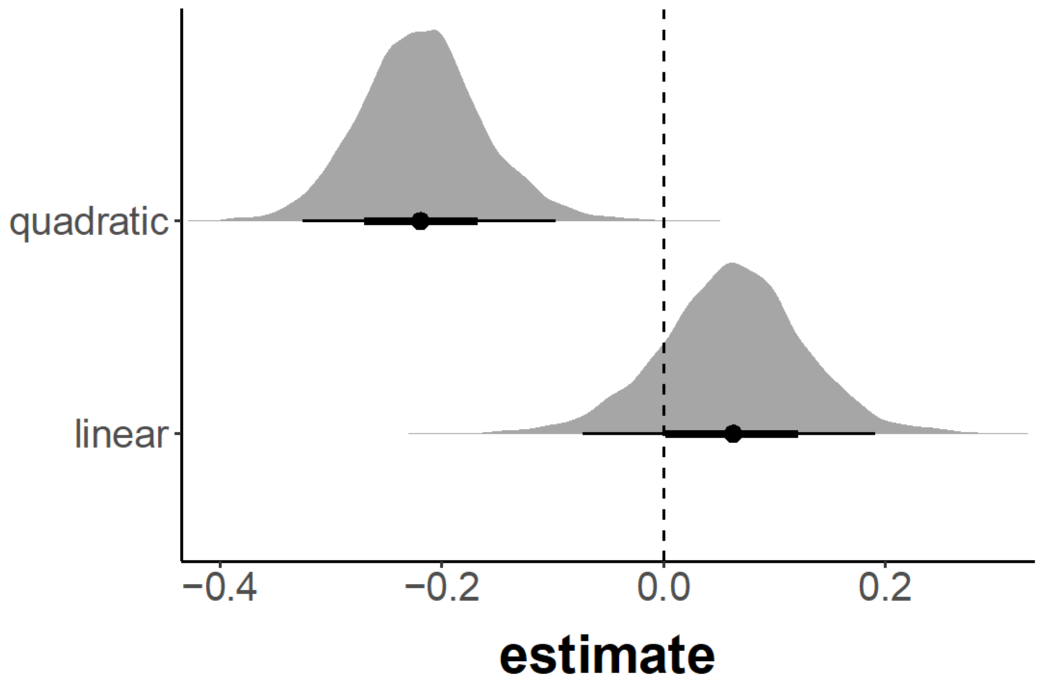

Figure 2. Posterior distributions for the linear and quadratic coefficient of the Bayesian linear mixed effects model; the dashed vertical line shows zero

Together, these results provide support for the idea that the emotional grounding hypothesis characterizes this set of languages: it appears as if more strongly emotionally valenced concepts — both positive and negative — are relatively more abstract in all of the languages except for Mandarin Chinese and Cantonese. By no means, however, does this allow concluding that the emotional grounding hypothesis is a cross-linguistic universal. Research in the tradition of linguistic typology generally requires many more languages from a much more diverse sample, including more languages from other language families. Given that rating data is only available for the small set of languages discussed here — all of are associated with large industrialized societies — we simply do not know whether the emotional grounding hypothesis is even more general and in particular, whether it would also characterize data from minority languages or more remote cultures. Nevertheless, the fact that the inverted U-shape emerges as a reliable effect when data is aggregated across languages suggests limited cross-linguistic generalizability.

4. Does the original data support emotional grounding?

4.1. Rationale and approach

Statements about emotional grounding are generally predicated upon *all* abstract concepts, e.g., "emotion provides grounding for abstract concepts" ([17], p. 2). An issue with the analyses conducted in the last section, as well as with the analyses conducted by Vigliocco et al. [54] and Ponari et al. [17], is that internal variation in abstract concepts is not captured by regression models that only incorporate a continuous rating scale without considering the potential presence of distinct subgroups of abstract concepts. Researchers studying abstract concepts have recently begun to emphasize more strongly that abstract concepts are characterized by heterogeneity [33,49,56]. A quick look at Figure 1 shows that for some of the languages considered here, clusters are readily visible to the naked eye. This is problematic for interpreting the quadratic effect in a continuous manner, as small subgroups of words can create quadratic patterns in the average. To demonstrate that this is actually a concern for the emotional grounding hypothesis, it is useful to briefly consider simulated data. In Figure 3a, random data was initialized by drawing 100 concreteness values that are uniformly distributed across the emotional valence scale, $Uniform(a = 1, b = 9)$, with concreteness values drawn from $Normal(\mu = 3.03, \sigma = 1.04)$ (means and standard deviations taken from the most extensive concreteness rating study conducted by Brysbaert et al. [98]). A regression model entering emotional valence as linear and quadratic predictor reveals no 'significant' quadratic effect in this randomly generated data (coefficient of quadratic effect: -0.003, *SE* = 0.02, *p* = 0.85).

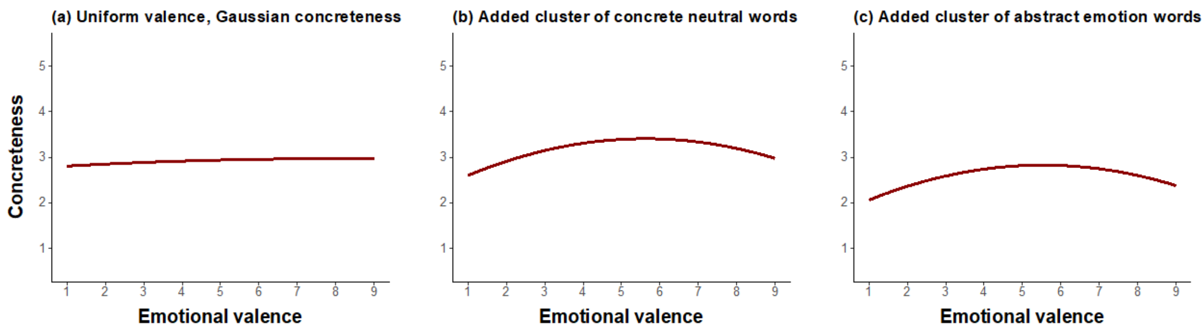

Figure 3. a) A random cluster of words showing no quadratic effect; b) a small group of concrete neutral words has been added to the data from a); c) a small group of abstract emotion words has been added to the data from a)

Small variations to this basic setup can create apparent quadratic effects. For example, if we add a small cluster of only 30 concrete neutral words to the existing 100 data points, this will exert a pull on the quadratic coefficient, creating the inverse U-shape pattern seen in Figure 3b. Although the cluster is barely visible in the plot, the quadratic relationship would be judged to be 'significant' (*b* = -0.04, *SE* = 0.16, *p* =

0.02). To some extent, this average quadratic trend is real and indeed an accurate reflection of the relationship between emotional valence and concreteness for this data. However, to some extent the quadratic trend is also spurious for this simulated data, given that we know that it does not characterize the whole concreteness rating scale but is instead driven by only a small group of words. The majority of words (those that are also shown in the original Figure 3a) do not actually follow the quadratic trend that is suggested by the regression model. Clearly, the general claim that "abstract concepts are more emotional" does not characterize the simulated data shown in Figure 3b very well.

An additional way of creating quadratic patterns to the data shown in Figure 3a is to add 20 negative words and 20 positive words with high abstractness, resulting in Figure 3b. Again, a quadratic pattern in the average is entirely created by these two small sets of words. Just as in Figure 3b, these clusters are barely visible to the naked eye, but they are enough to create an average quadratic effect in the corresponding regression model that would be judged to be 'significant' (b = -0.04, SE = 0.02, p = 0.04).

4.2. Applying cluster analysis to the original English data

The simulated data represents a proof-of-concept demonstration of the idea that clusters are a potential problem for the emotional grounding hypothesis. Whether there actually is statistical support for clusters in the data is a separate question. To assess the impact of clusters on the emotional grounding hypothesis, Figure 4 and 5 reproduce the analyses by Vigliocco et al. [54] (using the MRC concreteness ratings [99] and ANEW emotional valence ratings [100]) and Ponari et al. [17] (using the Brysbaert et al. concreteness ratings [98] and the Warriner et al. [101] emotional valence ratings). For both analyses, there were 'significant' quadratic effects (Vigliocco et al.: b = -13.12, SE = 1.14, p < 0.0001; Ponari et al.: b = -0.07, SE = 0.004, p < 0.0001).

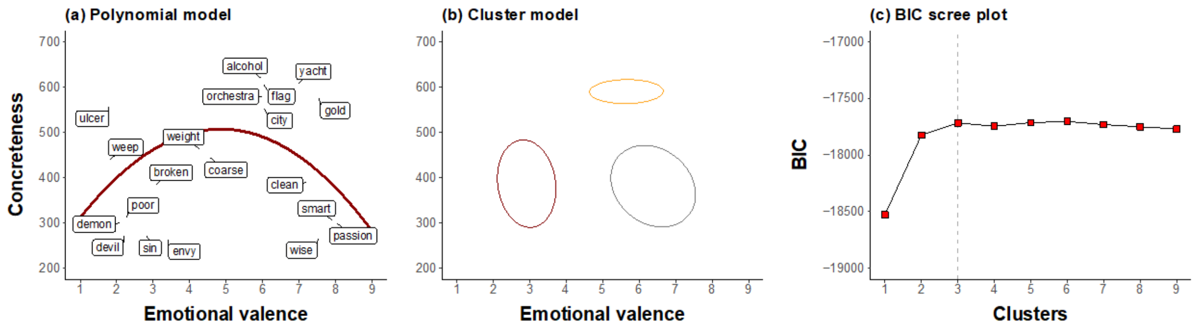

Figure 4. a) Polynomial model for the MRC concreteness ratings [99] and the ANEW emotional valence ratings [100], replicating Vigliocco et al. [54]; b) shows the same data with super-imposed Gaussian mixture models for a three-cluster solution, as determined by c) a scree plot of Bayesian Information Criterion against clusters (the mclust package uses a reversed BIC scale)

The `mclust` package version 5.4.7 [102] was used to perform cluster analyses using Gaussian mixture models over the two-dimensional space spanned by concreteness and emotional valence. Scree plots of Bayesian Information Criterion (BIC) values (`mclust` uses a reversed BIC scale) were used to determine cluster solutions for each dataset by looking at the maximum number of clusters before there is a bent in the scree plot.[3] For the data from Vigliocco et al. [54] a three-cluster emerged as optimal; for the data from Ponari et al. [17] a four-cluster solution emerged as optimal.

[3] All models were fit with `modelNames = 'VVV'`, allowing for the most flexible cluster shapes (ellipsoid that can have any orientation).

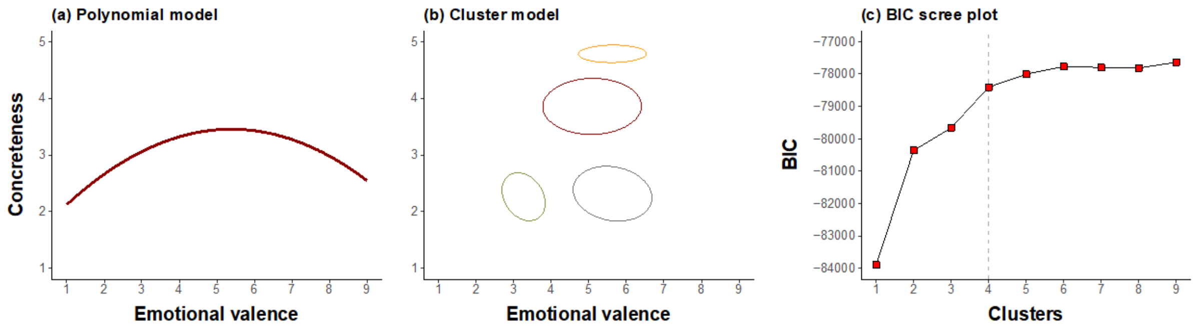

Figure 5. a) Polynomial model for the Brysbaert et al. [98] concreteness ratings and the Warriner et al. [101] emotional valence ratings, replicating Ponari et al. [17]; b) shows the same data with super-imposed Gaussian mixture models for a four-cluster solution, as determined by c) a scree plot of Bayesian Information Criterion against clusters (the mclust package uses a reversed BIC scale)

When concreteness ratings are regressed onto a three-cluster solution of the Gausian mixture model for the Vigliocco et al. [54] data, this predictor explains considerably more variance (adjusted $R^2 = 0.69$). If clusters and linear and quadratic effects are simultaneously entered into the same regression model, the quadratic effect ceases to be 'significant' ($b = 0.90$, $SE = 0.85$, $p = 0.29$). For the data from Ponari et al. [17], the variance explained by a four-cluster solution vastly supersedes the variance explained by the polynomial model (adjusted $R^2 = 0.69$), and when both polynomials and clusters are simultaneously entered into the same regression, there is a quadratic effect in the *opposite* direction ($b = 0.005$, $SE = 0.002$, $p = 0.007$).[4]

The same Gaussian mixture model approach was applied to all cross-linguistic datasets. For every single language except for Cantonese, a three-cluster solution emerged as the providing the best balance between model fit and

[4] It should be pointed out that regression models using clusters that are derived from the data in a bottom-up fashion will naturally capture more variation given that the expectation maximization algorithm used by `mclust` is designed to find the best-fitting clusters. In response to the concern that the higher R^2 of a model with clusters is statistically inevitable, it should be highlighted that 1) the very existence of clusters (regardless of their subsequent use in regression models), and 2) the fact that there are no traces of inverted U-shapes within clusters both independently speak to limitations of the emotional grounding hypothesis.

complexity. For Cantonese, the Gaussian mixture model indicates that a one-cluster solution (i.e., no cluster) is the best fit. The clusters did not clearly align across languages, i.e., it was not the case that there was always a cluster of abstract emotion words, for example. Just as was the case with the English data, when concreteness was regressed onto the cluster predictor, this described between 36% and 82% more variance than the corresponding quadratic model, with the exception of Cantonese (no excess variance given that there were no clusters for this language). When both clusters and polynomials were simultaneously entered into the same regression model, there were 'significant' negative quadratic effects for Dutch, Indonesian, German, French, and Mandarin Chinese, 'significant' *positive* quadratic effects for Croatian and Italian, as well as no 'significant' effects for Spanish, Polish, and Cantonese. See the online open science framework repository (https://osf.io/8p2an/) for detailed results for all individual languages.

4.3. Effect size considerations

The cross-linguistic data shown in Figure 1 as well as the original English data shown in Figures 4 and 5 also suggests another reason for concern. The intense scatter suggests that broad sweep claims such as "emotion provides grounding for abstract concepts" ([17], p. 2) need to be qualified. It is clear from the plots shown here that the majority of abstract concepts are *not* captured by the quadratic trend: across the whole range of the emotional valence spectrum, we find words of all concreteness levels, a pattern that is particularly striking for the Ponari et al. [17] data shown in Figure 5b. Standardized effect size measures paint a similar picture. For the data from Vigliocco et al. [54], there is a quadratic effect that is associated with 10.5% partial variance (as computed by the `rsq` package version 2.2 [103]). For the data by Ponari et al. [17], partial R^2 was even lower, with only 2.8% of the overall variance in concreteness ratings being attributable to the quadratic effect. Together with the visual impression suggested by the scatter plots, these relatively low effect sizes show that emotional grounding clearly fails to account for most of the variation in concreteness ratings. The partial R^2 values for the cross-linguistic data shown in Table 2 suggest a similarly humbling picture, with 6 out of 10 regression models describing less 10% of the variance.

5. Triangulating concreteness using different rating scales

So far, we have explored the generalizability of the emotional grounding hypothesis with respect to different languages (section 3) and analysis methods (section 4). Another important aspect of assessing the generalizability of this hypothesis is the extent to which it depends on using a particular operationalization of concreteness, specifically concreteness ratings. There are many different ways of operationalizing concreteness [20,21,33,104], and other measures have been shown to outperform concreteness ratings when it comes to accounting for word processing times [58].

Here, we are going to focus on the idea that concreteness captures whether a concept is accessible to the senses. If we take accessibility to the senses as a primary component of the concrete/abstract distinction, scales specifically focused on sensory experience [105] or perceptual strength [47,58,106,107] are viable alternative operationalizations. To see whether the inverted U-shape holds for such other measures, we combined emotional valence ratings from Warriner et al. [101] with sensory experience ratings from Juhasz and Yap [105] and the Lancaster sensorimotor norms [106].

Figure 6a shows that for sensory experience ratings from Juhasz and Yap [105], there was a *positive* rather than negative quadratic coefficient ($b = 0.12$, $SE = 0.007$, $p < 0.0001$), i.e., the U-shaped curve is not inverted. The same was the case for the maximum sensorimotor strength association from the Lancaster modality norms [106] ($b = 0.03$, $SE = 0.003$, $p < 0.0001$), as well as for the sum of all sensorimotor ratings from the same data ($b = 0.49$, $SE = 0.02$, $p < 0.0001$).[5] And even though we used the same emotional valence data, all R^2 values for these quadratic trends exceed

[5] For the Lancaster data, the maximum and sum are based on both the perceptual and the motor rating scales. Similar results are obtained if only the perceptual rating scales are used.

the effect size reported in Ponari et al. [17], ranging from sensorimotor max (3%) over sensory experience ratings (6%) to sensorimotor sensorimotor sum (8%). These new results are diametrically opposed to the idea of emotional grounding. It appears that once we move away from concreteness ratings, it is concepts that are *more* concrete that are also more emotionally valenced.

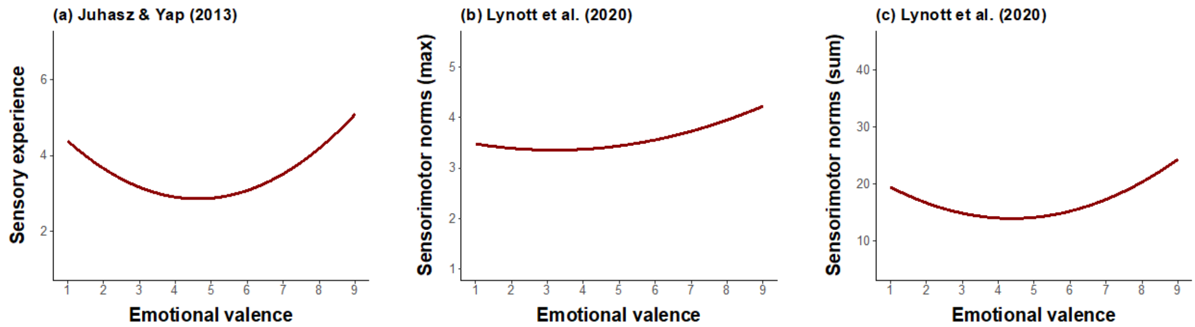

Figure 6. Emotional valence from Warriner et al. [101] plotted against sensory experience ratings [105] and the maximum and sum of the Lancaster sensorimotor norms [106], showing a non-inverted U-shaped pattern going against the emotional grounding hypothesis

6. Discussion

6.1. General discussion

The results presented here generalize the idea of emotional grounding to a larger set of languages, but they also suggest strong limitations. The following sums up all data-driven results against the emotional grounding of abstract concepts:

- Scatter plots suggest that the quadratic effect does not characterize *most* abstract concepts
- This is also suggested by the relatively weak effect sizes (R^2 = 0.028 in the biggest English dataset)

- A proof-of-concept demonstration with simulated data shows that clusters can create apparent quadratic effects in the kinds of polynomial regression models originally used to support emotional grounding
- Cluster analyses performed on the original data and the new cross-linguistic data reveal clear subgroups of words that a regression model by necessity averages over
- Regression models with clusters entered as categorical predictor variables outperform polynomial models and at least for English (but not for all of the other languages), quadratic effects would not be judged as 'significant' any more once clusters are entered into the model
- Finally, triangulating concreteness ratings with other ways of operationalizing the concept of sensory accessibility (sensory experience ratings, sensorimotor norms) reveals results that have the opposite sign from what is expected based on the emotional grounding hypothesis

On top of the limitations suggested by the present data, it is important to point out that papers on the emotional grounding hypothesis rarely ever mention that there are behavioral effects with opposing sign in the literature [7,66,67,69], as was reviewed in Section 2. In addition to the present set of results, these contravening behavioral results suggests that general statements such as "emotion provides grounding for abstract concepts" ([17], p. 2) need to be clarified, or at least defended vis-à-vis the existing evidence against emotional grounding, including the new evidence presented here.

6.2. Emotional grounding versus language?

As mentioned in the opening section, the most widely accepted proposal in this field of research is that abstract concepts are relatively more language-based. Ponari et al. [17] construe a linguistic basis of abstract concepts as being in opposition to an affective/emotional basis, as reflected in statements such as "emotion, *rather than* *language*, may provide a bootstrapping mechanism for the development of abstract words and concepts" (p. 2; italics our own). This deviates from earlier formulations of the emotional grounding hypothesis that clearly saw both affective and linguistic content as relevant (e.g., Kousta et al. [16]). There clearly is no need for linguistic information and affective/emotional information to be in opposition with each other. In fact, in this section we will argue that the idea of emotional grounding may in fact be epiphenomenal, an outgrowth of the linguistic nature of abstract concepts.

Lenci and colleagues [55] provide a direct test of the idea that the distributional semantic structure of language could explain emotional grounding, showing that affective information inferred from language statistics is a strong predictor of concreteness ratings in Italian. They conclude that "the strong affective content of abstract words might itself be a consequence of their linguistic distribution" (p. 19). More indirect evidence for the idea that emotional grounding effects are actually linguistic effects in disguise comes from considering the details of the available reaction time evidence for emotional grounding: It is known that valence effects arise early on in processing, often before imagistic information could plausibly be involved [69,108]. As was discussed above, Kousta et al. [16] is one of the key reaction time studies supporting the idea of emotional grounding. It has been noted that the reaction times in this story are below 600ms, which is very short [108]. This is consistent with the idea that this study actually types into shallow lexical processing, i.e., largely language-based processing. Second, the task used by Kousta et al. [16] is a lexical decision task, and it has generally been shown that emotional valence effects emerge in simple lexical decision tasks [61] rather than deep lexical processing.

Another piece of evidence for a linguistic locus of emotional grounding comes from considering the nature of the Brysbaert et al. [98] concreteness scale more closely. As discussed by Lupyan and Winter [109], this scale clearly defines abstractness as a language-dependent construct, directly instructing participants that the meaning of abstract concepts "depends on language", and that abstract words are easiest to explain "by using other words". From this perspective, the abstract end of the concreteness scale may actually measure language dependence, and the U-shaped curve observed by Ponari et al. [17] may in fact show that *linguistic* concepts — which also happen to be often abstract — are more emotional than non-linguistic ones. Consistent with this view, the present paper demonstrates that scales measuring accessibility to the senses that are less linguistically contaminated, show the opposite effect of what is expected under the emotional grounding hypothesis.

Given the evidence for the pivotal role of language in the representation and processing of abstract concepts reviewed earlier, it seems likely that language is the primary representational format of abstract concepts, and that the weak effect sizes for emotional grounding piggyback on the linguistic nature by which lexical concepts develop emotional connotations (see also [110]). On top of the available behavioral and neuro-imaging evidence against emotional grounding reviewed earlier, the present results suggests that most abstract concepts are not grounded in emotion.

Acknowledgments

I thank Anna Borghi, members of the BALLAB and ESLP conferences, Louise Connell, Melvin Yap, and Barbara Juhasz for useful comments and suggestions. Bodo Winter was supported by the UKRI Future Leaders Fellowship MR/T040505/1.

References

1. Binder JR, Westbury CF, McKiernan KA, Possing ET, Medler DA. 2005 Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience* **17**, 905–917. (doi:10.1162/0898929054021102)

2. James CT. 1975 The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance* **1**, 130. (doi:10.1037/0096-1523.1.2.130)

3. Kroll JF, Merves JS. 1986 Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **12**, 92.

4. Pauligk S, Kotz SA, Kanske P. 2019 Differential impact of emotion on semantic processing of abstract and concrete words: ERP and fMRI evidence. *Sci Rep* **9**, 14439. (doi:10.1038/s41598-019-50755-3)

5. Rubin DC. 1980 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior* **19**, 736–755. (doi:10.1016/S0022-5371(80)90415-6)

6. Whaley CP. 1978 Word—nonword classification time. *Journal of Verbal Learning and Verbal Behavior* **17**, 143–154. (doi:10.1016/S0022-5371(78)90110-X)

7. Yao B, Keitel A, Bruce G, Scott GG, O'Donnell PJ, Sereno SC. 2018 Differential emotional processing in concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **44**, 1064–1074. (doi:10.1037/xlm0000464)

8. Bottini R, Morucci P, D'Urso A, Collignon O, Crepaldi D. 2021 The concreteness advantage in lexical decision does not depend on perceptual simulations. *Journal of Experimental Psychology: General* **47**, 685–704. (doi:10.1037/XGE0001090)

9. Allen R, Hulme C. 2006 Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language* **55**, 64–88. (doi:10.1016/j.jml.2006.02.002)

10. Schwanenflugel PJ, Stowe RW. 1989 Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly* **24**, 114–126. (doi:10.2307/748013)

11. Doest L ter, Semin G. 2005 Retrieval contexts and the concreteness effect: Dissociations in memory for concrete and abstract words. *European Journal of Cognitive Psychology* **17**, 859–881. (doi:10.1080/09541440540000031)

12. Fliessbach K, Weis S, Klaver P, Elger CE, Weber B. 2006 The effect of word concreteness on recognition memory. *NeuroImage* **32**, 1413–1421. (doi:10.1016/j.neuroimage.2006.06.007)

13. Paivio A. 1965 Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior* **4**, 32–38. (doi:10.1016/S0022-5371(65)80064-0)

14. Paivio A, Csapo K. 1969 Concrete image and verbal memory codes. *Journal of Experimental psychology* **80**, 279. (doi:10.1037/h0027273)

15. Bergelson E, Swingley D. 2013 The acquisition of abstract words by young infants. *Cognition* **127**, 391–397. (doi:10.1016/j.cognition.2013.02.011)

16. Kousta S-T, Vigliocco G, Vinson DP, Andrews M, Del Campo E. 2011 The representation of abstract words: why emotion matters. *Journal of Experimental Psychology: General* **140**, 14.

17. Ponari M, Norbury CF, Vigliocco G. 2018 Acquisition of abstract concepts is influenced by emotional valence. *Developmental Science* **21**, e12549. (doi:10.1111/desc.12549)

18. Schwanenflugel PJ. 1991 Why are abstract concepts hard to understand? In *The psychology of word meanings* (ed PJ Schwanenflugel), pp. 235–262. Hillsdale, NJ: Lawrence Erlbaum.

19. Löhr G. 2021 What are abstract concepts? On lexical ambiguity and concreteness ratings. *Review of Philosophy and Psychology* , 1–18. (doi:10.1007/s13164-021-00542-9)

20. Dunn J. 2015 Modeling abstractness and metaphoricity. *Metaphor and Symbol* **30**, 259–289.

21. Barsalou LW, Dutriaux L, Scheepers C. 2018 Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170144. (doi:10.1098/rstb.2017.0144)

22. Borghi AM, Binkofski F, Castelfranchi C, Cimatti F, Scorolli C, Tummolini L. 2017 The challenge of abstract concepts. *Psychological Bulletin* **143**, 263. (doi:10.1037/bul0000089)

23. Mahon BZ, Caramazza A. 2008 A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris* **102**, 59–70. (doi:10.1016/j.jphysparis.2008.03.004)

24. Barsalou LW. 1999 Perceptual symbol systems. *Behavioral and Brain Sciences* **22**, 577–660.

25. Barsalou LW. 2008 Grounded cognition. *Annu. Rev. Psychol.* **59**, 617–645.

26. Glenberg AM, Kaschak MP. 2002 Grounding language in action. *Psychonomic Bulletin & Review* **9**, 558–565. (doi:10.3758/BF03196313)

27. Stanfield RA, Zwaan RA. 2001 The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science* **12**, 153–156.

28. Zwaan RA, Stanfield RA, Yaxley RH. 2002 Language comprehenders mentally represent the shapes of objects. *Psychological Science* **13**, 168–171.

29. Pulvermüller F. 2005 Brain mechanisms linking language and action. *Nature Reviews Neuroscience* **6**, 576–582.

30. Pulvermüller F. 2013 How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences* **17**, 458–470. (doi:10.1016/j.tics.2013.06.004)

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

32. Crutch S, Troche J, Reilly J, Ridgway G. 2013 Abstract conceptual feature ratings: the role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience* **7**, 186. (doi:10.3389/fnhum.2013.00186)

33. Borghi AM, Barca L, Binkofski F, Castelfranchi C, Pezzulo G, Tummolini L. 2019 Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews* **29**, 120–153. (doi:10.1016/j.plrev.2018.12.001)

34. Meteyard L, Cuadrado SR, Bahrami B, Vigliocco G. 2012 Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* **48**, 788–804. (doi:10.1016/j.cortex.2010.11.002)

35. Vigliocco G, Meteyard L, Andrews M, Kousta S. 2009 Toward a theory of semantic representation. *Language and Cognition* **1**, 219–247.

36. Louwerse MM. 2018 Knowing the meaning of a word by the linguistic and perceptual company it keeps. *Topics in Cognitive Science* **10**, 573–589. (doi:10.1111/tops.12349)

37. Begg I, Paivio A. 1969 Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior* **8**, 821–827. (doi:10.1016/S0022-5371(69)80049-6)

38. Paivio A. 1971 *Imagery and verbal processes*. New York, NY: Holt, Rinehart and Winston.

39. Paivio A. 1991 Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie* **45**, 255–287. (doi:10.1037/h0084295)

40. Paivio A, Walsh M, Bons T. 1994 Concreteness effects on memory: When and why? *Journal of Experimental Psychology: Learning, Memory, and Cognition* **20**, 1196.

41. Paivio A. 1986 *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

42. Binder JR, Desai RH, Graves WW, Conant LL. 2009 Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* **19**, 2767–2796. (doi:10.1093/cercor/bhp055)

43. Binney RJ, Hoffman P, Lambon Ralph MA. 2016 Mapping the multiple graded contributions of the anterior temporal lobe representational hub to abstract and social concepts: evidence from distortion-corrected fMRI. *Cerebral Cortex* **26**, 4227–4241. (doi:10.1093/cercor/bhw260)

44. Sabsevitz DS, Medler DA, Seidenberg M, Binder JR. 2005 Modulation of the semantic system by word imageability. *Neuroimage* **27**, 188–200. (doi:10.1016/j.neuroimage.2005.04.012)

45. Lund TC, Sidhu DM, Pexman PM. 2019 Sensitivity to emotion information in children's lexical processing. *Cognition* **190**, 61–71. (doi:10.1016/j.cognition.2019.04.017)

46. Recchia G, Jones M. 2012 The semantic richness of abstract concepts. *Frontiers in Human Neuroscience* **6**, 315. (doi:10.3389/fnhum.2012.00315)

47. Connell L, Lynott D, Banks B. 2018 Interoception: the forgotten modality in perceptual grounding of abstract and concrete concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170143.

48. Barsalou LW, Wiemer-Hastings K. 2005 Situating abstract concepts. In *Grounding cognition: The role of perception and action in memory, language, and thinking* (eds D Pecher, RA Zwaan), Cambridge, UK: Cambridge University Press.

49. Villani C, Lugli L, Liuzza MT, Borghi AM. 2019 Varieties of abstract concepts and their multiple dimensions. *Language and Cognition* **11**, 403–430. (doi:10.1017/langcog.2019.23)

50. Jamrozik A, McQuire M, Cardillo ER, Chatterjee A. 2016 Metaphor: Bridging embodiment to abstraction. *Psychonomic Bulletin & Review* **23**, 1080–1089. (doi:10.3758/s13423-015-0861-0)

51. Moffat M, Siakaluk PD, Sidhu DM, Pexman PM. 2015 Situated conceptualization and semantic processing: effects of emotional experience and context availability in semantic categorization and naming tasks. *Psychonomic Bulletin & Review* **22**, 408–419. (doi:10.3758/s13423-014-0696-0)

52. Newcombe PI, Campbell C, Siakaluk PD, Pexman PM. 2012 Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience* **6**, 275. (doi:10.3389/fnhum.2012.00275)

53. Siakaluk PD, Knol N, Pexman PM. 2014 Effects of emotional experience for abstract words in the Stroop task. *Cognitive Science* **38**, 1698–1717. (doi:10.1111/cogs.12137)

54. Vigliocco G, Kousta S-T, Della Rosa PA, Vinson DP, Tettamanti M, Devlin JT, Cappa SF. 2014 The neural representation of abstract words: the role of emotion. *Cerebral Cortex* **24**, 1767–1777.

55. Lenci A, Lebani GE, Passaro LC. 2018 The emotions of abstract words: A distributional semantic analysis. *Topics in Cognitive Science* **10**, 550–572. (doi:10.1111/tops.12335)

56. Muraki EJ, Sidhu DM, Pexman PM. 2020 Heterogenous abstract concepts: is “ponder” different from “dissolve”? *Psychological Research* (doi:10.1007/s00426-020-01398-x)

57. Strik Lievers F, Bolognesi M, Winter B. 2021 The linguistic dimensions of concrete and abstract concepts: lexical category, morphological structure, countability, and etymology. *Cognitive Linguistics* (doi:10.1515/cog-2021-0007)

58. Connell L, Lynott D. 2012 Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition* **125**, 452–465.

59. Pollock L. 2018 Statistical and methodological problems with concreteness and other semantic variables: A list memory experiment case study. *Behavior research methods* **50**, 1198–1216.

60. Kousta S-T, Vinson DP, Vigliocco G. 2009 Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition* **112**, 473–481. (doi:10.1016/j.cognition.2009.06.007)

61. Vinson D, Ponari M, Vigliocco G. 2014 How does emotional content affect lexical processing? *Cognition and Emotion* **28**, 737–746. (doi:10.1080/02699931.2013.851068)

62. Etkin A, Egner T, Peraza DM, Kandel ER, Hirsch J. 2006 Resolving emotional conflict: A role for the rostral anterior cingulate cortex in modulating activity in the amygdala. *Neuron* **51**, 871–882. (doi:10.1016/j.neuron.2006.07.029)

63. Kanske P, Kotz SA. 2011 Emotion speeds up conflict resolution: a new role for the ventral anterior cingulate cortex? *Cerebral Cortex* **21**, 911–919. (doi:10.1093/cercor/bhq157)

64. Skipper LM, Olson IR. 2014 Semantic memory: Distinct neural representations for abstractness and valence. *Brain and Language* **130**, 1–10. (doi:10.1016/j.bandl.2014.01.001)

65. Kim JM, Sidhu DM, Pexman PM. 2020 Effects of emotional valence and concreteness on children's recognition memory. *Frontiers in Psychology* **11**. (doi:10.3389/fpsyg.2020.615041)

66. Kanske P, Kotz SA. 2007 Concreteness in emotional words: ERP evidence from a hemifield study. *Brain Research* **1148**, 138–148. (doi:10.1016/j.brainres.2007.02.044)

67. Künecke J, Sommer W, Schacht A, Palazova M. 2015 Embodied simulation of emotional valence: Facial muscle responses to abstract and concrete words. *Psychophysiology* **52**, 1590–1598. (doi:10.1111/psyp.12555)

68. Larsen JT, Norris CJ, Cacioppo JT. 2003 Effects of positive and negative affect on electromyographic activity over zygomaticus major and corrugator supercilii. *Psychophysiology* **40**, 776–785. (doi:10.1111/1469-8986.00078)

69. Palazova M, Sommer W, Schacht A. 2013 Interplay of emotional valence and concreteness in word processing: An event-related potential study with verbs. *Brain and Language* **125**, 264–271. (doi:10.1016/j.bandl.2013.02.008)

70. Ponari M, Norbury CF, Vigliocco G. 2020 The role of emotional valence in learning novel abstract concepts. *Developmental Psychology* **56**, 1855–1865. (doi:0.1037/dev0001091)

71. Wierzbicka A. 2014 *Imprisoned in English: The hazards of English as a default language*. Oxford, UK: Oxford University Press.

72. Henrich J, Heine SJ, Norenzayan A. 2010 Most people are not WEIRD. *Nature* **466**, 29–29. (doi:10.1038/466029a)

73. Jackson JC, Watts J, Henry TR, List J-M, Forkel R, Mucha PJ, Greenhill SJ, Gray RD, Lindquist KA. 2019 Emotion semantics show both cultural variation and universal structure. *Science* **366**, 1517–1522. (doi:10.1126/science.aaw8160)

74. Kövecses Z. 2005 *Metaphor in culture: Universality and variation.* Cambridge, UK: Cambridge University Press.

75. Thompson B, Roberts SG, Lupyan G. 2020 Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour* **4**, 1029–1038. (doi:10.1038/s41562-020-0924-8)

76. Yee LT. 2017 Valence, arousal, familiarity, concreteness, and imageability ratings for 292 two-character Chinese nouns in Cantonese speakers in Hong Kong. *PloS one* **12**, e0174569.

77. Yao Z, Wu J, Zhang Y, Wang Z. 2017 Norms of valence, arousal, concreteness, familiarity, imageability, and context availability for 1,100 Chinese words. *Behav Res* **49**, 1374–1385. (doi:10.3758/s13428-016-0793-2)

78. Ćoso B, Guasch M, Ferré P, Hinojosa JA. 2019 Affective and concreteness norms for 3,022 Croatian words. *Quarterly Journal of Experimental Psychology* **72**, 2302–2312. (doi:10.1177/1747021819834226)

79. Moors A, De Houwer J, Hermans D, Wanmaker S, van Schie K, Van Harmelen A-L, De Schryver M, De Winne J, Brysbaert M. 2013 Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behav Res* **45**, 169–177. (doi:10.3758/s13428-012-0243-8)

80. Brysbaert M, Stevens M, De Deyne S, Voorspoels W, Storms G. 2014 Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta psychologica* **150**, 80–84. (doi:10.1016/j.actpsy.2014.04.010)

81. Monnier C, Syssau A. 2014 Affective norms for french words (FAN). *Behav Res* **46**, 1128–1137. (doi:10.3758/s13428-013-0431-1)

82. Bonin P, Méot A, Bugaiska A. 2018 Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. *Behav Res* **50**, 2366–2387. (doi:10.3758/s13428-018-1014-y)

83. Kanske P, Kotz SA. 2010 Leipzig Affective Norms for German: A reliability study. *Behavior Research Methods* **42**, 987–991. (doi:10.3758/BRM.42.4.987)

84. Sianipar A, van Groenestijn P, Dijkstra T. 2016 Affective meaning, concreteness, and subjective frequency norms for Indonesian words. *Front. Psychol.* **7**. (doi:10.3389/fpsyg.2016.01907)

85. Montefinese M, Ambrosini E, Fairfield B, Mammarella N. 2014 The adaptation of the affective norms for English words (ANEW) for Italian. *Behavior Research Methods* **46**, 887–903. (doi:10.3758/s13428-013-0405-3)

86. Imbir KK. 2016 Affective norms for 4900 Polish words reload (ANPW_R): Assessments for valence, arousal, dominance, origin, significance, concreteness, imageability and, age of acquisition. *Frontiers in Psychology* **7**, 1081. (doi:10.3389/fpsyg.2016.01081)

87. Guasch M, Ferré P, Fraga I. 2016 Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behav Res* **48**, 1358–1369. (doi:10.3758/s13428-015-0684-y)

88. Bürkner P-C. 2017 brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* **80**, 1–28.

89. Bürkner P-C. 2018 Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* **10**, 395–411.

90. Wickham H *et al.* 2019 Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686. (doi:10.21105/joss.01686)

91. R Core Team. 2019 *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

92. Kay M. 2021 tidybayes: Tidy data and geoms for Bayesian models. *R package version 3.0.1* (doi:10.5281/zenodo.1308151)

93. Pedersen TL. 2020 *patchwork: The composer of Plots. R package version 1.1.1*. See https://CRAN.R-project.org/package=patchwork.

94. Winter B, Grice M. 2021 Independence and generalizability in linguistics. *Linguistics* **59**, 1251–1277. (doi:10.1515/ling-2019-0049)

95. Jaeger TF, Graff P, Croft W, Pontillo D. 2011 Mixed effect models for genetic and areal dependencies in linguistic typology. *Linguistic Typology* **15**, 281–319. (doi:10.1515/lity.2011.021)

96. Lemoine NP. 2019 Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses. *Oikos* **128**, 912–928. (doi:10.1111/oik.05985)

97. McElreath R. 2020 *Statistical rethinking: A Bayesian course with examples in R and Stan*. 2nd edn. New York: CRC press.

98. Brysbaert M, Warriner AB, Kuperman V. 2014 Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods* **46**, 904–911.

99. Coltheart M. 1981 The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A* **33**, 497–505. (doi:10.1080/14640748108400805)

100. Bradley MM, Lang PJ. 1999 Affective norms for English words (ANEW): Instruction manual and affective ratings.

101. Warriner AB, Kuperman V, Brysbaert M. 2013 Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior research methods* **45**, 1191–1207.

102. Scrucca L, Fop M, Murphy TB, Raftery AE. 2016 mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J* **8**, 289–317.

103. Zhang D. 2021 rsq: R-Squared and Related Measures. *R package version 2.2*

104. Langland-Hassan P, Faries FR, Gatyas M, Dietz A, Richardson MJ. 2021 Assessing abstract thought and its relation to language with a new nonverbal paradigm: Evidence from aphasia. *Cognition* **211**, 104622. (doi:10.1016/j.cognition.2021.104622)

105. Juhasz BJ, Yap MJ. 2013 Sensory experience ratings for over 5,000 mono-and disyllabic words. *Behavior Research Methods* **45**, 160–168.

106. Lynott D, Connell L, Brysbaert M, Brand J, Carney J. 2019 The Lancaster Sensorimotor Norms: multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods* , 1–21. (doi:10.3758/s13428-019-01316-z)

107. Lynott D, Connell L. 2009 Modality exclusivity norms for 423 object properties. *Behavior Research Methods* **41**, 558–564.

108. Paivio A. 2013 Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General* **142**, 282–287. (doi:10.1037/a0027004)

109. Lupyan G, Winter B. 2018 Language is more abstract than you think, or, why aren't languages more iconic? *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170137.

110. Snefjella B, Kuperman V. 2016 It's all in the delivery: Effects of context valence, arousal, and concreteness on visual word processing. *Cognition* **156**, 135–146.