Journal of Memory and Language 127 (2022) 104369

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

What can size tell us about abstract conceptual processing?

Bo Yao [a,*], Jack E. Taylor [b], Sara C. Sereno [b,*]

[a] Department of Psychology, Fylde College, Lancaster University, UK
[b] School of Psychology and Neuroscience, University of Glasgow, UK

ARTICLE INFO

Keywords:
Size
Emotion
Embodied cognition
Semantic processing
Abstract concepts

ABSTRACT

Embodied cognition theories propose that abstract concepts are grounded in a variety of exogenous and endogenous experiences which may be flexibly activated across contexts and tasks. In three experiments, we explored how semantic size (i.e., the magnitude, dimension or extent of an object or a concept) of abstract (vs concrete) concepts is mentally represented. We show that abstract size is metaphorically associated with the physical size of concrete objects (Experiment 1) and can produce a semantic-font size congruency effect comparable to that demonstrated in concrete words during online lexical processing (Experiment 2). Critically, this size congruency effect is large when a word is judged by its semantic size but significantly smaller when it is judged by its emotionality (Experiment 3), regardless of concreteness. Our results suggest that semantic size of abstract concepts can be grounded in visual size, which is activated adaptively under different task demands. The present findings advocate flexible embodiment of semantic representations, with an emphasis on the role of task effects on conceptual processing.

Introduction

How do we understand concepts? Traditional views of conceptual processing hold that concepts are represented in arbitrarily defined symbols like *car* or *love* (Fodor, 1975; Pylyshyn, 1984). Such symbolic representations are ungrounded – that is, without intrinsic links to their referents (cf. the symbol grounding problem; Harnad, 1990). To address this issue, recent embodied theories of cognition propose that concepts must be represented in our multimodal experiences (Barsalou, 1999, 2008). This hypothesis has been supported by a vast body of empirical evidence linking language with perception and action (e.g., Fischer & Zwaan, 2008; Glenberg & Gallese, 2012; Glenberg & Kaschak, 2002; Pulvermüller, 2005; Speer et al., 2009). Behaviourally, language comprehension of perceptual- and action-related concepts is often modulated, respectively, by concurrent perceptual and motor tasks (Glenberg & Kaschak, 2002; Zwaan et al., 2002; Zwaan & Taylor, 2006). Neuroanatomically, processing sensory- and action-related concepts typically recruits neural substrates that are engaged, respectively, in perception and action (Kiefer et al., 2008; Pulvermüller, 2005, 2013). While most empirical evidence favours an (at least partially) embodied view of conceptual processing, such data have been predominantly gathered in relation to concrete concepts, with some embodiment effects failing to replicate (e.g., Morey et al., 2021). Critically, it remains inconclusive how abstract concepts, such as *love* and *freedom*, are mentally represented under an embodied framework (Borghi et al., 2017).

While some theorists have proposed complementary dis-embodied (e.g., linguistic) representations for abstract concepts (Andrews et al., 2014; Borghi et al., 2018b; Dove, 2011), several proposals have also been put forward to explore the embodiment of abstract concepts (Barsalou & Wiemer-Hastings, 2005; Connell et al., 2018; Kousta et al., 2011; Lakoff, 1987; Lakoff & Johnson, 1999) – either through exogenous sensory and motor experiences, or through endogenous affective, introspective experiences.

Diverse experiential grounding in abstract concepts

The "exogenous grounding" hypothesis proposes that abstract concepts may nonetheless be grounded in sensorimotor experiences of the external world. Here, it is proposed that, despite (or because of) not having physical referents, abstract ideas are often understood and expressed as metaphorical extensions from their concrete cousins (Gentner & Asmuth, 2019; Lakoff & Johnson, 1999). For instance, *love* is often explained in metaphors such as "journey" or "sweet", and may

* Corresponding authors at: Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK (B. Yao). School of Psychology and Neuroscience, 62 Hillhead Street, University of Glasgow, Glasgow G12 8QB, UK (S.C. Sereno).

E-mail addresses: b.yao1@lancaster.ac.uk (B. Yao), sara.sereno@glasgow.ac.uk (S.C. Sereno).

https://doi.org/10.1016/j.jml.2022.104369
Received 16 November 2021; Received in revised form 4 August 2022; Accepted 5 August 2022

actually be grounded in the experience of making a journey or tasting sweetness. This hypothesis is supported by some behavioural evidence. For instance, Boroditsky and Ramscar (2002) demonstrated that the understanding of the abstract concept *time* could be influenced by representations of spatial experience such as making an air journey or waiting in a lunch line. In other words, *time* may simply 'borrow' experiences of physical activity (occurring over time). Similarly, mental representations of abstract information transfer (e.g., "*You are delegating responsibilities to Anna*") may metaphorically originate from experiences of physical movements (Glenberg et al., 2008).

The "endogenous grounding" hypothesis postulates that abstract concepts may instead rely on experiences fundamentally different to those associated with concrete concepts. While concrete concepts are closely linked to experiences of the external world, abstract concepts may rely more on endogenous experiences such as inner speech (Borghi et al., 2018b), emotion (Kousta et al., 2011), introception (Connell et al., 2018) and introspection (Barsalou & Wiemer-Hastings, 2005). For instance, Kousta et al. (2011) showed that abstract concepts were overall more emotionally charged than concrete ones, enjoying a residual processing advantage once differences in imageability and context availability were accounted for. Using fMRI, Vigliocco et al. (2014) further demonstrated higher activation of the rostral anterior cingulate cortex (rACC; an area associated with emotional processing) during lexical processing of abstract relative to concrete words. Barsalou and Wiemer-Hastings (2005) proposed that, in comparison to their concrete cousins, abstract concepts rely more on introspection and on social aspects of situations. They asked participants to generate properties of highly abstract (e.g., *truth*, *freedom*, *invention*), highly concrete (e.g., *sofa*, *car*, *bird*), and intermediate (e.g., *cooking*, *farming*, *carpeting*) concepts that were preceded by short scenarios. For concrete concepts, participants tended to generate physical properties of the concepts and pick out other concrete objects that were associated with the concepts. In contrast, participants focused more on the social, situational, or introspective aspects of the abstract concepts.

Although these grounding approaches are useful in their own right, they may be restricted to *subsets* of abstract concepts under investigation (e.g., temporal concepts, concepts with emotional contents, etc.), and may not be generalisable to *all* abstract concepts. For example, recent challenges to the affective grounding approach argue that not all abstract concepts may be grounded in emotion. In an fMRI study, Skipper and Olson (2014) showed that the increased activation in the rACC in Vigliocco et al.'s (2014) study may have reflected unbalanced emotional valence between abstract and concrete words, as it did not respond to abstractness once valence was matched between abstract and concrete words. Although emotional valence may be stronger in abstract than concrete words overall (Kousta et al., 2011), it may not characterise the concreteness of a concept diagnostically. For abstract words that lack emotional content (e.g., *thought*, *logic*), other experiential grounding (e.g., in the motor system) may play a more dominant role (Dreyer & Pulvermüller, 2018).

One way to address the generalisability of these grounding approaches is to recognise that they are not mutually exclusive. For instance, while *love* can be metaphorically linked to a journey or the taste of sweetness, it can also be represented in feelings of happiness and in social situations of a romantic dinner or a wedding. The recent emergence of multidimensional views of conceptual representations (Binder et al., 2016; Borghi et al., 2018a; Conca et al., 2021; Crutch et al., 2013; Fernandino et al., 2016; Harpaintner et al., 2018; Villani et al., 2019) has permitted alternative accounts of experiential grounding to coexist in abstract concepts. For instance, Crutch et al. (2013) collected "abstract conceptual feature" (ACF) ratings to quantify the relevance of 12 semantic attributes (e.g., sensation, action, thought, emotion, etc.) to a given word. They found that semantic relatedness in this 12-dimensional space (as measured in Euclidean distances between words) reliably predicted an aphasic patient's ability to semantically discriminate two words. In contrast, neither word usage-based semantic metrics (Latent Semantic Analysis (LSA) cosines; Landauer & Dumais, 1997) nor individual semantic attributes were predictive. Troche et al. (2014) reduced these 12 dimensions into three latent dimensions of perceptual salience, affective association, and magnitude. They found that concrete and abstract words could be represented in this single semantic space with overlapping yet distinct topographies. The most comprehensive brain-based multidimensional semantic space to date is provided by Binder et al. (2016). They normed the salience of 65 experiential attributes, including sensory, motor, social, emotional, and cognitive attributes, for over 500 English words. They found that abstract entities could be distinguished from concrete ones on 57 of the 65 attributes, and were particularly salient on attributes related to temporal, causal, social, and emotional experiences. In a follow-up fMRI study, the 65-dimensional semantic model of word meaning successfully predicted neural activation patterns of trained and novel single words, as well as sentences that contained these words (Anderson et al., 2017).

However, these multidimensional approaches have so far focused on stable and invariant aspects of conceptual representations (cf. Machery, 2015), which are captured via semantic ratings on decontextualised words. Although heterogenous experiential groundings can be captured simultaneously in multidimensional semantic spaces, little is known how they might interact with contextual factors such as the environment/situation or task conditions. One recent study observed that multidimensional conceptual representations can shift in response to significant events such as the Covid-19 pandemic (Mazzuca et al., 2022). This resonates with the idea that concepts are dynamic and flexible in nature (Barsalou, 1993; Connell & Lynott, 2014), and may be instantiated by differential activation of semantic representations across contexts (Lupyan & Casasanto, 2015).

Testing flexibility of abstract conceptual representations

Indeed, a key prediction of the embodied perspective emphasises that experience-based representations are inherently flexible and can take many forms across contexts (Lebois et al., 2015; Yee & Thompson-Schill, 2016) and task demands (Wilson & Golonka, 2013). Abstract conceptual representations may be particularly flexible because they can be simultaneously grounded in a more diverse range of loosely associated experiences, events, and situations (Barsalou & Wiemer-Hastings, 2005; Borghi et al., 2019; Kousta et al., 2011; Lakoff, 1987; Schwanenflugel, 1991; Schwanenflugel et al., 1988). The relative contributions of these diverse experiences would depend on the communicative and situational context an abstract concept is embedded in (Zwaan, 2014). For example, *love* could be simultaneously represented in the taste of sweetness and feelings of happiness. When choosing chocolates for Valentine's Day, *love* may be conceptualised more in the taste of sweetness, whose relevance may wane considerably in the context of hugging a baby. Understanding how grounded representations can be flexibly formed across contexts may provide key insights into reconciling alternative theories of abstractness. Such findings may help reframe the discussion on conceptual processing in general, promoting a paradigm shift from invariantism (e.g., Machery, 2015) to contextualism (e.g., Barsalou, 1993; Connell & Lynott, 2014).

To this end, we tested the hypothesis that abstract conceptual representations are more flexible than concrete representations. Specifically, we examined how semantic size, a universal semantic property, may be flexibly represented in abstract (vs concrete) concepts across contexts and tasks. Semantic size is a measure of something's dimensions, magnitude, or extent, and is a latent semantic factor (labelled "magnitude") underlying multidimensional semantic spaces (Troche et al., 2014, 2017). Based on a norming database of 5,500 English words, semantic size is relatively evenly distributed along the concreteness continuum, and can be either big (*mountain*, *infinity*) or small (*caterpillar*, *zero*) (Scott et al., 2019). It is therefore a common semantic dimension that can be examined in both abstract and concrete concepts, while being simultaneously grounded in distinct embodied experiences.

Research to date has shown that semantic size influences conceptual processing, particularly in concrete concepts that can be measured in physical size. Using a modified Stroop task, Rubinsten and Henik (2002) found a size congruency effect between semantic size and physical size. Participants were shown pairs of animal names that differed in semantic and font size (e.g., *ant – LION*[1]) and were explicitly asked to judge which of the two words presented was larger, in either semantic or font size. In both tasks, response times were faster on size congruent (*ant – LION*) than size incongruent stimuli (*ANT – lion*), suggesting an interaction between the activation of semantic size and the perception of physical (i.e., font) size. Similarly, Setti et al. (2009) asked participants to decide whether a pair of prime and target words (e.g., *elephant - giraffe*) belonged to the same category. They found that target words (e.g., *giraffe*) were responded to faster when preceded by a same-size prime (e.g., *elephant*) than a different-size prime (e.g., *hare*), suggesting that object size is elicited by concrete nouns. Using a more implicit lexical decision task, Sereno et al. (2009) observed that words representing big objects (*cathedral*, *dinosaur*, *ocean*) were recognised faster than words representing small objects (*cigarette*, *parasite*, *apple*). Although this effect could not be directly replicated in the US (Kang et al., 2011; Larranaga et al., 2022), it was reliably observed once word familiarity (bigger = less familiar) and gender association (bigger = more masculine) differences are controlled for (Larranaga et al., 2022), suggesting that nuanced socio-cultural differences (e.g., expressed via somewhat different word perceptions and associations in the US) may have masked the size effect in Kang et al.'s (2011) replication. Indeed, a mega-analysis of 4,568 English words showed that lexical decision times from the English Lexicon Project (Balota et al., 2007) were significantly and negatively predicted by semantic size, with familiarity, gender, and eight other lexico-semantic variables controlled for (Scott et al., 2019).[2] The bigger-is-faster advantage highlights a cognitive bias favouring larger objects over small ones and demonstrates that size may be an integral, automatically accessed aspect of concrete word processing.

Interestingly, size is also widely used at the more abstract end of the concreteness spectrum. We often call the event of a wedding the "*big day*" and make statements like "I like *big ideas*" or "That's a *small mistake*". Intuitively, concepts like *trust*, *eternal* and *crisis* can be classified as "big", with concepts like *trace*, *impulse* and *humble* classified as "small", as has been empirically confirmed by the semantic size ratings we collected in the Glasgow Norms (Scott et al., 2019). In a word recognition study, Yao et al. (2013) found that, similar to concrete words, semantically bigger abstract concepts (e.g., *truth*) were processed faster than semantically smaller abstract concepts (e.g., *trace*). This processing advantage was *partially* mediated by the emotions these concepts induced – a *crisis* feels "big" partly because it induces a strong emotion of fear; an *aspect* feels "small" partly because it lacks emotion. However, a significant proportion of the size effects was not accounted for by emotion.

Can other embodied experiences also contribute to abstract size? Because there is a sense of heft or breadth with certain abstract words, which is generally expressed in terms of greater visual size in concrete words, abstract words may 'borrow' similar expressions of embodiment. Although abstract concepts such as a *crisis* cannot be seen or touched in the physical world, their meanings can be developed over repeated analogical comparison or metaphor use (Gentner & Asmuth, 2019). For example, "*With endless queues at the petrol stations, the country is having a fuel supply crisis*" compares the representation of *crisis* analogically to the visual experience of large crowds; "*A midlife crisis is like a great shifting of earth*" aligns the representation of *crisis* metaphorically to the sensory experience of an earthquake. Intuitively, semantically big abstract concepts (e.g., *crisis*) may be more commonly compared with exogenous experiences of larger size (e.g., *an earthquake*), whereas semantically small abstract concepts (e.g., *remark*) may be more commonly compared with exogenous experiences of smaller size (e.g., *a drop in the sea*). Over time, abstract concepts may become more structurally aligned with concrete metaphors (Gentner, 2010), with their semantic size more strongly grounded in exogenous experiences of physical size.

The current study

The current study aimed to understand the extent to which semantic size is flexibly represented in abstract (vs concrete) concepts. Building upon Yao et al.'s (2013) finding that abstract size is partially grounded in endogenous experience of emotion, we tested whether and how abstract size may be flexibly grounded in exogenous experience of physical (visual) size across contexts and tasks. We asked three specific questions: (1) Can abstract size be represented in physical size? (2) Is abstract size *automatically* represented in visual size during lexical processing? (3) Can abstract size be *flexibly* represented in visual size under different task conditions? We addressed these questions in three experiments. In Experiment 1, we tested whether abstract concepts can be metaphorically associated with concrete objects of various physical sizes in a forced association task. We predicted that if abstract size can be grounded in physical size, semantically big abstract concepts would more likely be associated with big physical objects, whereas semantically small abstract concepts would more likely be associated with small physical objects. In Experiment 2, we examined the interaction between semantic size and visual (font) size during lexical processing of abstract and concrete words. We used a more implicit lexical decision task to test if visual size could be *automatically* activated in word recognition of abstract concepts without being forced to make exogenous associations. We predicted that visual size would be activated in lexical processing of both abstract and concrete concepts, producing a semantic-visual size congruency effect (i.e., faster lexical decisions when semantic and visual size match than when they mismatch; cf. Rubinsten & Henik, 2002). In Experiment 3, we investigated the flexibility of the semantic-visual size interaction via explicit judgements of a word's size versus its emotionality. We predicted that if experiential grounding is more flexible in abstract than in concrete concepts, the semantic-visual size congruency effect would be stronger when judging a word's size rather than its emotionality. We also predicted that the size congruency effect would be more task-dependent for abstract concepts, with concrete concepts less affected by task demands.

Data availability

The materials, data and analysis codes are freely available at: https://osf.io/vpkze/.

Experiment 1

Experiment 1 aimed to probe potential metaphorical associations between abstract concepts and concrete objects of various physical sizes in *offline* forced choices. All participants gave written informed consent and the experimental procedure was approved by the University Research Ethics Committee at the University of Manchester.

[1] Here uppercase letters represent a lower-case word presented in larger font size.

[2] In addition to a replication study, Kang et al. (2011) also reported a lack of a size effect on lexical decision times from the English Lexicon Project. The discrepant results may be explained by (1) a relatively smaller sample of words analysed (N=324 vs 4568 in Scott et al., 2019); (2) a dichotomous classification of size (95 "big" vs 238 "small" words), which may not capture the nuanced size differences between individual words; and (3) a conservative hierarchical regression analysis where the size effects were examined on the residuals *after* effects of controlling variables were accounted for (vs a simultaneous regression analysis where the size effects were examined *with* controlling variables; cf. Scott et al., 2019).

Participants

As there were no similar studies to estimate the expected effect sizes, the sample size was determined by a pilot study (N = 72) on a less controlled set of concrete words and by convenience sampling of psychology undergraduate students at the time of research.

Sixty-seven native English speakers (all females, M_{age} = 20.39, SD_{age} = 2.91) from the University of Manchester participated in the experiment for one course credit. All participants had normal or corrected-to-normal vision and had not been diagnosed with any learning/language disorders (e.g., dyslexia). The experiment took approximately 10 min.

Design and materials

The experiment employed a 2 (Abstract Word Size: Big, Small) × 2 (Concrete Triplet Type: Varying, Matched) within-participant design. Target words consisted of 110 abstract words from Yao et al. (2013). Half of the words described relatively big concepts (e.g., *disaster*) while the other half relatively small concepts (e.g., *incident*). Words were matched across conditions for (written) word frequency (log occurrences per million) according to the British National Corpus (BNC; http://www.natcorp.ox.ac.uk/; Davies, 2004), and word length (number of letters) on an item-by-item basis.

Each abstract word pair (e.g., *disaster-incident*) was accompanied by three concrete metaphors. These three metaphors were either size-varying, referring to objects that were big (*stadium*), medium (*costume*), and small (*nucleus*), or size-matched, referring to items with similar semantic size ratings (*tomato*, *candle*, *potato*). They were all nouns (tagged as “substantive” in the BNC) with a mean concreteness rating of at least 5 out of 7 in the Glasgow Norms (Scott et al., 2019). Metaphors within a triplet were matched for word length exactly, and for word frequency within ± 0.2 Zipf from each other (van Heuven et al., 2014). The size categories for size-varying triplets were determined based on the words’ semantic size ratings in the Glasgow Norms (<3 for small; 3.5 ≤ medium ≤ 4.5; >5 for big), whereas size-matched triplets consisted of three random words matched on their semantic size within ± 0.3 rating points from each other. The semantic size of size-matched triplets, although equivalent within a triplet, varied between triplets. To ensure that size-dependent metaphorical associations were not confounded by semantic associations, each concrete word had a cosine similarity of ≤ 0.2 with their respective Big and Small abstract words, according to the GloVe pre-trained word vector trained on 840 billion tokens of web data, varying on 300 dimensions (Pennington et al., 2014). A large pool of candidate triplets was identified by running this pipeline iteratively using LexOPS (Taylor et al., 2020). Words within each triplet were inspected and hand-picked to ensure they were: (1) all animate or all inanimate; (2) not ambiguous (i.e., did not have both big and small meanings); and (3) not used more than three times across the stimulus set. The target words were embedded in sentence fragments that ended with *“is like a(n)…”*. The specifications of the word stimuli are summarised in Fig. 1. All sentence fragments and word stimuli are listed in Table S1.

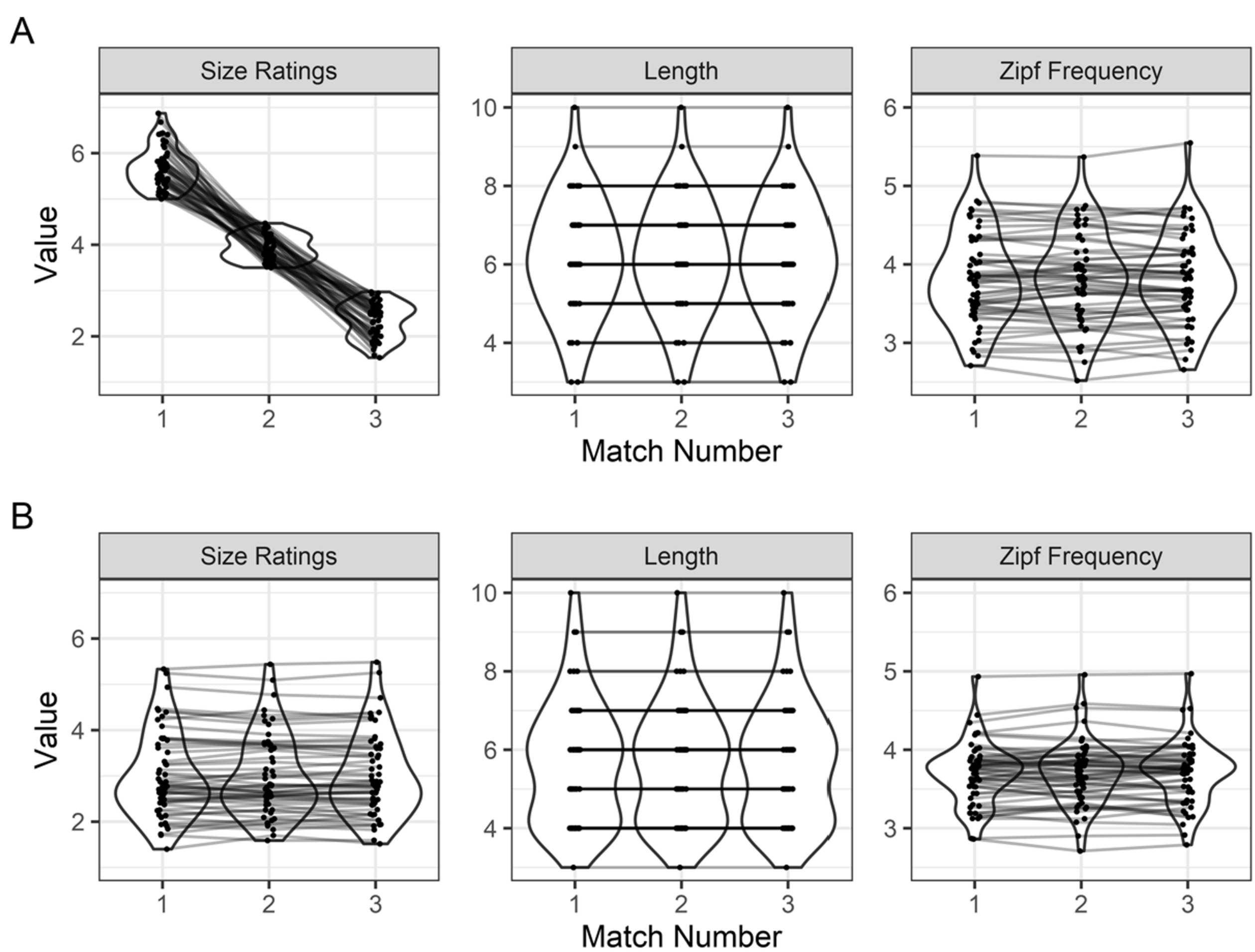

Fig. 1. Semantic size, length, and frequency (Zipf) values for matched triplets of concrete words in the (A) size-varying and (B) size-matched conditions. Points represent individual words, with matched triplets joined by lines. Violin plots show the distributions of variables for each matched item. Match numbers in A, for size-varying triplets, correspond to 1 = “big”, 2 = “medium”, and 3 = “small” concrete words. Match Numbers in B, for size-matched triplets, correspond to 1 = word1, 2 = word2, and 3 = word3 concrete words.

Procedure

The experiment was implemented online using Qualtrics (https://www.qualtrics.com). Participants were presented with a series of sentence fragments with the target word highlighted in bold (e.g., *A **crisis** is like a(n)…*). Each sentence fragment was followed by a triplet of candidate words in a random order (e.g., *stadium*, *costume*, *nucleus* for the size-varying condition; or *tomato*, *candle*, *potato* for the size-matched condition). Participants were instructed to select one of the three words that they thought best completed the sentence metaphorically. They were told to use their first impression in making this decision, and that there were no correct answers.

Participants were presented with one trial at a time. After making their choice, they clicked on the "Next" button to proceed to the next trial. The trials were presented in a random order for each participant. Participants' choices were recorded for analysis.

Results

Participants' word choices were coded as "big", "medium" or "small" based on their semantic size. The percentages of "big", "medium" and "small" responses are summarised in Table 1. Descriptively, Big abstract concepts were more likely to be associated with "big" (49%) than "medium" (26%) or "small" (25%) concrete objects. Small abstract words were more likely to be associated with "small" (46%) than "medium" (25%) or "big" (30%) concrete objects.

We tested whether the observed size-congruent associations were statistically meaningful. We modelled the probabilities for choosing each response in a cumulative link mixed-effect model (CLMM) in R (https://www.r-project.org/) using the 'ordinal' package (https://CRAN.R-project.org/package=ordinal; Christensen, 2015). The CLMM was fitted with the Laplace approximation, with a logit link function and "flexible" thresholds (i.e., threshold locations were not constrained to be, e.g., symmetric or equidistant). The responses were coded in the order of 1 = "big"/word1, 2 = "medium"/word2 and 3 = "small"/word3. The two fixed factors (Abstract Word Size, Concrete Triplet Type) were deviation-coded (0.5 = Big, −0.5 = Small; 0.5 = Varying, −0.5 = Matched). The fixed effect structure included the two main effects and their interaction. The random effect structure included *by-Subject* and *by-Target Word* random intercepts and *by-Subject* random slopes for both fixed factors and their interaction, using the maximal random effects structure justified by the design (Barr et al., 2013). We tested significance via likelihood-ratio model comparisons between the full model and a model lacking each fixed effect of interest, yielding Likelihood-Ratio Chi-squared values (LRχ^2) and associated p values.

The two-way interaction was significant, b = -0.80, SE = 0.13, LR$\chi^2(1) = 30$, $p < .001$. Exploring the interaction, the effect of Abstract Word Size was significant when Concrete Triplet Size was Varying, b = -.97, *95%CI* = [-1.22, -.73], but not when Concrete Triplet Size was Matched, b = -.17, *95%CI* = [-.41, .07]. This suggests that participants' choices of concrete metaphors were significantly biased by Abstract Word Size when there was a size difference within the triplets. When the triplets were matched in size, responses were significantly less affected by Abstract Word Size, and the observed frequencies were close to those expected by chance. As a result of the Abstract Word Size effect in the size-varying triplet condition, the main effect of Abstract Word Size was also significant, b = -.57, SE = 0.10, LR$\chi^2(1) = 26.4$, $p < .001$, with bigger abstract words associated with bigger concrete metaphors and smaller abstract words associated with smaller concrete metaphors. The main effect of Concrete Triplet Type was not significant, b = -0.13, SE = 0.07, LR$\chi^2(1) = 2.77$, $p = 0.096$.

Table 1
Percentage of responses by condition in Experiment 1.

Abstract Word	Size-Varying Triplet "big"	Size-Varying Triplet "medium"	Size-Varying Triplet "small"	Size-Matched Triplet word1	Size-Matched Triplet word2	Size-Matched Triplet word3
Big	49%	26%	25%	34%	34%	32%
Small	30%	25%	46%	31%	33%	36%

Discussion

Experiment 1 explored whether semantic size of abstract concepts may be metaphorically linked to physical size. Participants chose one of three concrete objects that best described an abstract concept metaphorically. Participants were significantly more likely to select concrete words referring to bigger objects to describe semantically big abstract concepts, and concrete words referring to smaller objects to describe semantically small abstract concepts. Their choices were at chance level when the concrete objects were matched in size. Importantly, the size congruency effects could not be explained by differences in word frequency or length between the candidate words, or by differences in semantic associations between the target word and each of the three concrete objects. The results therefore support our prediction that abstract concepts can be metaphorically associated with concrete objects in terms of their size. This suggests that physical size must be activated for both abstract and concrete words to favour size-congruent associations. It remains unclear, however, whether these associations reflect automatic activation of visual size in abstract concepts or were artificially induced by the forced-choice nature of the task. That is, representations of physical size may not necessarily constitute the word's semantic meaning during lexical processing.

Experiment 2

Experiment 2 tested whether representations of physical size are automatically activated during *online* lexical processing of abstract (vs concrete) concepts. In a lexical decision task, we examined the effects of semantic size and visual font size on word recognition of abstract versus concrete words. We chose the more implicit task of lexical decision because it requires relatively superficial semantic activation and does not explicitly require participants to associate abstract concepts with concrete metaphors. If a word's canonical semantic representations comprise embodied visual size, we should observe faster response times when semantic size and font size are congruent (i.e., big word-large font, small word-little font) as opposed to when they are incongruent. As previously observed by Rubinsten and Henik (2002), we predicted such a size congruency effect in concrete words, given that visual size is an integral part of concrete objects. As reasoned in the Introduction and supported by findings from Experiment 1, we also predicted a similar size congruency effect for abstract words, assuming that visual size can indeed be *automatically* activated during lexical processing. Should a size congruency effect not be observed for abstract words, it would instead suggest that visual size is *not* a canonical abstract conceptual representation and that the size-congruent associations demonstrated in Experiment 1 were mostly likely driven by the forced-choice nature of the task.

Participants

As estimated in G*Power 3.1 (Faul et al., 2009), a minimum of 49 participants was required to replicate the size effect reported by Yao et al., (2013) (mean Cohen's f = .38 in a lexical decision task), in a multiple regression model with .05 alpha (2-tailed) and .8 power.

Fifty-six native English speakers from the University of Glasgow participated in the experiment for two course credits. Three participants were excluded because their mean percentage error rate was greater than 15%, suggesting their lexical decisions were unusually poor. Three further participants were excluded because their mean RTs were more than 2 SDs longer than the group average, suggesting they were not responding based on first impressions.

The remaining 50 participants (40 females, 10 males, $M_{age} = 21.9$, $SD_{age} = 4.7$) had normal or corrected-to-normal vision and had no diagnosed learning/language disorders (e.g., dyslexia). All participants gave written informed consent and the experimental procedure was approved by the College of Science and Engineering Ethics Committee at the University of Glasgow. The experiment took about 30 min to complete.

Design and materials

The experiment employed a 2 (Concreteness: Concrete, Abstract) × 2 (Semantic Size: Big, Small) × 2 (Font Size: Large, Little) within-participants design. It comprised a total of 440 stimuli (220 words and 220 nonwords) from Yao et al. (2013). Half of the 220 words had relatively concrete meanings (e.g., *volcano*) while the other half had relatively abstract meanings (e.g., *dynasty*). Within both Concreteness conditions, half of the words described relatively big objects or concepts (e.g., *volcano* or *dynasty*) while the other half described relatively small objects or concepts (e.g., *apricot* or *literal*). Across levels of Concreteness and Semantic Size, words were matched on an item-by-item basis for word frequency and word length. Word frequencies were obtained from the BNC database of 90 million written word tokens (Davies, 2004). All word stimuli are presented in the Supporting Information of Yao et al. (2013). Nonwords comprised pronounceable, orthographically legal pseudowords (e.g., *tabanol*) matched in length to word stimuli.

Two stimulus lists were generated. In each list, half of the words in each condition were presented in large font size (90 pixels) and half were in little font size (30 pixels). Words presented in large font in one list were presented in little font in the other list and vice versa. Each list was assigned to 25 participants.

Procedure

The experiment was run using the OpenSesame software (Mathôt et al., 2012). The visual stimuli were presented on a grey background on a 24″ monitor (60 Hz, 1920 × 1080 resolution; Dell Optiplex 9030 AIO Series) and the viewing distance was approximately 60 cm. The responses were made on a QWERTY keyboard, and reaction times (RTs) were recorded with millisecond accuracy.

Participants were tested individually. They were instructed that they would be presented with a series of letter strings and that they should decide as quickly and as accurately as possible whether each item was a real word or nonword by pressing the corresponding keys – the right CTRL key for words and the left CTRL key for nonwords. Participants were also told that items would be presented in larger or smaller font sizes to determine how font size affected their speed of recognition. Participants were first presented 16 practice trials to familiarise themselves with the task. They were then presented with the 440 experimental items (220 words, 220 nonwords), with three break periods scheduled at regular intervals.

Each trial consisted of the following events. A blank screen was initially presented for 750 ms, followed by the word "NEXT" in the centre of the screen for 200 ms. "NEXT" was displayed in blue, in a Serif font, at a size of 45 pixels, to act as a baseline point of reference between large and little font sizes. Another blank screen was presented for 500 ms and was replaced by a letter string presented centrally on the screen until the participant responded. Letter strings were presented in black Sans Serif font, using a size of either 90 or 30 pixels. Correct responses triggered the next trial. If participants made an error, "INCORRECT!" appeared centrally in 45-pixel red Serif font for 500 ms before the next trial began. Trials were presented in a different random order for each participant.

Results

Before analyses, RT data were pre-processed in line with previous studies (Kang et al., 2011; Sereno et al., 2009; Yao et al., 2013). Trials were excluded if responded to incorrectly (3.9% of word trials). The remaining trials with RTs of less than 250 ms or greater than 1500 ms were also removed. In addition, for each participant in each condition, trials with RTs more than two standard deviations greater than the mean for that participant in that condition were then excluded (with an additional average data loss of 5.4%). These procedures (error and outlier removal) resulted in an average RT data loss of 9.3% per participant. The median RT and %Correct data across Concreteness, Semantic Size, and Font Size conditions are presented in Table 2.

As RTs were positively skewed, we fitted a Gamma family generalised linear mixed-effect model (identity link) of RTs with the *glmer* function in the lme4 package (Bates et al., 2015) in R. We deviation-coded the three fixed factors (Concreteness, Semantic Size, Font Size). We included all main effects and interactions between the three factors in the fixed effect structure and employed the maximal random effect structure with Subject and Word as crossed random factors. The p-values for fixed effects were computed using Satterthwaites's approximation using the lmerTest package (Kuznetsova et al., 2017). The results are reported in Table 3.

There were significant main effects of Concreteness and Semantic Size, as well as a Semantic Size × Font Size interaction. In line with the literature, we found that Concrete words (573 ms)[3] were recognised faster than Abstract words (590 ms). Similarly, responses to semantically Big words (576 ms) were faster than those to semantically Small words (587 ms). Importantly, Font Size differentially affected RTs for semantically Big versus Small words, respectively. Big words were recognised significantly faster when they were presented in Large font (571 ms) than in Little font (580 ms), b = -9.0 ms, 95%CI = [-15.7, –2.3]. The opposite pattern, although not significant, applied to Small words: they were processed numerically slower when they were presented in Large font (589 ms) than in Little font (586 ms), b = 2.8 ms, 95%CI = [-4.0, 9.6]. The interaction is illustrated in Fig. 2.

Discussion

Experiment 2 examined whether visual size was automatically activated during lexical processing of concrete and abstract words. By manipulating the font size of the presented stimuli, we observed a significant size congruency effect, as evidenced by the Semantic Size × Font Size interaction. Regarding the interaction, we found that semantically Big words were recognised significantly faster when they were presented in Large versus Little font; however, semantically Small words were only recognised numerically faster when presented in Little versus Large font. It is possible that size congruency failed to reach significance for Small words because the Little font size (30 pixels) represented only a 33% reduction in size from the reference font (45 pixels). We could not

Table 2
Median RTs (in ms) and %Correct across experimental conditions. Parentheses for RTs contain the interquartile range, while brackets for %Correct contain 95% binomial confidence intervals.

		Semantically Big		Semantically Small	
		Large Font	**Little Font**	**Large Font**	**Little Font**
Concrete	**RT**	529 (118)	533 (126)	535 (132)	536 (126)
	% Correct	98 [97, 98]	98 [97, 98]	96 [95, 97]	96 [94, 97]
Abstract	**RT**	535 (122)	545 (110)	553 (145)	550 (138)
	% Correct	96 [95, 97]	95 [94, 96]	95 [94, 96]	95 [94, 96]

[3] All conditional means in the results section are model-estimated marginal means calculated using R package *emmeans*().

Table 3
Generalised linear mixed-effect model estimates of fixed effects on RTs in Experiment 2.

Fixed effects	*b*	*S.E.*	*t*	*p*
CNC	**−17.71**	**3.64**	**−4.86**	**<.001**
SemSize	**−11.36**	**3.49**	**−3.25**	**.001**
FontSize	−3.11	2.57	−1.21	.226
CNC × SemSize	2.98	6.99	.43	.670
CNC × FontSize	4.93	4.82	1.02	.307
SemSize × FontSize	**−11.75**	**4.61**	**−2.55**	**.011**
CNC × SemSize × FontSize	−2.50	9.30	-.27	.788

Note: CNC = Concreteness; SemSize = Semantic Size; FontSize = Font Size. Significant effects are highlighted in bold.

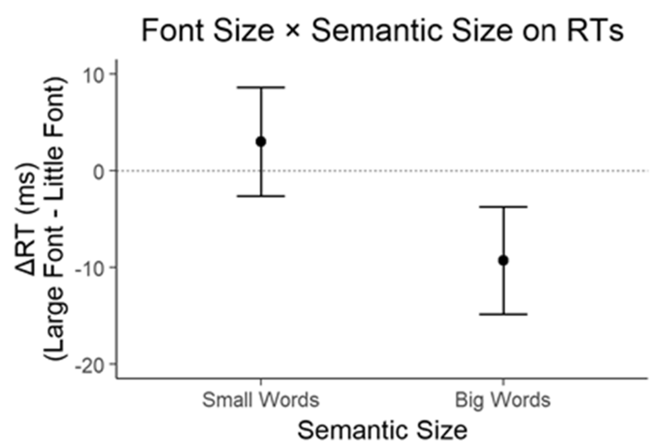

Fig. 2. The estimated effects of font size on RTs by semantic size. Note: The *y*-axis shows the coefficients of the Font Size effects (effectively the RT difference between Large and Little Font conditions). Error bars represent the 95% confidence intervals.

use font sizes smaller than 30 pixels with a 1920 × 1080 resolution as the stimuli would be too small to see clearly from a typical viewing distance. In contrast, the Large font size (90 pixels) was 100% bigger than the reference font. On reflection, it would have been more prudent to select a reference font size of 60 pixels, splitting the difference between Little and Large font sizes.

Critically, the size congruency effect (Semantic Size × Font Size interaction) did not depend on Concreteness. Size congruency in Concrete words is in line with previous research (Rubinsten & Henik, 2002), showing that processing concrete concepts activates sensory experience of the visual size of the denoted object. Size congruency in Abstract words is comparably less straightforward to interpret. One interpretation would be that visual size information must be activated during lexical processing of abstract words, which then interacts with the perceived font size of the text. However, an alternative interpretation might argue that size congruency in abstract words is not necessarily of a visual nature, given that abstract concepts do not have physical referents. Instead, the observed effects may be mediated by congruency between the emotional make-up of abstract words and the emotion elicited by font size itself. It has been reported, for example, that larger font sizes may enhance emotional processing of written words (Bayer et al., 2012). Given that semantically big abstract concepts are more emotionally charged than small abstract concepts (Yao et al., 2013), their processing may be facilitated when font size-elicited emotion corresponds with their semantic emotionality. For instance, *truth* may be considered big in part because it elicits strong emotion (i.e., high arousal or absolute valence), which is more congruent with a higher level of emotion elicited by larger compared to smaller font sizes. As such, the size congruency effect observed in abstract concepts may not be driven solely by congruency of visual size, but also or instead by congruency of semantically- and font size-elicited emotion.

Experiment 3

Experiment 3 probed the nature and flexibility of the observed size congruency effect under different task conditions. Specifically, we manipulated the relevance of size and emotion by asking participants to explicitly judge the size of the presented word (Size Judgement; "Big" vs "Small") and the emotionality (emotional arousal or absolute valence) of the presented word (Emotion Judgement; "Emotional" vs "Neutral"),[4] respectively. We predicted similar size congruency effects in Experiment 3, depending on the nature of the effects and the foci of the tasks. If the size congruency effect in Experiment 2 reflected interactions between semantic and visual (font) size (i.e., a size congruency hypothesis), participants should be more likely and faster to judge semantically big words as "Big" when they are presented in a large font, and judge semantically small words as "Small" when they are in a little font. In contrast, if the size congruency effect in Experiment 2 reflected an interaction between semantic- and font size-elicited emotions (i.e., driven by emotion congruency), participants should be more likely and faster to judge semantically big words as "Emotional" (i.e., high arousal or absolute valence) when they are presented in a large font, and judge semantically small words as "Neutral" when they are presented in a little font. Crucially, we also predicted the task-dependence of size congruency effects would be stronger in abstract over concrete concepts, due to their hypothesised more flexible experiential grounding.

Participants

As estimated in G*Power 3.1 (Faul et al., 2009), a minimum of 39 participants was required to replicate the size congruity effect in Rubinsten and Henik (2002) (Cohen's $d = .93$ in a size judgement task), in a multiple regression model with .05 alpha (2-tailed) and .8 power.

Fifty-two native English speakers from the University of Manchester community participated in the experiment for £3. Ten participants were excluded from the final analysis: one participant responded with the same key throughout the whole experiment; two participants' mean RTs were more than 2 SDs longer than the group average, suggesting that they were not responding based on their first impressions; seven further participants' size judgements on *concrete* objects had a hit rate lower than 70%, suggesting they were not attending to the demands of the task.

The remaining 42 participants (27 female, 15 male, $M_{age} = 24.4$, $SD_{age} = 7.3$) all had normal or corrected-to-normal vision and had no diagnosed learning/language disorders (e.g., dyslexia). All participants gave written informed consent and the experimental procedure was approved by the University Research Ethics Committee at the University of Manchester. The experiment took approximately 25 min.

Design and materials

The design and materials were the same as in Experiment 2 except that the nonwords were excluded. Four stimulus lists were generated to counterbalance for the font size manipulation and task order.

[4] We did not distinguish between positive and negative valence in this study for three reasons. First, abstract size is associated with emotional arousal (Yao et al., 2013). Second, both positively and negatively valenced words demonstrate facilitation relative to neutral words, regardless of polarity or assumed mechanisms (approach vs avoidance) (Kousta et al., 2009). Third, binary choices (Emotional-Neutral) were needed to match the binary choices in the Size Judgement task (Big-Small).

Procedure

The experiment was run on a Dell Optiplex lab computer using the OpenSesame software (Mathôt et al., 2012). Participants received the same 220 words in the Size and Emotion Judgement blocks.

In the Size Judgement block, participants were asked to decide, as quickly as possible and using first impressions, whether each word represented a BIG or SMALL thing or concept. They should press the right CTRL key if they felt the word represented something that was relatively big, or press the left CTRL key if the word represented something relatively small.

In the Emotion Judgement block, they were asked to decide, as quickly as possible and using first impressions, whether each word represented an EMOTIONAL or NEUTRAL thing or concept. They were told that if something was emotionally charged, it can provoke either a *positive* or *negative* response. They were instructed to press the right CTRL key if they felt the word represented something that was strongly emotional, or press the left CTRL key for words that were relatively neutral.

In each block, participants were first presented with 8 practice trials to familiarise themselves with the task, followed by the 220 experimental trials. Each trial consisted of the same events and timings that took place in Experiment 2, only without the post-response feedback. Within each block, trials were presented in a random order, with three breaks scheduled at regular intervals between the trials. Participants' RTs and responses were recorded for analysis.

Results

We expected RTs in this experiment to be slower than those acquired during lexical decision because of the increased complexity of the judgement involved. Nevertheless, as participants had been instructed to respond as quickly as possible using first impressions, similar to Experiment 2, we excluded data based on (slightly longer) cut-offs. Responses with RTs less than 250 ms or greater than 3000 ms were removed (1.7% of trials). In addition, for each participant in each condition, RTs that were more than 2 SDs away from the mean were excluded (with a further data loss of 5.2%). These procedures resulted in an average RT data loss of 6.8% per participant.

Size and emotion judgements

We first examined how Big and Small words were judged by their size and emotion across Concreteness conditions. It is worth noting that there were no right or wrong answers in our judgement tasks. However, because we had predicted Big words to be judged as "Big" or "Emotional" and Small words would be judged as "Small" or "Neutral", we coded participants' responses as "Consistent" when participants categorised words in accordance with the hypothesised correspondence between semantic size and font size (Big-Large, Small-Little) in the Size Judgement task, and between semantic size and emotion (Big-Emotional, Small-Neutral) in the Emotion Judgement task. We quantified the hit rate of each hypothesised correspondence in %Consistent responses. This enabled us to compare the two tasks and the size versus emotional congruency hypotheses on the same measure. To better illustrate the key size congruency effect and simplify the statistical model, we combined the Semantic Size and Font Size factors into a single Size Congruency factor. Size Congruency had two levels: Congruent (when Semantic Size and Font Size were congruent; i.e., Big-Large, Small-Little) and Incongruent (when Semantic Size and Font Size were incongruent; i.e., Big-Little, Small-Large). The mean %Consistent responses and their standard deviations by task and conditions are presented in Table 4.

We fitted a binomial family generalised linear mixed model of binary response accuracy (1 = "consistent", 0 = "inconsistent") with a logit link function. We included a full factorial fixed effect structure with deviation-coded Task, Size Congruency and Concreteness, and a maximal random effect structure with Subject and Word as crossed random factors. The results are reported in Table 5.

Table 4
Mean %consistent responses (95% binomial confidence intervals in brackets) by task, size congruency, and concreteness.

	Size Judgement		Emotion Judgement	
	Congruent	Incongruent	Congruent	Incongruent
Concrete	90 [89, 92]	86 [85, 88]	56 [54, 58]	55 [53, 57]
Abstract	87 [85, 88]	78 [77, 80]	67 [65, 69]	64 [62, 66]

Note: Congruent denotes Semantic-Font Size conditions of Big-Large and Small-Little. Incongruent denotes Semantic-Font Size conditions of Big-Little and Small-Large.

There were significant main effects of Task and Size Congruency: %Consistent responses were significantly higher in the Size Judgement task (93%)[5] than in the Emotion Judgement task (69%), and was higher when semantic and font size were Congruent (86%) than when they were Incongruent (81%). The Size Congruency effect significantly depended on Task, which was significant in the Size Judgement task (Congruent: 95% vs Incongruent: 91%), $b = .61$, 95%CI = [.43, .79], but not in the Emotion Judgement task (Congruent: 70% vs Incongruent: 67%), $b = .16$, 95%CI = [-.01, .32]. This Size Congruency × Task interaction is illustrated in Fig. 3**A**. There was also a significant interaction between Task and Concreteness: %Consistent responses were higher when judging the Size of Concrete words (95%) versus Abstract words (89%), $b = .89$, 95%CI = [.31, 1.46], but was lower when judging the Emotionality of Concrete words (65%) versus Abstract words (72%), $b = -.32$, 95%CI = [-.88, .24]. This interaction is illustrated in Fig. 3**B**.

RTs of size-congruent vs size-incongruent judgements

Next, we examined how Task and Size Congruency influenced RTs. Since the overall %Consistent responses was far from being perfect in the Size Judgement task (86%) and especially in the Emotion Judgement task (60%), we coded the *Effective* Size Congruency by combining Font Size with participants' subjective judgement of the word. That is, when a word in Large font was judged as "Big" or "Emotional" or when a word in Little font was judged as "Small" or "Neutral", we considered it as a size congruent trial, regardless of whether this word was hypothesised to be Big or Small in the first place. This enabled us to tap into the *effective* size congruency effect according to participants' actual responses. The median RTs across Concreteness, Task, and Effective Size Congruency conditions are presented in Table 6.

We fitted a Gamma family generalised linear mixed model of RTs (with an identity link function) using the *glmer* function in R. We included a full factorial fixed effect structure with deviation-coded Task, Effective Size Congruency, and Concreteness in the fixed effect structure and a maximal random effect structure with Subject and Word as crossed

Table 5
Generalised linear mixed-effect model estimates of fixed effects on %Consistent responses in Experiment 3.

Fixed effects	b	S.E.	z	p
Task	**1.78**	**.19**	**9.21**	**<.001**
Size Congruency	**.38**	**.07**	**5.17**	**<.001**
CNC	.27	.24	1.15	.252
Task × Size Congruency	**.45**	**.10**	**4.67**	**<.001**
Task × CNC	**1.21**	**.32**	**3.74**	**<.001**
Size Congruency × CNC	−.11	.10	−1.17	.240
Task × Size Congruency × CNC	−.23	.19	−1.20	.230

Note: CNC = Concreteness. Significant effects are highlighted in bold.

[5] Percentages reported are model-estimated marginal means.

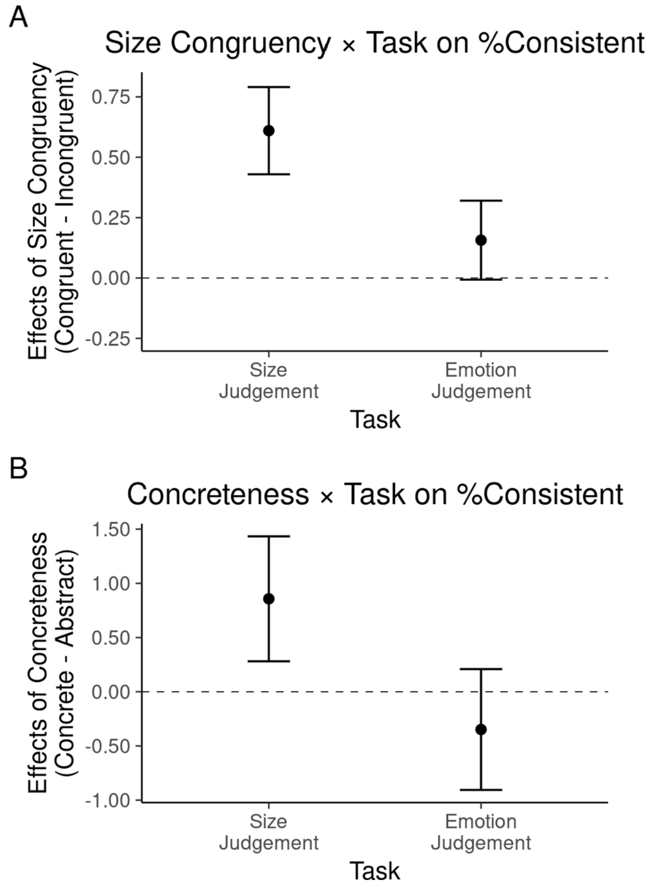

Fig. 3. The estimated effects of size congruency (Panel A) and concreteness (Panel B) on %consistent responses by task. Note: Panel A. The *y*-axis shows the coefficients of the size congruency effect. A positive coefficient means % consistent responses is higher when size is congruent than incongruent. Panel B. The *y*-axis shows the coefficients of the concreteness effect. A positive coefficient means %consistent responses is higher in concrete words than in abstract words. A negative coefficient means %consistent responses is higher in abstract words than in concrete words.

Table 6
Median RTs (in ms) across experimental conditions. Parentheses for RTs contain the interquartile range.

	Size Judgement		Emotion Judgement	
	Effective Size Congruent	Effective Size Incongruent	Effective Size Congruent	Effective Size Incongruent
Concrete	692 (264)	729 (277)	631 (279)	640 (285)
Abstract	755 (355)	800 (372)	718 (309)	737 (312)

Table 7
Generalised linear mixed-effect model estimates of fixed effects on RTs in Experiment 3.

Fixed effects	*b*	*S.E.*	*t*	*p*
Task	**67.18**	**10.18**	**6.60**	**<.001**
Effective Size Congruency	**−22.37**	**4.30**	**−5.20**	**<.001**
CNC	**−90.09**	**9.11**	**−9.89**	**<.001**
Task × Effective Size Congruency	**−26.75**	**8.82**	**−3.03**	**.002**
Task × CNC	−5.55	13.92	−.40	.690
Effective Size Congruency × CNC	8.98	7.68	1.17	.242
Task × Effective Size Congruency × CNC	−2.02	13.00	−.16	.877

Note: CNC = Concreteness. Significant effects are highlighted in bold.

random factors. The results are reported in Table 7.

All three main effects were significant (*ps* < .001). Size Judgements (845 ms) were significantly slower than Emotion Judgements (778 ms). Judgements on Concrete words (766 ms) were faster than judgements on Abstract words (857 ms), and judgements were faster when the perceived size/emotionality and font size were congruent (801 ms) than when they were incongruent (823 ms). Importantly, this Effective Size Congruency effect significantly interacted with Task, as it was significant in the Size Judgement task (Congruent: 828 ms vs Incongruent: 864 ms), b = -36.0, 95%CI = [-48.5, -23.5], but not in the Emotion Judgement task (Congruent: 774 ms vs Incongruent: 783 ms), b = -9.2, 95%CI = [-20.9, 2.4]. This Size Congruency × Task interaction is illustrated in Fig. 4.

The non-significant effect of Effective Size Congruency observed in the Emotion Judgement task could be either due to an absence of an effect, or an absence of evidence for an effect. To better describe and assess evidence for this effect, we fit a Bayesian mixed effects model of RTs using the *brm* function from the brms package (Bürkner, 2017) for R. The same maximal mixed-effects structure was used as in the generalised linear mixed model fit with lme4, but we modelled the data with an Exponentially Modified Gaussian (ex-Gaussian) distribution family. We modelled the full mixed-effects model formula for the distribution's *mu* (central tendency), and modelled population-level intercepts for the *sigma* (dispersion) and *beta* (skewness). We modelled *mu* with an identity link function, and modelled *sigma* and *beta* with log link functions. While Gamma family models are often fit with lme4, as a common distribution type supported by the package which is generally sensitive to large effects, ex-Gaussian distributions more accurately describe the shape of the distribution of RTs (Dawson, 1988). We fitted three variants of the model: a model where all fixed effects were deviation coded, and two models dummy coded to estimate effects within each task. All models were fitted using the same weakly informative normal priors for fixed effects: the prior for the intercept for *mu* was specified as centred on 1000 ms with an *SD* of 250; the priors for the intercepts of *sigma* and *beta* were specified as centred on 0 with an *SD* of 250; and the priors for all fixed effects (on *mu*) were specified as centred on 0 with an *SD* of 50. The priors for the variance of both random effects distributions (subject and item random effects) were specified as a weakly informative Student's *t*-distribution centred on 0 (df = 2, mu = 0, $sigma$ = 100). All Bayesian models were fitted with five chains, each consisting of 10,000 (7500 warmup and 2500 sampling) iterations. To facilitate model convergence, the *adapt_delta* parameter was set to.99. Fixed effects estimated from the models' posterior distributions are reported in Table 8.

The interaction between effective size congruency and task, as estimated by the Bayesian ex-Gaussian model, is illustrated in Fig. 5. The

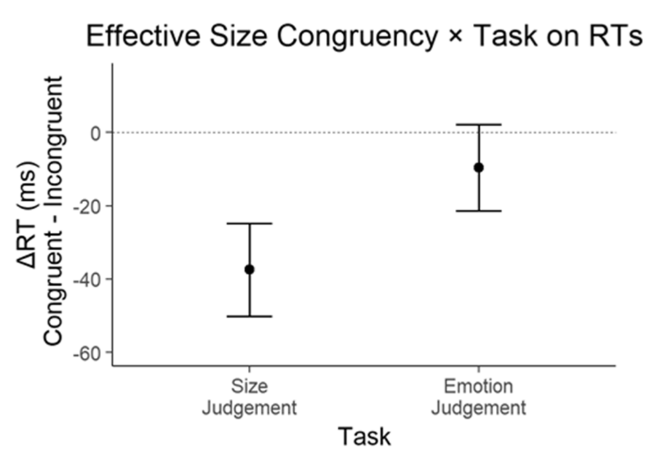

Fig. 4. The estimated effects of effective size congruency on RTs by task. Note: The *y*-axis shows the coefficients of the effective size congruency effects (i.e., the RT difference between congruent and incongruent conditions). Error bars represent the 95% confidence intervals.

Table 8
Bayesian ex-Gaussian mixed-effect model estimates of fixed effects (median of posterior samples) on RTs and 95% Credible Intervals (CrI; highest density intervals) in Experiment 3.

Fixed effects	*b*	*95% CrI*
Task	52.70	[34.10, 71.50]
Effective Size Congruency	−23.60	[−28.00, −19.20]
CNC	−35.20	[−45.80, −24.50]
Task × Effective Size Congruency	−14.10	[−25.70, −2.71]
Task × CNC	9.25	[−6.70, 25.00]
Effective Size Congruency × CNC	11.30	[2.62, 19.70]
Task × Effective Size Congruency × CNC	−7.42	[−23.70, 9.57]

Note: CNC = Concreteness; CrI = Credible Interval.

posterior distributions confirm that the effect of Effective Size Congruency is strongly modulated by task, as it is substantially larger in the Size Judgement task than in the Emotion Judgement task. However, they also show an additional, unexpected finding of a smaller but non-zero effect of Effective Size Congruency in the Emotion Judgement task.

Discussion

Experiment 3 probed the nature and flexibility of the size congruency effect observed in Experiment 2 by manipulating the task relevance of size and emotion. We assessed how consistently semantic size was mapped to visual size (big-large, small-little) or to emotion (big-emotional, small-neutral) by explicitly asking participants to judge the size and emotionality of a given word, respectively. We also measured how fast the judgements were made.

We replicated the size congruency effect on both measures and found that it did depend on task conditions (i.e., focusing on size or emotion aspects of words). In the Size Judgement task, semantically big words were more likely to be judged as “Big” when they were presented in a large rather than little font, while semantically small words were more likely to be judged as “Small” when they were presented in a little rather than large font. Size judgements were also faster when words were perceived as “Big” in a large font or were perceived as “Small” in a little font. By contrast, the size-emotion congruency effect was significantly weaker in the Emotion Judgement task. However, the effect of effective size congruency in the Emotion Judgement task was not significantly different from 0 in the generalised linear mixed-effects model, but was different from 0 in the Bayesian mixed model. This discrepancy was primarily due to the different distribution families used (Gamma vs ex-Gaussian). Taken together, the findings suggest that the size congruency effect observed in Experiment 2 *predominantly* reflected congruency between mentally simulated visual size (semantic processing) and visually perceived font size (visual processing). It remained inconclusive whether or not the size congruency effect may also be partially mediated by the congruency between the emotional make-up of abstract words and the emotion elicited by font sizes.

However, the task-dependent size congruency effects did not interact with concreteness and was significant in both abstract and concrete words. This suggests that, contrary to our hypothesis, concrete conceptual representations are as flexible as abstract ones, at least in how semantic size is represented. This finding highlights the need to move beyond characterising distinctions between abstract and concrete concepts, and to study all concepts in situational contexts (Barsalou et al., 2018).

We also observed several other effects which we will discuss here. With regard to the “consistent” judgements made, we found a significant main effect of Task. Semantically big and small words were more consistently judged as “Big” and “Small”, respectively, in the Size Judgement task while the correspondence between semantic size and emotion (big-emotional, small-neutral) was significantly less consistent in the Emotion Judgement task. The Task effect significantly interacted with Concreteness. There were significantly more “consistent” judgements on concrete rather than abstract words in the size judgement task but relatively more “consistent” judgements on abstract rather than concrete words in the Emotion Judgement task. These findings are in line with previous findings that concrete size is grounded in sensorimotor experience of physical size whereas abstract size is more rooted in the magnitude of emotional experience (Yao et al., 2013). They also highlight that while abstract words may be semantically more diverse (Hoffman et al., 2013) and situationally less systematic (Davis et al., 2020) than concrete words overall, they can be flexibly represented in a focused and systematic manner under specific contexts and tasks. Finally, size judgements were much slower than emotion judgements overall. The extra processing time in the former may reflect greater Stroop-like interference between simulated visual size and perceived font size, which is more obvious when attention is focused on size as opposed to emotion. We also observed faster processing times on concrete than abstract words, which echoes the widely reported processing speed differences between concrete and abstract words (Paivio, 1990).

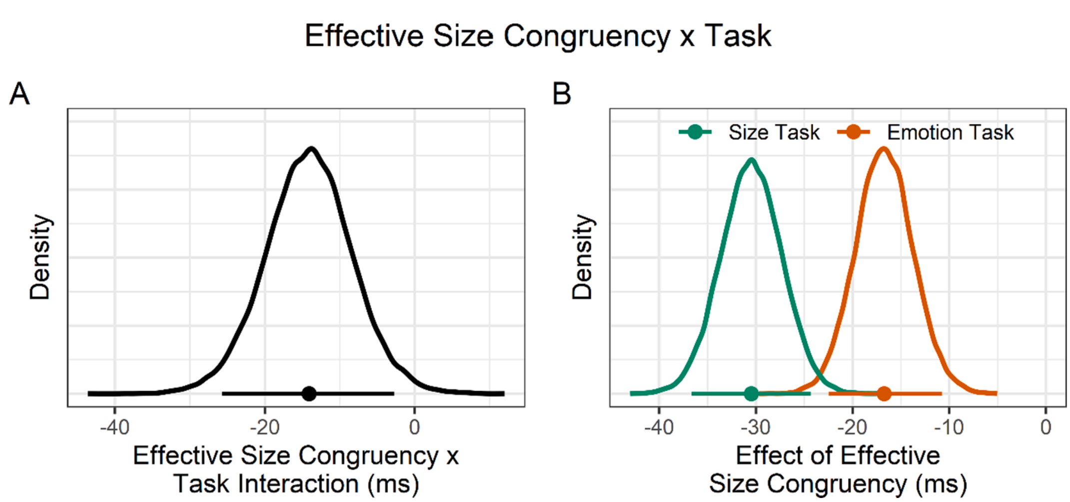

Fig. 5. The posterior distributions for the estimated effects of (A) the interaction between task and effective size congruency, and (B) the effect of effective size congruency in the size judgement (*green*) and emotion judgement (*orange*) tasks. In both panels, density plots depict the posterior distributions, points depict effect estimates (median of the posterior distribution), and horizontal whiskers depict 95% credible intervals (highest density intervals). The scale for density (*y*-axis) is identical across panels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

General discussion

In three experiments, we systematically examined whether and how semantic size in abstract concepts may be flexibly grounded in physical (visual) size across contexts and tasks. Using a forced association task, we found that semantically big abstract concepts were metaphorically associated with physically bigger concrete objects while semantically small abstract concepts were associated with physically smaller objects. These metaphorical associations were *specifically* driven by size congruency and could not be explained by lexical variables such as word frequency, length, or semantic associations between abstract concepts and concrete objects. This size congruency effect was also observed during online lexical processing. Regardless of a word's concreteness, visual recognition of semantically big words was significantly faster when they were presented in a larger rather than smaller font while recognition times of semantically small words were numerically faster in smaller rather than larger font. In a third experiment, we showed that the size congruency effect was replicated when participants explicitly judged the semantic size of the words but to a significantly lesser extent when judging the emotionality of the words. However, contrary to our hypothesis, the task-dependent size congruency effect did not vary by concreteness. These findings confirmed that semantic size of abstract concepts can indeed be represented in physical (visual) size, as a function of context (font size variation) and task demands (forced associations, lexical decision, size vs emotion judgements). They also revealed that semantic size of concrete concepts is representationally as flexible and context- and task-dependent as that of abstract concepts.

The current dataset makes three important contributions to the literature. First, it provides novel evidence for the "exogenous grounding" hypothesis of abstract concepts, beyond what has been demonstrated in specific concepts of time (Boroditsky & Ramscar, 2002) and transfer (Glenberg et al., 2008). Across three experiments, we showed that abstract concepts, despite not having physical referents, can be grounded in visual experience of size. Critically, the grounding in visual size seems automatic (as observed in lexical decisions) and thus may be included in abstract concepts' canonical representations in lexical processing. How do abstract concepts develop such grounding in visual experience of size? Intuitively, abstract concepts are unlikely to be directly associated with visual size. For example, in our daily life, big abstract words like *trust*, *eternal*, and *crisis* are not systematically written in larger fonts than small abstract words such as *trace*, *impulse*, and *humble*. Instead, abstract concepts may be more likely to 'acquire' the sense of size through repeated analogic or metaphoric associations with concrete objects (Gentner & Asmuth, 2019; Lakoff, 1987). Via the mechanism of structural alignment (Gentner, 2010), the structural commonality (e.g., in magnitude) between abstract and concrete concepts becomes progressively salient, thereby establishing more stable connections between abstract concepts and visual experiences of size. It is worth noting that the current data cannot address how much of this grounding process is driven by metaphoric use or other forms of analogic associations. What it does demonstrate is that exogenous experiential grounding of abstract concepts is pervasive (beyond special cases of time and transfer), with the *post hoc* inference that it must have developed through interactions between abstract and concrete concepts in language use.

Second, the current dataset demonstrates that exogenous experiential grounding of abstract concepts depends on contexts and tasks. In particular, we showed that abstract conceptual grounding in visual size was significantly stronger when the task focused on size rather than non-size aspects of the word. Our finding speaks against the idea of *invariant* conceptual cores (Machery, 2015), which assumes that conceptual representations are fixed and context-independent. Rather, it supports the more recent shift towards contextualism, which argues that conceptual representations are flexible, and can be heavily influenced by context and task (Barsalou, 1993; Connell & Lynott, 2014, p. 20; Lebois et al., 2015; Lupyan & Casasanto, 2015; Mazzuca et al., 2022; Wilson & Golonka, 2013; Yee & Thompson-Schill, 2016). This has important methodological implications as to whether decontextualised tasks like the lexical decision task (LDT) or semantic rating tasks are sufficient for studying conceptual processing. For example, the LDT has dominated the psycholinguistic literature as a 'all-purpose' task to study lexical processing, as it is context-independent and requires minimal semantic activation. If embodied experience is activated during the LDT, it must constitute a word's "conceptual core" – the very essence of a word's meaning, which should be automatically activated across contexts and tasks (Moors & De Houwer, 2006). Our evidence, however, contradicts this belief. Although we showed a clear interaction between a word's semantic size and visual size in an LDT (Experiment 2) and when participants judged the size of a word (Experiment 3), this size congruency effect was significantly weaker when participants judged the emotionality of a word (Experiment 3). The evidence suggests that abstract semantic size is unlikely to have a "conceptual core" as such (Machery, 2015). The sense of size or magnitude is more likely to be created "on the fly", with embodied experiences that are made available under particular task demands. Our finding suggests that the quest for invariant conceptual cores may be misguided (Connell & Lynott, 2014; Lebois et al., 2015) and that context and task demands may hold the keys in unlocking the true nature of conceptual representations.

The third contribution of the current study extends representational flexibility to concrete concepts, advocating a need to move beyond dichotomous distinctions between abstract and concrete concepts (cf. Barsalou et al., 2018). Contrary to our hypothesis, the size congruency effects in concrete concepts were as flexible and task dependent as those in abstract concepts. This is surprising because concrete conceptual representations are thought to be more bounded and stable than abstract concepts. Visual size, as an inherent part of visual experience of concrete objects, should always be relevant in representing concrete words such as a *mountain* or *caterpillar*. However, our findings illustrated that visual size representations of concrete objects can be muted when the task at hand deems it irrelevant, suggesting that flexibility is an inherent property of *all* concepts, regardless of concreteness.

In conclusion, our findings provide new evidence that abstract concepts can be represented via exogenous concrete experiences. Specifically, we show that semantic size in concepts can be represented in visual size and that these experiences are flexibly engaged under different task demands, regardless of concreteness. Our results suggest that psycholinguistic traditions of looking for the conceptual cores of words might have been misguided. Future research should focus on contextual and task effects on conceptual processing, with flexibility at the heart of its theoretical motivation and experimental design.

CRediT authorship contribution statement

Bo Yao: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Jack E. Taylor:** Methodology, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – review & editing, Visualization. **Sara C. Sereno:** Conceptualization, Methodology, Resources, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

An early version of this work was presented at the 10th Embodied and Situated Language Processing (ESLP) conference. This research was partly supported by Economic and Social Research Council Grant [ES/

N002784/1] to B. Yao. We thank the three reviewers and editor for their constructive feedback which has greatly improved this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2022.104369.

References

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Aguilar, M., Wang, X., Doko, D., & Raizada, R. D. (2017). Predicting neural activity patterns associated with sentences using a neurobiologically motivated model of semantic representation. *Cerebral Cortex, 27*(9), 4379–4395. https://doi.org/10.1093/cercor/bhw240

Andrews, M., Frank, S., & Vigliocco, G. (2014). Reconciling embodied and distributional accounts of meaning in language. *Topics in Cognitive Science, 6*(3), 359–370. https://doi.org/10.1111/tops.12096

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barsalou, L. W. (1993). *Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols*. In Theories of Memory: Psychology Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W., Dutriaux, L., & Scheepers, C. (2018). Moving beyond the distinction between concrete and abstract concepts. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170144. https://doi.org/10.1098/rstb.2017.0144

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In *Grounding cognition: The role of perception and action in memory, language, and thought* (pp. 129–163). Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bayer, M., Sommer, W., & Schacht, A. (2012). Font size matters—Emotion and attention in cortical responses to written words. *PLoS ONE, 7*(5), e36042.

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology, 33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Borghi, A. M., 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. https://doi.org/10.1016/j.plrev.2018.12.001

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018a). Varieties of abstract concepts: Development, use and representation in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170121. https://doi.org/10.1098/rstb.2017.0121

Borghi, A. M., Barca, L., Binkofski, F., & Tummolini, L. (2018b). Abstract concepts, language and sociality: From acquisition to inner speech. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1752), 20170134. https://doi.org/10.1098/rstb.2017.0134

Borghi, A. M., Binkofski, F., Castelfranchi, C., Cimatti, F., Scorolli, C., & Tummolini, L. (2017). The challenge of abstract concepts. *Psychological Bulletin, 143*(3), 263–292. https://doi.org/10.1037/bul0000089

Boroditsky, L., & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13*(2), 185–189. https://doi.org/10.1111/1467-9280.00434

Bürkner, P.-C. (2017). brms: An R package for bayesian multilevel models using stan. *Journal of Statistical Software, 80*, 1–28. https://doi.org/10.18637/jss.v080.i01.

Christensen, R. H. B. (2015). Ordinal—Regression models for ordinal data. *R Package Version, 28*, 2015.

Conca, F., Borsa, V. M., Cappa, S. F., & Catricalà, E. (2021). The multidimensionality of abstract concepts: A systematic review. *Neuroscience & Biobehavioral Reviews, 127*, 474–491. https://doi.org/10.1016/j.neubiorev.2021.05.004

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

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*(1752), 20170143. https://doi.org/10.1098/rstb.2017.0143

Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (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. https://doi.org/10.3389/fnhum.2013.00186

Davies, M. (2004). BYU-BNC: The British National Corpus. Available at Http://Corpus.Byu.Edu/Bnc.

Davis, C. P., Altmann, G. T. M., & Yee, E. (2020). Situational systematicity: A role for schema in understanding the differences between abstract and concrete concepts. *Cognitive Neuropsychology, 37*(1–2), 142–153. https://doi.org/10.1080/02643294.2019.1710124

Dawson, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, & Computers, 20*(1), 54–57. https://doi.org/10.3758/BF03202603

Dove, G. (2011). On the need for embodied and dis-embodied cognition. *Frontiers in Psychology*, *1*, 242. https://doi.org/tps://doi.org/10.3389/fpsyg.2010.00242.

Dreyer, F. R., & Pulvermüller, F. (2018). Abstract semantics in the motor system? – An event-related fMRI study on passive reading of semantic word categories carrying abstract emotional and mental meaning. *Cortex, 100*, 52–70. https://doi.org/10.1016/j.cortex.2017.10.021

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex, 26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the role of the motor system in language comprehension. *Quarterly Journal of Experimental Psychology, 61*(6), 825–850. https://doi.org/10.1080/17470210701623605

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science, 34*(5), 752–775. https://doi.org/10.1111/j.1551-6709.2010.01114.x

Gentner, D., & Asmuth, J. (2019). Metaphoric extension, relational categories, and abstraction. *Language, Cognition and Neuroscience, 34*(10), 1298–1307. https://doi.org/10.1080/23273798.2017.1410560

Glenberg, A. M., & Gallese, V. (2012). Action-based language: A theory of language acquisition, comprehension, and production. *Cortex, 48*(7), 905–922. https://doi.org/10.1016/j.cortex.2011.04.010

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*(3), 558–565. https://doi.org/10.3758/BF03196313

Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G. (2008). Processing abstract language modulates motor system activity. *Quarterly Journal of Experimental Psychology, 61*(6), 905–919. https://doi.org/10.1080/17470210701625550

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42*(1), 335–346. https://doi.org/10.1016/0167-2789(90)90087-6

Harpaintner, M., Trumpp, N. M., & Kiefer, M. (2018). The semantic content of abstract concepts: A property listing study of 296 abstract words. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/article/10.3389/fpsyg.2018.01748.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods, 45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Kang, S. H. K., Yap, M. J., Tse, C.-S., & Kurby, C. A. (2011). Rapid communication: Semantic size does not matter: "Bigger" words are not recognized faster. *Quarterly Journal of Experimental Psychology, 64*(6), 1041–1047. https://doi.org/10.1080/17470218.2011.575947

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience, 28*(47), 12224–12230. https://doi.org/10.1523/JNEUROSCI.3579-08.2008

Kousta, S.-T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General, 140*(1), 14–34. https://doi.org/10.1037/a0021446

Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2009). Emotion words, regardless of polarity, have a processing advantage over neutral words. *Cognition, 112*(3), 473–481. https://doi.org/10.1016/j.cognition.2009.06.007

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. Basic Books.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Larranaga, D., Minolli, A. M., & Sereno, A. B. (2022). *Bigger is better: Resolving conflicting behavioral evidence for semantic size bias in the lexical decision task*. Chicago: Poster at Association for Psychological Science Annual Convention.

Lebois, L. A. M., Wilson-Mendenhall, C. D., & Barsalou, L. W. (2015). Are automatic conceptual cores the gold standard of semantic processing? The context-dependence of spatial meaning in grounded congruency effects. *Cognitive Science, 39*(8), 1764–1801. https://doi.org/10.1111/cogs.12174

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

Machery, E. (2015). By default: Concepts are accessed in a context-independent manner. In *The conceptual mind: New directions in the study of concepts* (pp. 567–588). MIT press.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Mazzuca, C., Falcinelli, I., Michalland, A.-H., Tummolini, L., & Borghi, A. M. (2022). Bodily, emotional, and public sphere at the time of COVID-19. An investigation on concrete and abstract concepts. *Psychological Research*. https://doi.org/10.1007/s00426-021-01633-z.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297.

Morey, R. D., Kaschak, M. P., Díez-Álamo, A. M., Glenberg, A. M., Zwaan, R. A., Lakens, D., … Ziv-Crispel, N. (2021). A pre-registered, multi-lab non-replication of the action-sentence compatibility effect (ACE). *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01927-8

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532–1543). https://doi.org/10.3115/v1/D14-1162

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6*(7), 576–582. https://doi.org/10.1038/nrn1706

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences, 17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica, 111*(1), 141–154. https://doi.org/10.1016/S0001-6918(02)00047-1

Schwanenflugel, P. J. (1991). *Why are abstract concepts hard to understand. In The psychology of word meanings* (Vol. 11,, 223–250.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*(5), 499–520. https://doi.org/10.1016/0749-596X(88)90022-8

Scott, G. G., Keitel, A., Becirspahic, M., Yao, B., & Sereno, S. C. (2019). The Glasgow Norms: Ratings of 5,500 words on nine scales. *Behavior Research Methods, 51*(3), 1258–1270. https://doi.org/10.3758/s13428-018-1099-3

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *Quarterly Journal of Experimental Psychology, 62*(6), 1115–1122. https://doi.org/10.1080/17470210802618900

Setti, A., Caramelli, N., & Borghi, A. M. (2009). Conceptual information about size of objects in nouns. *European Journal of Cognitive Psychology, 21*(7), 1022–1044. https://doi.org/10.1080/09541440802469499

Skipper, L. M., & Olson, I. R. (2014). Semantic memory: Distinct neural representations for abstractness and valence. *Brain and Language, 130*, 1–10. https://doi.org/10.1016/j.bandl.2014.01.001

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science, 20*(8), 989–999. https://doi.org/10.1111/j.1467-9280.2009.02397.x

Taylor, J. E., Beith, A., & Sereno, S. C. (2020). LexOPS: An R package and user interface for the controlled generation of word stimuli. *Behavior Research Methods, 52*(6), 2372–2382. https://doi.org/10.3758/s13428-020-01389-1

Troche, J., Crutch, S. J., & Reilly, J. (2017). Defining a conceptual topography of word concreteness: Clustering properties of emotion, sensation, and magnitude among 750 English words. *Frontiers in Psychology*, *8*. https://www.frontiersin.org/article/10.3389/fpsyg.2017.01787.

Troche, J., Crutch, S., & Reilly, J. (2014). Clustering, hierarchical organization, and the topography of abstract and concrete nouns. *Frontiers in Psychology, 5*, 360.

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Vigliocco, G., Kousta, S.-T., Della Rosa, P. A., Vinson, D. P., Tettamanti, M., Devlin, J. T., & Cappa, S. F. (2014). The neural representation of abstract words: The role of emotion. *Cerebral Cortex, 24*(7), 1767–1777. https://doi.org/10.1093/cercor/bht025

Villani, C., Lugli, L., Liuzza, M. T., & Borghi, A. M. (2019). Varieties of abstract concepts and their multiple dimensions. *Language and Cognition, 11*(3), 403–430. https://doi.org/10.1017/langcog.2019.23

Wilson, A., & Golonka, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology, 4*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00058.

Yao, B., Vasiljevic, M., Weick, M., Sereno, M. E., O'Donnell, P. J., & Sereno, S. C. (2013). Semantic size of abstract concepts: It gets emotional when you can't see it. *PLoS ONE, 8*(9), e75000.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review, 23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

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

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language comprehenders mentally represent the shapes of objects. *Psychological Science, 13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

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