Cognition 125 (2012) 452–465

Contents lists available at SciVerse ScienceDirect

Cognition

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

Strength of perceptual experience predicts word processing performance better than concreteness or imageability

Louise Connell [a,*], Dermot Lynott [b]

[a] School of Psychological Sciences, Oxford Road, Manchester M13 9PL, UK
[b] Decision and Cognitive Sciences Research Centre, Manchester Business School, MBS East, Booth Street West, Manchester M15 6PB, UK

A R T I C L E I N F O

Article history:
Received 22 February 2012
Revised 6 July 2012
Accepted 11 July 2012
Available online 27 August 2012

Keywords:
Abstract and concrete concepts
Concreteness effects
Imageability
Perceptual strength
Lexical decision
Word naming
Situated simulation
Dual coding
Context availability

A B S T R A C T

Abstract concepts are traditionally thought to differ from concrete concepts by their lack of perceptual information, which causes them to be processed more slowly and less accurately than perceptually-based concrete concepts. In two studies, we examined this assumption by comparing concreteness and imageability ratings to a set of perceptual strength norms in five separate modalities: sound, taste, touch, smell and vision. Results showed that concreteness and imageability do not reflect the perceptual basis of concepts: concreteness ratings appear to be based on two different intersecting decision criteria, while imageability ratings are visually biased. Analysis of lexical decision and word naming performance showed that maximum perceptual strength (i.e., strength in the dominant perceptual modality) consistently outperformed both concreteness and imageability ratings in accounting for variance in response latency and accuracy. We conclude that so-called concreteness effects in word processing emerge from the perceptual strength of a concept's representation and discuss the implications for theories of conceptual representation.

1. Introduction

What exactly constitutes an abstract concept? Traditionally, abstract words such as *truth* or *impossible* are assumed to refer to things that are not perceptually experienced, while concrete words such as *chair* or *turquoise* are assumed to refer to perceptible, material entities (Heidbreder, 1945; Piaget & Inhelder, 1969; Reed & Dick, 1968). A long history of research has examined processing differences between such abstract and concrete concepts. In particular, concreteness effects refer to a behavioral advantage for words that refer to concrete concepts, which are processed more quickly and accurately than abstract concepts in lexical decision tasks (Binder, Westbury, McKiernan, Possing, & Medler, 2005; Bleasdale, 1987; de Groot, 1989; Howell & Bryden, 1987; James, 1975; Kroll & Merves, 1986; Rubin, 1980; Schwanenflugel, Harnishfeger, & Stowe, 1988; Schwanenflugel & Shoben, 1983; Whaley, 1978), word naming (de Groot, 1989; Schwanenflugel & Stowe, 1989), and recall (Allen & Hulme, 2006; Fliessbach, Weis, Klaver, Elger, & Weber, 2006; Paivio, Yuille, & Smythe, 1966; Romani, McAlpine, & Martin, 2007; Walker & Hulme, 1999).

A number of different theories have been proposed to account for concreteness effects in word processing performance. Dual coding theory (Paivio, 1971, 1986, 2007) holds that both concrete and abstract concepts have a verbal code representation, but that concrete concepts alone also have a nonverbal, imagistic representation available. These imagistic codes are perceptual and modality-specific (i.e., visual, auditory, haptic, etc.) in nature, and "give rise to conscious (reportable) imagery when activated" (Paivio, 2007, p. 39). Concreteness is defined as the directness of connections between these verbal and imagistic representations; concrete concepts have the most direct connections, whereas abstract concepts have only indirect

* Corresponding author. Tel.: +44 161 306 0447.
E-mail addresses: louise.connell@manchester.ac.uk (L. Connell), dermot.lynott@manchester.ac.uk (D. Lynott).

0010-0277/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2012.07.010

connections to images via other verbal codes (e.g., *religion* has no imagistic code of its own but nonetheless may be concretised via *church*, which does have an imagistic code). Abstract words are therefore slower to process because they can only be imaged indirectly, via related concrete words. Context availability theory (Schwanenflugel, 1991; Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1988) instead argues that the type of information is less important than the quantity, and that concrete concepts are strongly linked to a narrow range of supporting contexts in memory whereas abstract concepts are weakly linked to a wide range. People are slower to process abstract words because they find it more difficult to retrieve associated contextual information. More recently, situated simulation views of conceptual representation (Barsalou, Santos, Simmons, & Wilson, 2008; Barsalou & Wiemer-Hastings, 2005; see also Vigliocco, Meteyard, Andrews, & Kousta, 2009) have drawn together several aspects of both dual coding and context availability theories. Both concrete and abstract concepts are represented as situated simulations – that is, a partial re-enactment of the neural activation in perceptual, motor, affective and other systems that arises during experience of those concepts – but they differ in the focal content of their situations. Concrete concepts are represented in a narrow range of situations that focus more on perceptual and motor information, while abstract concepts have a wide range of situations that focus more on social, introspective and affective information. Abstract words are slower to process because people find it more difficult to access their complex, broadly-focused situations.

However, despite their reputation as a textbook effect, concreteness effects do not always reliably emerge in semantic processing. Null effects are rarely publishable and tend to languish in experimenters' file drawers (Rosenthal, 1979), but lack of concreteness effects in response times and error rates do occasionally appear (e.g., Barca, Burani, & Arduino, 2002), and are not uncommon in cognitive neuroscience studies where significant findings on other measures are reported alongside null behavioral results (Fiebach & Friederici, 2004; Papagno, Fogliata, Catricalà, & Miniussi, 2009; Sabsevitz, Medler, Seidenberg, & Binder, 2005; Tsai et al., 2009; Tyler, Russell, Fadili, & Moss, 2001). For example, Papagno et al. (2009) found that concrete and abstract words were processed equally quickly when participants were *not* undergoing TMS (i.e., their control condition), and Tsai et al. (2009) also found no speed or accuracy differences between abstract and concrete items in a lexical decision task on Chinese ideographs. Furthermore, while reverse concreteness effects – a processing advantage for abstract concepts rather than concrete – have been found in certain semantic memory disorders (e.g., Breedin, Saffran, & Coslett, 1994; Macoir, 2008; Warrington & Shallice, 1984; Yi, Moore, & Grossman, 2007), they also been found in studies of healthy adult participants. In a large-scale regression analysis, Kousta, Vigliocco, Vinson, Andrews, and Del Campo (2011) found that, when a large number of psycholinguistic variables were partialled out, lexical decision times for abstract words were faster than those for concrete words (see also Adelman, Brown, & Quesada, 2006). Pexman, Hargreaves, Edwards, Henry, and Goodyear (2007) also found an abstract word advantage in a semantic task that asked people to judge whether or not a word denoted a consumable (i.e., edible/drinkable) item, even though this task presumably required deeper conceptual processing than a word/nonword lexical decision task. Such null and reversed concreteness effects are problematic for theories that claim a fundamental processing advantage for concrete over abstract concepts.

One reason for inconsistencies in empirical tests of concreteness effects may be that the theoretical assumption is valid (i.e., that concepts with perceptual information are faster to process), but that the typical basis for selecting experimental items (i.e., concreteness or imageability ratings) does not offer an accurate measure of the perceptual basis of concepts. Most researchers employ a set of published norms when selecting abstract and concrete materials, and perhaps the most widely-used source is the MRC psycholinguistic database (Coltheart, 1981; Wilson, 1988; available online at http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm), which contains both concreteness and imageability ratings on 7-point scales. However, when we examined the original norming instructions used to collect these ratings, we found it questionable that participants would have simultaneously considered their sensory experience across all modalities and then managed to aggregate this experience into a single, composite rating per word.

For example, the instructions for MRC concreteness ratings originate with Spreen and Schultz (1966), who labeled the abstract end of the scale as "low concreteness", although much of the MRC norming data comes from Paivio, Yuille and Madigan's (1968) variant where it is labeled as "high abstractness":

> Any word that refers to objects, materials, or persons should receive a h*igh concreteness* rating; any word that refers to an abstract concept that cannot be experienced by the senses should receive a *high abstractness* rating. (p. 5)

While these instructions make direct reference to the extent to which concepts are experienced through the various senses, we thought it possible that defining concrete words as referring to "objects, materials, or persons" was insufficiently clear for participants to understand they should consider any sensory experience as a form of concreteness. The resulting ratings, therefore, may reflect different decision criteria at the concrete and abstract ends of the scale – one that judged relation to objects, materials and persons, and one that judged absence of sensory experience – which could add considerable noise to any dataset that assumed the ratings reflected a smooth continuum of perceptual experience. Indeed, the finding that the abstractness–concreteness ratings scale has a bimodal distribution (Kousta et al., 2011; Nelson & Schreiber, 1992) is consistent with this notion.

Imageability ratings are frequently used interchangeably with concreteness ratings in the experimental literature (e.g., Binder et al., 2005; Fliessbach et al., 2006; Sabsevitz et al., 2005) because of their high correlation and theoretical relationship in dual coding theory. Instructions for imageability ratings in the MRC database originate with Paivio et al. (1968):

> Any word which, in your estimation, arouses a mental image (i.e., a mental picture, or sound, or other sensory experience) very quickly and easily should be given a *high imagery* rating; any word that arouses a mental image with difficulty or not at all should be given a *low imagery* rating. (p. 4)

Here, we considered it possible that asking naïve participants to consider a "mental image", regardless of the mention of other senses, was likely to have led them to focus on vision at the expense of other modalities. Even though people experience the world through many senses, they are more practiced at consciously generating imagery through the mind's eye than, say, the mind's tongue or nose. Datasets that use imageability ratings to test for concreteness effects may therefore be subject to noise as expertise in visual imagery is confounded with perceptual experience of the concept.

Our goals in the present paper were twofold. First, we aimed to establish whether concreteness and imageability norms actually reflect the degree with which concepts are perceptually experienced, as is commonly assumed. Second, we examined whether classic concreteness effects in lexical decision and naming tasks are better predicted by concreteness/imageability ratings or by strength of perceptual experience. If the former, then forty years of empirical methodology have been validated, but the reasons for null and reverse concreteness effects remain unclear. If the latter, then concreteness and imageability ratings are not well-suited to the tasks in which they are employed, and null and reverse concreteness effects are likely to be due to the unreliability of perceptual information in such ratings. Since the theoretical assumptions of many theories of conceptual content are based on findings from concreteness and imageability norms (Barsalou & Wiemer-Hastings, 2005; Paivio, 1986, 2007; Vigliocco et al., 2009; cf. Schwanenflugel, 1991; Schwanenflugel & Shoben, 1983), the results of these studies will allow us to test whether the primary difference between concrete and abstract concepts is indeed one of perceptual information, and whether this difference underlies concreteness effects.

2. Study 1: do concreteness and imageability capture sensory experience?

Rather than ask participants to condense their estimations of sensory experience into a single concreteness or imageability rating, an alternative approach is to ask participants to explicitly consider each perceptual modality in turn. Modality-specific norms contain ratings of how strongly people experience a variety of concepts by hearing, tasting, feeling through touch, smelling and seeing (i.e., through each of five perceptual modalities in turn: auditory, gustatory, haptic, olfactory or visual: Lynott & Connell, 2009, in press). These norms have previously been used in a variety of studies of conceptual processing, such as examining the tactile modality disadvantage in word processing (Connell & Lynott, 2010), modality switching costs in both property verification (Collins, Pecher, Zeelenberg, & Coulson, 2011; Hald, Marshall, Janssen, & Garnham, 2011; Louwerse & Connell, 2011; Lynott & Connell, 2009) and conceptual combination (Connell & Lynott, 2011), and the differential activation of semantic information by perceptual attention (Connell & Lynott, in preparation).

In the current study, we compared ratings of concreteness and imageability to ratings of auditory, gustatory, haptic, olfactory and visual strength across a set of several hundred words. If concreteness and imageability are a fair reflection of the degree of perceptual information in a concept, then ratings of perceptual strength in all five modalities should be positively related to concreteness and imageability ratings, and these relationships should remain consistent across the rating scale. On the other hand, if we were correct in our hypothesis to the contrary, then we would expect some perceptual modalities to be neglected (i.e., no relationship) or even misinterpreted (i.e., negative relationship) in concreteness and imageability ratings. Specifically, concreteness norming instructions may have led to different decision criteria and therefore distinctly different modality profiles at each end of the scale, whereas imageability instructions may have led to a predominantly visual bias.

2.1. Method

2.1.1. Materials

A total of 592 words were collated that each had ratings of perceptual strength on five modalities as well as concreteness and imageability (see Table 1 for sample items). The only selection criteria for this word set was availability of all three norms. Perceptual strength norms came from Lynott and Connell (2009, in press and additional unpublished extensions), in which participants were asked to rate "to what extent do you experience something being WORD" (for adjectives) or "to what extent do you experience WORD" (for other words) through each of the five senses (i.e., "by hearing", "by tasting", "by feeling through touch", "by smelling" and "by seeing"), using separate rating scales for each modality. Perceptual strength ratings therefore took the form of a 5-value vector per word, ranging from 0 (low strength) to 5 (high strength).

Concreteness ratings were taken from the MRC psycholinguistic database for 522 words, with ratings for the remaining 70 words coming from Nelson, McEvoy, and Schreiber (2004), who replicated the instructions of Paivio et al. (1968) when collecting their norms (i.e., all ratings emerged from comparable instruction sets). Imageability ratings for 524 words also came from the MRC database, and were supplemented with ratings for a further 68 words from Clark and Paivio's (2004) extension of Paivio et al.'s (1968) original norms. Both concreteness and imageability ratings ranged from 100 (abstract or low-imageability) to 700 (concrete or high-imageability). The correlation between imageability and concreteness for this word set was high, $r(590) = .828$, $p < .0001$, and comparable to previous studies (e.g., $r = .83$ in Paivio et al., 1968).

2.1.2. Design and analysis

We first ran stepwise regression analyses across the full scales, with either concreteness or imageability rating as the dependent variable, and ratings of auditory, gustatory, haptic, olfactory and visual strength as competing predic-

Table 1
Sample words, used in Studies 1 and 2, for which perceptual strength matches or mismatches ratings of concreteness and imageability.

Word	Perceptual strength					Concreteness	Imageability
	Auditory	Gustatory	Haptic	Olfactory	Visual		
Strongly perceptual, high concreteness/imageability							
Hen	3.53	1.12	2.35	1.47	3.82	631	597
Soap	0.35	1.29	4.12	4.00	4.06	589	600
Yellow	0.15	0.05	0.00	0.05	4.90	518	597
Strongly perceptual, low concreteness/imageability							
Fear	2.18	0.71	1.88	0.82	3.47	326	394
Noisy	4.95	0.05	0.29	0.05	1.67	293	138
Quality	3.06	3.41	4.06	3.12	4.29	274	349
Weakly perceptual, high concreteness/imageability							
Air	1.06	1.47	2.12	2.53	1.35	581	450
Atom	1.00	0.63	0.94	0.50	1.38	481	499
Hell	2.47	0.24	1.06	0.71	1.24	355	519
Weakly perceptual, low concreteness/imageability							
Aspect	1.88	0.50	0.80	1.00	2.38	217	233
Factor	1.31	0.38	0.31	0.06	1.88	328	269
Republic	0.53	0.67	0.27	0.07	1.79	376	356

Note: perceptual strength ratings range from 0 to 5, concreteness and imageability ratings range from 100 to 700.

tors (inclusion criterion $p < .05$; exclusion criterion $p > .1$; see Table 2 for zero-order correlations). Second, for analysis of consistency across the scales, we split each scale at its midpoint and re-fitted the stepwise regression models. Concreteness ratings were split into abstract (rating 100–400, $N = 294$) and concrete (rating 401–700, $N = 298$) groups, whereas imageability ratings were split into low-imageability (rating 100–400, $N = 167$) and high-imageability (rating 401–700, $N = 425$) groups. *A priori* sensitivity analysis confirmed that the sample size of the smallest group (low-imageability words) was still large enough to capture even a low degree of fit (minimum $R^2 = .074$) in a five-predictor regression model at power of 0.8. Here and elsewhere, all reported *p*-values are two-tailed.

2.2. Results and discussion

2.2.1. Concreteness

Analysis showed clear dissociations between concreteness and modality-specific perceptual experience. While perceptual strength on all five modalities contributed to the regression model of concreteness ratings, $F(5,586) = 42.63$, $p < .0001$, $R^2 = .267$, the direction of the relationship varied (see Table 3). Olfactory strength ratings showed the strongest positive relation to concreteness, followed by visual and haptic ratings. However, auditory and gustatory ratings were negatively related to concreteness: the more strongly a concept related to sound or taste experience (e.g., *noisy*, *bitter*), the more *abstract* it was considered to be.

When split into abstract and concrete groups, there was little consistency in which perceptual modalities contributed most to concreteness ratings (see Fig. 1). Abstract words' ratings were predicted by three of the five modalities, $F(3,290) = 8.64$, $p < .0001$, $R^2 = .082$, but with a low degree of fit and inconsistency in the direction of the relationship: positively related to vision, and negatively to auditory and olfactory strength. In contrast, concrete words' ratings were predicted positively by olfactory and visual strength, $F(2,295) = 33.52$, $p < .0001$, $R^2 = .185$, but these two perceptual modalities offered a higher degree of fit than the model for abstract words.

Overall, results are consistent with the idea that concreteness ratings reflect different decision criteria at the concrete and abstract ends of the scale. Only one perceptual modality showed consistent behavior, with visual strength positively predicting concreteness ratings in both concrete and abstract groups. The other modalities either failed to retain a consistent relationship with concreteness across the scale (auditory, olfactory) or lost their predictive

Table 2
Zero-order correlations between concreteness, imageability, and perceptual strength ratings for each modality predictor in Study 1 (N = 592).

Modality	Concreteness	Imageability	Auditory	Gustatory	Haptic	Olfactory	Visual
Concreteness	–	.828**	−.233**	.035	.355**	.246**	.400**
Imageability		–	−.142**	.028	.314**	.236**	.496**
Auditory			–	−.081*	−.242**	−.035	−.118**
Gustatory				–	.235**	.688**	.000
Haptic					–	.287**	.468**
Olfactory						–	.217**
Visual							–

* $p < .05$.
** $p < .01$.

Table 3
Standardized regression coefficients for each modality of perceptual strength as model predictors of concreteness and imageability ratings in Study 1.

Rating	Auditory	Gustatory	Haptic	Olfactory	Visual
Concreteness	−0.197**	−0.241**	+0.164**	+0.321**	+0.230**
Imageability	−0.113*	−0.154*	+0.072	+0.253**	+0.428**

* $p < .05$.
** $p < .01$.

value entirely when analysed within abstract or concrete groups (gustatory, haptic). However, the most serious conflict concerned the inversion of the olfactory effect. As one moved from the abstract to concrete ends of the scale, the effect of olfactory strength flipped from negative to positive: more olfactory meant more abstract, but more olfactory also meant more concrete. Such inconsistency in behavior poses serious problems for the assumption that abstractness and concreteness represent two ends of the same continuum (Paivio et al., 1968), and rather indicates that participants behaved differently as they tended towards the abstract or concrete end of the original norming scale. One possible reason for this inconsistency is that participants did not interpret the norming instructions to understand they should consider any sensory experience as a form of concreteness. Specifically, instructing participants that an abstract concept "cannot be experienced by the senses" (Paivio et al., 1968, p. 5) appears to have led them to neglect the senses of touch and taste and misconstrue the senses of hearing and smell. In other words, if something could be heard or smelled but not seen, participants were likely to judge it as abstract, but given that only 8% of the concreteness variance in abstract words was explained by perceptual information, many other criteria also played a role in their decision. Furthermore, instructing participants that a concrete word "refers to objects, materials, or persons" (Paivio et al., 1968, p. 5; Spreen & Schultz, 1966, p. 460) led them to overemphasize vision and smell, and neglect taste, touch and hearing experience completely. While perceptual strength can explain more than twice the concreteness variance in concrete words (19%) as it did in abstract words (8%), participants are clearly basing their concreteness decision on non-perceptual information. It therefore appears that participants in abstractness–concreteness norming studies treated the scale as two intersecting continua, neither of which reliably reflects the extent of sensory experience.

2.2.2. Imageability

Analysis of imageability showed a clear visual bias at the expense of other perceptual modalities. All modalities except haptic strength contributed to the model of imageability ratings across the full scale, $F(4,587) = 58.04$, $p < .0001$, $R^2 = .283$, but, as found for concreteness, the direction of the relationships varied (see Table 3). The best predictor was visual strength, followed by olfactory strength; both of which were positively related to imageability. However, as with concreteness, auditory and gustatory ratings were negatively related to imageability: the more strongly a concept related to sound or taste experience (e.g., *noisy*, *bitter*), the less imageable it was considered to be.

Imageability ratings fared little better than concreteness ratings regarding consistency in behavior across the scale (see Fig. 2). Ratings of low-imageability words were predicted by two perceptual modalities: visual strength (positively) and olfactory strength (negatively), $F(2,164) = 16.42$, $p < .0001$, $R^2 = .167$. High-imageability ratings, on the other hand, were related to three perceptual modalities with a similar degree of fit, $F(3,421) = 36.32$, $p < .0001$, $R^2 = .206$: positively for both visual and olfactory information, and negatively for gustatory.

In short, people do not find it equally easy to generate imagery across the range of modalities that constitute

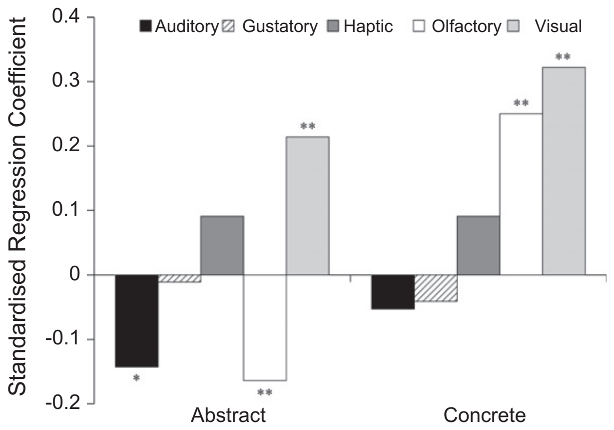

Fig. 1. Standardized regression coefficients for modality predictors of concreteness ratings in Study 1 at the abstract and concrete ends of the scale (*$p < .05$, **$p < .01$).

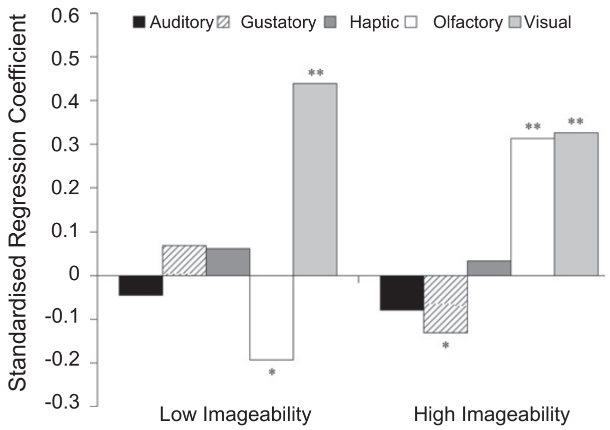

Fig. 2. Standardized regression coefficients for modality predictors of imageability ratings in Study 1 at the low- and high-imageability ends of the scale (*$p < .05$, **$p < .01$).

perceptual experience. Participants tend to rely on visual experience when generating imageability ratings: visible things are highly imageable and invisible things are not. However, this focus on vision led other modalities to be neglected or misinterpreted. Haptic experience was not reflected at either end of the scale, or across the scale as a whole (i.e., both low- and high-imageability concepts are equally likely to be touchable), which may be because visual imagery subsumed haptic imagery: most things that can be touched can also be seen. Gustatory and olfactory experience are often closely coupled (e.g., flavor and food concepts), but there were major differences in how easily participants generated their imagery: people tended to misconstrue olfactory information and ignore gustatory information for low-imageability concepts, yet follow olfactory strength while misinterpreting gustatory strength for high-imageability concepts. One reason for this difference is that the multimodal experience of olfactory and gustatory modalities is not isomorphic: many concepts are high in olfactory strength (e.g., *paint*, *air*) while having little presence on the gustatory modality. Finally, even though people can generate auditory imagery when a task requires it (Hubbard, 2010), auditory information was overall negatively related to general ratings of imageability: concepts that were strongly sound-related tended to be misconstrued as low-imageability. Together, these results indicate that participants in imageability norming studies were unable to consider all sensory modalities even though they were explicitly instructed to do so. Participants seem to have had difficulty in extending the meaning of "image" beyond its conventional interpretation as a visual impression; the modality in which frequent use of the mind's eye has lent them the most expertise.

3. Study 2: does perceptual strength outperform concreteness and imageability ratings in predicting word processing performance?

Since neither concreteness nor imageability ratings reflect the full range of sensory experience, it raises the question of whether textbook concreteness effects in word processing are actually due to (a) the degree of perceptual information in each referent concept's representation, or (b) some other conceptually meaningful information that makes up most of the variance in concreteness and imageability ratings. The present study aimed to resolve this question by comparing the unique predictive abilities of concreteness ratings, imageability ratings and perceptual strength in lexical decision and word naming performance. We have shown elsewhere (Connell & Lynott, in preparation) that visual strength is the best modality predictor of lexical decision (a task which focuses visual attention on word forms) and that auditory strength is the best modality predictor of word naming (a task which focuses auditory attention on pronunciation). However, it is convenient for statistical and other reasons, including fair comparison with concreteness and imageability ratings, to have a single predictor of perceptual strength in place of the five-modality vector used in Study 1. We therefore utilized maximum perceptual strength[1], which represents the highest rating in the concept's dominant modality and the maximum component of the vector.

Although different theories concur that concrete concepts have a greater reliance on perceptual information than abstract concepts (e.g., Barsalou & Wiemer-Hastings, 2005; Paivio, 1986, 2007; Vigliocco et al., 2009), it remains unclear to what extent perceptual differences constitute concreteness effects. If so-called concreteness effects are due to the degree of perceptual information in each referent concept's representation, then perceptual strength should outperform concreteness and imageability in predicting latency and accuracy in word processing. In this case, we would expect perceptual strength to exhibit an independent effect in the presence of concreteness/imageability predictors, but not vice versa, because most of the variance in concreteness and imageability ratings reflects decision criteria that are unrelated to processing performance. On the other hand, if concreteness effects are actually due to some other non-perceptual representation that is captured by concreteness and imageability ratings, then they would maintain an independent effect even when perceptual strength has already been partialled out. In this sense, concreteness effects would subsume perceptibility effects, because the variance in concreteness and imageability ratings would reflect conceptually meaningful information in addition to perceptual differences.

3.1. Method

3.1.1. Materials

The same set of 592 words from Study 1 was used in this study, along with lexical decision and naming data from the Elexicon database (Balota et al., 2007; available online at http://elexicon.wustl.edu), which also provided lexical characteristics for each word to act as independent regression variables (see below).

3.1.2. Design and analysis

Hierarchical regression analyses determined the proportion of variance each candidate rating could explain. The dependent variables were mean lexical decision and naming times for each word (M = 633 ms, SD = 64 ms; M = 622 ms, SD = 53 ms; respectively), and their accompanying mean accuracy rates (M = 96.6%, SD = 5.4%; M = 98.9%, SD = 3.3%). As well as raw RT in ms, we also analysed standardized RT based on the mean z-scores of the original participants in the Elexicon data, which offers a more reliable measure of latency by partialing out individual differences in overall speed and variability (Balota et al., 2007; Faust, Balota, Spieler, & Ferraro, 1999). As independent variables in all regressions, we used a basic model found by Brysbaert and New (2009) to provide the best fit

[1] We considered a range of possible methods of compressing the five-value vector of perceptual strength into a single variable, including vector length (i.e., Euclidean distance of the five-modality vector from the origin), summed or average strength across all modalities, and principal components reduction to one dimension. Maximum perceptual strength was considerably better than any other method in predicting lexical decision and naming performance, and so we selected it to compete against concreteness and imageability.

Table 4
Correlations between predictors in regression models of Study 2 ($N = 592$).

Predictor	C	I	P	L	S	Log CD	Log² CD
(C)oncreteness	–	.828**	.427**	−.334**	−.330**	−.108**	−.103*
(I)mageability		–	.502**	−.295**	−.300**	−.024	−.004
(P)erceptual strength			–	−.139**	−.144**	.117**	.191**
Number of (L)etters				–	.853**	−.263**	−.200**
Number of (S)yllables					–	−.224**	−.161**
Log contextual diversity (CD)						–	.515**
Log² contextual diversity (CD)							–

* $p < .05$.
** $p < .01$.

for RT and accuracy: log contextual diversity (LgSUBTLCD variable in Elexicon: $M = 2.83$, $SD = 0.61$), log² contextual diversity ($M = 8.40$, $SD = 3.36$), number of letters in the word ($M = 5.63$, $SD = 2.00$), and number of syllables in the word ($M = 1.63$, $SD = 0.88$). Contextual diversity (i.e., proportion of documents in which a particular word form appears) was used instead of global word form frequency because it is a better predictor of lexical decision and naming times and accuracy (Adelman et al., 2006; Brysbaert & New, 2009).[2] In addition, log² contextual diversity was included in order to capture variance in performance for very high frequency words, which tend to be subject to floor effects on a logarithmic curve.

After fitting the basic model to each dependent variable, we separately examined the simple effects of each predictor: concreteness ($M = 427$, $SD = 107$), imageability ($M = 461$, $SD = 92$), and maximum perceptual strength ($M = 3.78$, $SD = 0.75$). We then added a subsequent predictor to each model in order to test for independent effects: specifically, whether an increase in fit resulted from adding concreteness or imageability to a model that already contained perceptual strength, or from adding perceptual strength to a model that already contained concreteness or imageability.

3.2. Results and discussion

3.2.1. Comparison of predictors

Table 4 shows zero-order correlations between predictors used in all models. Concreteness and imageability correlated well in this sample, with 68% shared variance, which is comparable with previous studies (e.g., 69%: Paivio et al., 1968). In contrast, maximum perceptual strength correlated relatively poorly with both concreteness and imageability, sharing only 18% and 25% of variance, respectively. Fig. 3 illustrates how ratings vary across a representative sample of words.

One other striking difference emerged. Previous research has found that contextual diversity is inversely correlated with concreteness and imageability (i.e., abstract words appear in a greater variety of contexts than do concrete words: Galbraith & Underwood, 1973; Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008; Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1988). Our data were consistent with this established pattern: concreteness and imageability were negatively correlated with contextual diversity (although imageability's weaker relationship was non-significant). Yet, in sharp contrast, perceptual strength was *positively* correlated with contextual diversity. That is, although although concrete-rated words have a narrower variety of contexts than abstract-rated words, perceptually strong words have a wider variety than perceptually weak words. We return to this issue in the general discussion.

3.2.2. Regressions

Overall, perceptual strength outperformed concreteness and imageability in accounting for variance in lexical decision and naming (with no problems of multicollinearity: all VIFs < 2). In simple effects (i.e., over and above a basic model of contextual diversity and word length[3]), maximum perceptual strength provided the best increase in fit for all dependent measures except naming accuracy, where no predictor outside the basic model was reliable (see Fig. 4). Imageability ratings acted as a significant predictor only for lexical decision data (raw RT, standardized RT, and accuracy), but failed to explain any variance in word naming performance except for a marginal contribution to standardized RT. Concreteness ratings fared quite poorly and predicted only a single dependent variable: lexical decision raw RT. In all cases, significant predictors operated in the expected direction (see Table 5), with higher perceptual strength, imageability and concreteness all leading to faster latencies and greater accuracy. Effects of concreteness and imageability, where they appeared, were comparable in size to past

[2] Importantly, Adelman et al. (2006) show that contextual diversity independently facilitates word processing even when word frequency has been partialled out, but not vice versa. When we test for the independent effect of contextual diversity in our own item set (i.e., add log contextual diversity as a predictor to a model already containing log word frequency, length in letters, and number of syllables), we replicate Adelman et al.'s findings that contextual diversity facilitates lexical decision and naming times, and improves accuracy in both tasks, independent of word frequency. Finally, when we add our variables of interest to this model that includes both frequency and contextual diversity, we replicate the simple and independent effects of perceptual strength, concreteness, and imageability that we report in Figs. 4 and 5. In other words, contextual diversity subsumes frequency, and the inclusion of frequency makes no difference to the results. We thank Diane Pecher for this suggestion.

[3] The basic model explained the largest proportion of variance in lexical decision raw RT [42.4%, $F(4,587) = 108.10$, $p < .0001$], standardized RT [52.1%, $F(4,587) = 159.91$, $p < .0001$] and accuracy [36.8%, $F(4,587) = 85.89$, $p < .0001$], as well as in word naming raw RT (33.7%, $F(4,587) = 74.63$, $p < .0001$], standardized RT (34.7%, $F(4,587) = 78.11$, $p < .0001$] and accuracy (9.1%, $F(4,587) = 14.77$, $p < .0001$].

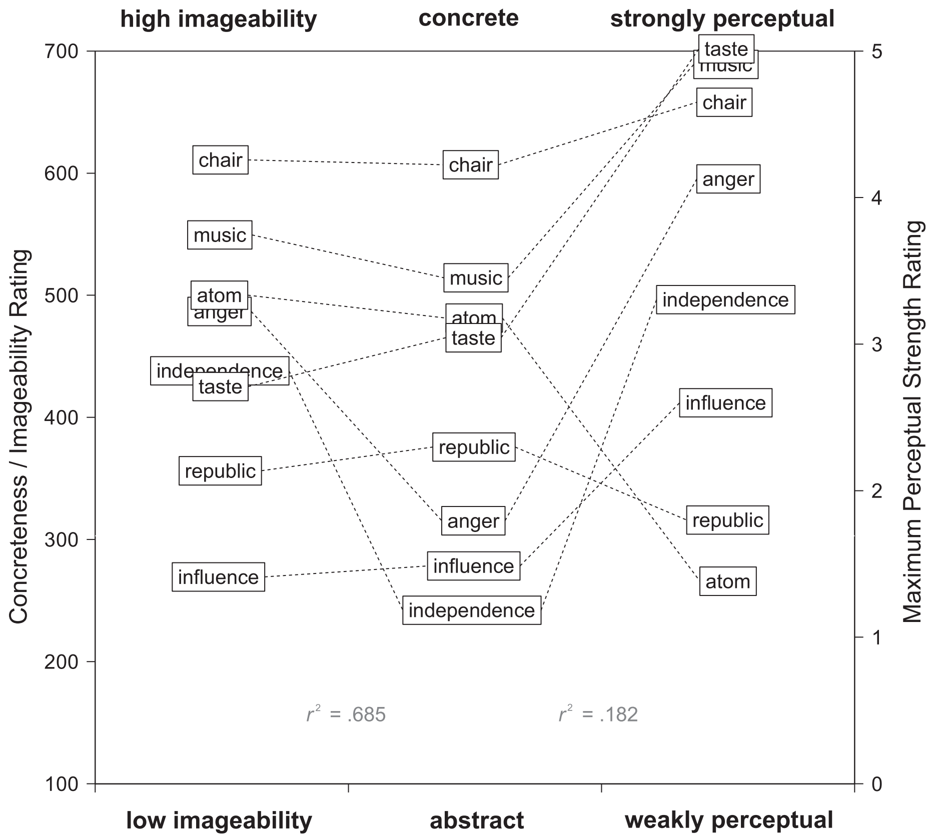

Fig. 3. Illustration of how imageability, concreteness and maximum perceptual strength ratings vary across a sample of words. R^2 values refer to correlations of concreteness with imageability or maximum perceptual strength in the dataset of Study 2 ($N = 592$).

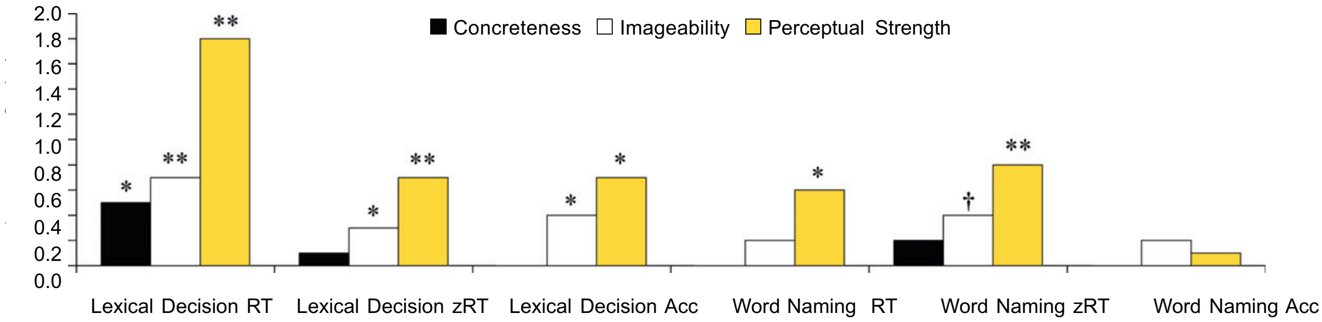

Fig. 4. Simple effects of each predictor in Study 2, showing proportion of explained variance (R^2 change in %) of Elexicon reaction time and accuracy data, over and above a basic model of contextual diversity, word length in letters, and number of syllables ($^{\dagger}p < .1$, $^{*}p < .05$, $^{**}p < .01$). Flatline bars (e.g., concreteness in naming RT) represent 0% contribution.

reports of classic concreteness effects. For example, based on the concreteness coefficient in the model for lexical decision RT, the difference between a fully abstract word of rating 100 and a fully concrete word of rating 700 would amount to approximately 28 ms, which is similar to previous factorial studies (e.g., 33 ms: Binder et al., 2005; 17 ms: Kroll & Merves, 1986, Experiment 2; 49 ms: Schwanenflugel & Shoben, 1983, Experiment 2). For comparison, the predicted RT difference between a low- and high-imageability word (ratings 100–700) was larger at 37 ms, whereas the difference between a weakly- and strongly-perceptual word (ratings 0–5) exceeded both at 59 ms. The respective differences for word naming RT followed the same trajectory, with concreteness at 5 ms, imageability at 17 ms, and perceptual strength at 28 ms.

In analysis of independent effects, only perceptual strength emerged as a unique predictor of variance (see Fig. 5). When concreteness had already been included in the model, perceptual strength still accounted for an extra proportion of variance in lexical decision raw RT (i.e., the only model in which concreteness was a useful predictor), as well as in all other measures except naming accuracy. Similarly, perceptual strength acted as an independent predictor in models that already included imageability: both where imageability had performed well as a simple predictor (lexical decision data), and where it had not

Table 5
Standardized regression coefficients in models of Elexicon reaction time and accuracy data in Study 2. Models of simple effects include a single predictor over and above the basic model, while models of independent effects include two predictors: perceptual strength with (A) concreteness or (B) imageability.

Predictor	Lexical decision RT	Lexical decision zRT	Lexical decision Acc	Word naming RT	Word naming zRT	Word naming Acc
Simple effects						
Concreteness	−0.078*	−0.039	+0.015	−0.018	−0.043	−0.005
Imageability	−0.089**	−0.060*	+0.070*	−0.050	−0.066†	+0.048
Perceptual strength	−0.138**	−0.089**	+0.084*	−0.078*	−0.094**	+0.028
Independent effects A						
Concreteness	−0.016	−0.004	−0.031	+0.025	+0.002	−0.023
Perceptual strength	−0.132**	−0.091**	+0.097**	−0.089*	−0.095*	+0.038
Independent effects B						
Imageability	−0.023	−0.018	+0.036	−0.013	−0.024	+0.045
Perceptual strength	−0.127**	−0.080*	+0.067†	−0.072†	−0.082*	+0.006

Note: RT = raw reaction times, zRT = standardized reaction times, Acc = Accuracy.
† $p < .1$.
* $p < .05$.
** $p < .01$.

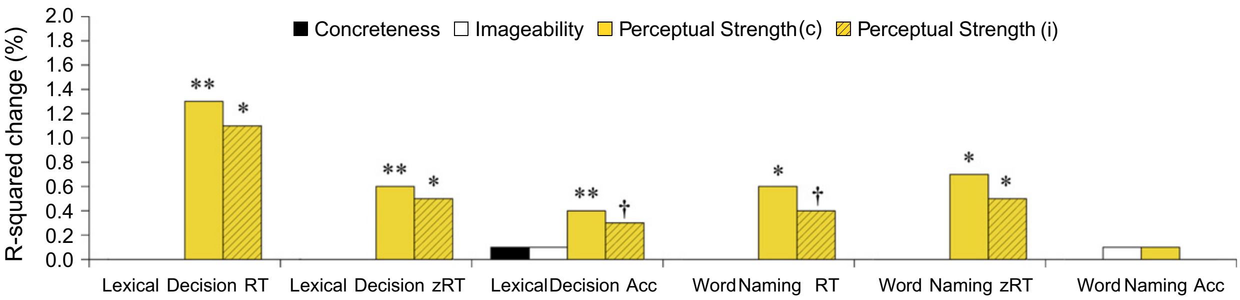

Fig. 5. Independent (unique) effects of each predictor in Study 2, showing proportion of explained variance (R^2 change in %) of Elexicon reaction time and accuracy data, over and above that of perceptual strength, concreteness (c) or imageability (i) (†$p < .1$, *$p < .05$, **$p < .01$). Flatline bars (e.g., concreteness in naming RT) represent 0% contribution.

(word naming times). Critically, the inverse was not true. When maximum perceptual strength was included as a predictor, there was no model where the addition of concreteness or imageability produced an increase in fit. Again, it is important to note that perceptual strength was in all cases acting in the expected direction (see Table 5): RT decreased and accuracy increased with higher perceptual strength. In other words, the independent predictive ability of perceptual strength never counteracted any facilitation by concreteness or imageability. Maximum perceptual strength thus captures meaningful information about conceptual structure that other ratings do not, and this information impacts directly on word processing performance.

4. General discussion

In the present paper, we show that concreteness and imageability ratings do not accurately reflect the perceptual basis of concepts, and that so-called concreteness effects in lexical decision and naming are better predicted by perceptual strength ratings than by concreteness or imageability ratings. These findings support the intuition that strongly perceptual concepts are faster to process, and show that textbook concreteness effects in word processing are actually a function of the degree of perceptual information in each referent concept's representation. However, our results also suggest that concreteness and imageability ratings are often unsuitable for the tasks in which they are employed, because most of their variance comes from non-perceptual decision criteria that is unrelated to word processing performance. Concreteness effects could therefore be better characterized as perceptibility effects, which can be sometimes nullified or inverted (e.g., Kousta et al., 2011; Papagno et al., 2009) when elicited from relatively noisy concreteness or imageability ratings.

While the connection between concreteness effects and perceptual information might at first glance seem like old news (e.g., Barsalou & Wiemer-Hastings, 2005; Paivio, 1971, 1986, 2007; Vigliocco et al., 2009; Wiemer-Hastings & Xu, 2005), the present findings have some important ramifications for how such effects should be interpreted. Concreteness effects, by their very name, are assumed to result from an ontological difference between concrete and abstract concepts carrying through to a representational difference that affects speed and accuracy of processing. Labeling a word as "concrete" or "abstract" has an intuitive appeal, but we would argue that these terms lacked proper operationalization during norming and hence it is unclear exactly what information is captured by concreteness ratings. Of course, any set of ratings can only ever be an approximation of an underlying representation, and we are not suggesting that one should expect a perfect fit between concreteness ratings and behavioral effects. That said, the poor performance of concreteness ratings in the current data lies in sharp contrast to the robust performance of perceptual strength ratings. We suggest that the concrete / abstract ontological distinction must be disentangled from concreteness and imageability norms because empirical concreteness effects are not in themselves well predicted by such ratings.

Furthermore, perceptual strength is not just an incremental improvement over concreteness or imageability ratings in predicting word processing performance. An incrementally-refining predictor would correlate well with its alternative predictors and display the same pattern of relationships with other variables. In that sense, imageability is an incremental improvement over concreteness ratings: it correlates well (over two-thirds shared variance in our dataset and others), and shows the same pattern of relationships with other independent variables (although strength of relationships vary, their direction remains the same), while ultimately being a better predictor of word processing performance. However, perceptual strength does neither of these things. In contrast to the strong relationship between concreteness and imageability (68% shared variance), perceptual strength correlates relatively poorly with both (18–25% shared variance), and shows the opposite relationship with contextual diversity (contextual diversity increases with perceptual strength, but decreases with concreteness and imageability), yet perceptual strength outperforms both these variables in predicting lexical decision and naming performance. The upshot of this directional flip is that concreteness and imageability are effectively fighting against the powerful facilitatory effect of contextual diversity while perceptual strength has an additive effect.

4.1. What do ratings reflect?

Study 1 showed that the concreteness rating scale appears to be based on two different, intersecting sets of non-perceptual decision criteria. While it is therefore unclear exactly what information is captured by concreteness ratings, we suggest two possibilities of what they might actually reflect. First, concreteness norms may reflect people's judgement of whether a concept is inherently "abstract" or "concrete" according to layperson definitions. In this view, words like *dog* and *chair* are regarded as inherently concrete because they refer to entities with solid mass; something that has an objective existence in a particular place at a particular time (e.g., "existing in a material or physical form; real or solid": New Oxford American Dictionary, 2009). On the other hand, words like *failure* and *democracy* are regarded as inherently abstract because of their subjective or uncertain nature; something that lacks objective existence (e.g., "existing in thought or as an idea but not having a physical or concrete existence": New Oxford American Dictionary, 2009). Such judgements resemble psychologically essentialist beliefs in the concreteness or abstractness of an entity (Medin & Ortony, 1989): *dog* and *chair* are concrete because they share some ineluctable essence of "concreteness", *failure* and *democracy* share an essence of "abstractness", and *substitute* is in the middle of the scale because it partially has elements of both. Alternatively, our second possibility is that concreteness ratings may reflect the typicality of membership of the category specified as concrete in norming instructions: objects, materials or persons.[4] Words like *dog*, *cotton* and *lady* are regarded as concrete because they are all fairly typical of the target category. In contrast, *failure* and *democracy* are regarded as abstract because they are atypical members, while *substitute* or *kingdom* lie in the middle of the scale. Both of these possible judgement criteria may have been employed by participants in concreteness norming studies as people attempted to create their own ad hoc operationalisation of the terms "abstract" and "concrete".

Imageability ratings, at least, appear more straightforward in that there is no reason to think that they reflect anything other than the ease of consciously generating imagery. However, Study 1 showed that people do not find it equally easy to generate imagery across the range of modalities that constitute perceptual experience, and that imageability ratings are visually biased. We can think of two reasons why people may over-rely on vision when rating imageability at the expense of other modalities. First is expertise: people have a lot of practice in imagining what things look like, so that the mind's eye gets a lot more exercise than the mind's ear, nose, tongue or skin. It is difficult to consolidate the differential ease of multimodal perceptual imagery into a single rating, and so imageability ratings predominantly reflect ease of generating imagery in the dominant visual modality. A second reason, which may operate in addition to the first, is that the use of the word "image" was always going to bias people towards the visual modality, regardless of experimental instructions to consider other perceptual modalities. To a naïve participant, images are things that can be seen. Since many words refer to objects that can be visually experienced, this approach works well much of the time. However, when it comes to words with little visual grounding such as *advice* (auditory-dominant) or *mild* (gustatory-dominant), people accord them low imageability ratings even though they are very strongly perceptual concepts.

The norming methodology for perceptual strength (Lynott & Connell, 2009, in press) asked people to consider their perceptual experience of the concepts in question (i.e., a direct, operationalized definition of the perceptual basis of a concept) for each of five separate modalities (i.e., not requiring participants to aggregate multimodal experience into one rating). These norms have previously been used in studies that show modality-specific perceptual simulations are involved in conceptual processing, such as eliciting modality-switching costs in property verification (Collins et al., 2011; Hald et al., 2011; Lynott & Connell, 2009) and conceptual combination (Connell & Lynott, 2011) tasks, or replicating the perceptual phenomenon whereby tactile stimuli are slower to process than stimuli from other modalities (Connell & Lynott, 2010). Consequently, we believe that perceptual strength ratings

[4] Norms from these category-based instructions include those of Clark and Paivio (2004), Gilhooly and Logie (1980), Nelson et al. (2004), Paivio et al. (1968), and Toglia and Battig (1978).

reflect the extent to which a particular perceptual modality is involved during experience of the referent concept, and hence the extent to which perceptual information is then involved in simulating that concept. In order to rate the strength of modality-specific experience for a given word, people generated a multimodal situated simulation of that concept and then judged the strength of each modality in their simulation. People do not seem to have been influenced by lexical relationships between the word to be rated and the modality words mentioned in the ratings scales (i.e., hearing, tasting, smelling, seeing, feeling through touch). For example, even synaesthetic metaphoric expressions like "taste of success" did not lead people to regard *success* as tastable[5] (gustatory rating 0.65 out of 5), nor did "smell of fear" lead *fear* to be considered smellable (olfactory rating 0.82). Furthermore, even though one can "feel pride", participants still granted *pride* a low haptic rating (0.71) even though the words co-occur frequently. Rather than basing their response on shallow linguistic associations (see also Barsalou et al., 2008; Louwerse & Connell, 2011), it appears that people generated perceptual strength ratings on the basis of deeper perceptual simulations.

As in all cases of single-word presentation, most words in English are lexically ambiguous (e.g., Church, 1988; Rodd, Gaskell, & Marslen-Wilson, 2004), and have multiple possible representations and therefore multiple possible profiles of perceptual strength. For example, the object property *plain* is visual in a simulation of fabric, but gustatory in a simulation of food (van Dantzig, Cowell, Zeelenberg, & Pecher, 2011). When rating perceptual strength, two potential strategies were open to participants for such words: either to simulate and rate only one situation, or (with awareness of the ambiguity) simulate both and attempt to consider both while rating. In the interest of cognitive parsimony, we would assume that the first approach is most common, but the second approach is likely for some participants, with similar behaviors occurring in lexical decision and naming tasks.

4.2. Impact on existing studies

Most reports of concreteness effects in the word processing literature are based on materials selected with concreteness or imageability ratings (e.g., Binder et al., 2005; Bleasdale, 1987; de Groot, 1989; Howell & Bryden, 1987; Kroll & Merves, 1986; Schwanenflugel & Stowe, 1989; Schwanenflugel et al., 1988). We would argue that it is likely such studies represent cases where the abstract versus concrete materials also happen to reflect differences in perceptual strength, particularly when item sets are refined by experimenter intuition (e.g., rejecting *noisy* as an unsuitable abstract word, despite its low concreteness or imageability rating). In contrast, where studies show null or inverted concreteness effects in word processing performance (e.g., Adelman et al., 2006; Barca et al., 2002; Fiebach & Friederici, 2004; Kousta et al., 2011; Papagno et al., 2009; Sabsevitz et al., 2005; Tsai et al., 2009; Tyler et al., 2001), it is likely that the ratings for those particular item sets simply do not reflect the perceptual basis of the referent concepts. Even when groups of words differ in concreteness or imageability ratings, there may be little perceptual difference between the groups, or perceptual differences may even lie in the opposite direction (e.g., Table 1). Future work could explicitly test our proposal that such mixed effects in the concreteness literature are largely driven by perceptual strength. Furthermore, our focus in this paper was on concreteness effects in online word processing, rather than in recall and recognition tasks (Allen & Hulme, 2006; Fliessbach et al., 2006; Paivio et al., 1966; Romani et al., 2007; Walker & Hulme, 1999). We aim to investigate whether the perceptibility effects we found for lexical decision and naming also extend to memory performance, and whether concreteness or imageability ratings can predict an independent effect in memory if not in word processing.

While the present work focuses on the somewhat fragile nature of concreteness effects in behavioral data, results have been similarly equivocal in neuroimaging. Some studies have reported that processing concrete words led to more bilaterally distributed activation than abstract words (e.g., Binder et al., 2005; Sabsevitz et al., 2005), but others found the reverse (Pexman et al., 2007). Still others have found that abstract concepts tend towards a left-hemisphere processing pathway (Binder et al., 2005; Fiebach & Friederici, 2004), while others have argued for right-hemisphere dominance (Kiehl et al., 1999). Kousta et al. (2011) speculate that the lack of consistency in the imaging literature may be due to differing conceptual content in the items used, "which quite reasonably leads to activation of different brain networks in different studies or to lack of consistent areas of activation within the same study" (p. 16). Given that ostensibly abstract or concrete concepts can vary considerably in perceptual information (Study 1), and that rated concreteness or imageability is not a particularly informative predictor of word processing latency or accuracy (Study 2), it is uncertain what existing imaging studies of these tasks have actually localized. More research is needed to identify the neural correlates of perceptibility – rather than rated-concreteness – effects.

4.3. Theoretical accounts of concreteness effects

Theoretically, the present results poses some problems for dual coding, context availability, and situated simulation explanations of concreteness effects. It is a central tenet of dual coding theory that highly perceptual concepts are those with the most direct connections between the verbal and nonverbal imagery codes, and people therefore find it difficult to generate perceptual imagery for words that lack these direct connections (Paivio, 1986, 2007). However, imageability (i.e., the ease of consciously generating imagery) is not well related to perceptual experience (Study 1), and its effects were entirely subsumed by larger effects of perceptual strength (Study 2). In other words, it is the extent of perceptual information in a concept's representation that matters to word processing, not the ease of generating imagery, which casts some doubt on the idea that processing delays for abstract concepts emerge from their lack of direct inter-system connections.

[5] We thank Ken McRae for this idea.

Both context availability (Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1988) and situated simulation (Barsalou & Wiemer-Hastings, 2005; Barsalou et al., 2008) views share the idea that abstract words are slowed in processing because they have relatively weak connections to potential situational contexts, which makes it difficult for people to represent the concept. In explaining why abstract words come to have such weak contextual connections, Schwanenflugel (1991; Schwanenflugel & Shoben, 1983) puts forward the reason that abstract words appear in a greater diversity of contexts than do concrete words. This effect has been demonstrated several times both when contextual diversity was assessed by subjective ratings (i.e., asking people whether a word appears in a limited or great number of contexts: Galbraith & Underwood, 1973; Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1988), and when contextual diversity was calculated objectively using large-scale corpora (i.e., as the count or proportion of document-level contexts in which a word appears: Adelman et al., 2006; Pexman et al., 2008). In their situated simulation account, Barsalou and Wiemer-Hastings (2005) support the idea that contextual diversity may underlie context availability, and therefore concreteness effects, because the large number of potential contexts for abstract words compete and interfere with one other when the word is accessed, and thereby slow down representation of the concept and processing of the word. This idea, however, is not borne out by our data. Strongly perceptual concepts (i.e., those that are generally assumed to be concrete, regardless of what concreteness ratings say) actually have greater contextual diversity than weakly perceptual concepts (Study 2). Furthermore, and contrary to the above theoretical assumptions, greater contextual diversity actually speeds up processing rather than slowing it down (Study 2; see also Adelman et al., 2006; Brysbaert & New, 2009), meaning that contextual diversity is not a tenable underlying reason for context availability to produce concreteness effects. It therefore seems that processing delays for abstract (i.e., weakly perceptual) concepts are unrelated to the diversity or otherwise of their situational contexts. It is currently unclear whether or how such processing delays relate to context availability when perceptual strength ratings are utilized rather than concreteness or imageability ratings.

4.4. What affects word processing performance?

But why should perceptual strength facilitate processing? One possibility is that, since the situated simulation view shares commonalities with both dual coding and contextual availability theories, it can be adapted to explain our effects (and those in the wider literature). In the same way that *chairs*, *tendencies*, and *anger* are encountered as part of broader, situated experience, the representations that people create during conceptual processing of these words are also situated. Concepts are not represented in isolation even when their words are presented in isolation, which means that, even in relatively shallow tasks like lexical decision and naming, background situational information is activated and incorporated into the representations of both strongly- and weakly-perceptual concepts (Connell & Lynott, submitted for publication; see also Lynott & Connell, 2010). Strongly perceptual concepts such as *chair*, *music*, or *crimson* can be represented quickly because most of their conceptual content is a relatively simple and discrete package of perceptual information, and hence is easy to simulate. Since lexical decision tasks that require visual attention can facilitate simulation of visual information, and naming tasks that require auditory attention can facilitate simulation of auditory information (Connell & Lynott, in preparation), it may be the case that any type of language processing has the ability to facilitate representation of perceptual concepts.

Weakly perceptual concepts, on the other hand, tend to take longer to represent because they lack a neat package of perceptual information that can benefit from modality attention effects, and because much of their non-perceptual conceptual content involves pulling in other concepts as part of their broader situation (e.g., a *tendency* to do what? A *republic* of where?). This latter idea is closely related to Barsalou and Wiemer-Hastings' (2005) hypothesis that abstract concepts are more complex and broadly-focused than concrete concepts. However, it should be noted that one of the three abstract concepts that Barsalou and Wiemer-Hastings coded for conceptual content – *invention* – is actually strongly perceptual in our norms. Furthermore, none of the concepts in our perceptual strength norms were devoid of perceptual information. The perceptually-weakest word in the present item set is *atom*, which nonetheless attracted a visual strength rating of 1.38 on a 0–5 scale. Other more traditionally abstract words, like *republic* or *factor*, scored even higher. In other words, even when weakly-perceptual concepts require a complex situational context as part of their representation, they still have a definite perceptual basis. Perceptual strength and situational complexity are therefore not mutually exclusive characteristics of conceptual representations, but rather we assume a probabilistic relationship between the two such that strongly perceptual concepts are less likely to draw in complex situational contexts than weakly perceptual concepts. In addition, because strongly perceptual words appear in a wider variety of contexts than do weakly perceptual words, the tendency towards situational complexity is unlikely to depend on the increased contextual diversity assumed for abstract concepts by Schwanenflugel (1991; Schwanenflugel & Shoben, 1983) and discussed by Barsalou and Wiemer-Hastings. Strongly perceptual concepts are likely to be found both in a wide variety of contexts and in relatively simple situations, but the exact relationship between contextual diversity and situational complexity has yet to be fully investigated.

One issue we have not addressed in the present study is that of emotional affect. Kousta et al. (2011) show that any effects of imageability or concreteness on lexical decision times disappear once the affective valence of words is taken into account. Kousta et al.'s contention (see also Vigliocco et al., 2009) was that abstract concepts tend to have more affective associations than concrete concepts, which is consistent with the situated simulation view that abstract concepts tend to focus on introspective situational context more than concrete concepts. However, the interactions that Kousta and colleagues report are complex: words with affective associations were processed faster than neutral words, which overall lent abstract words an *advantage* over concrete words, meaning the precise mechanisms by which concreteness effects emerge are unclear. It remains to be seen what sort of influence affective valence and/or arousal will exert on word processing times when perceptual strength (rather than concreteness or imageability) is taken into account.

5. Conclusion

In sum, we believe that the operationalisation of abstract and concrete concepts deserves much closer scrutiny than it has received to date. Whether researchers want to investigate the ontological distinction between abstract and concrete concepts, or the variables that affect latency and accuracy in word processing, then they should reconsider the automatic tendency to reach for concreteness and imageability ratings that have little to do with the perceptual basis of concepts. Strength of perceptual experience has a powerful bearing on how people represent concepts during word processing, and these perceptibility effects are stronger than those elicited by concreteness or imageability.

Acknowledgement

The order of authorship is arbitrary. Some of this work was supported by the UK Economic and Social Research Council [Grant Number RES-000-22-3248].

References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17*, 814–823.

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64–88.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445–459.

Barca, L., Burani, C., & Arduino, L. S. (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods, Instruments, & Computers, 34*, 424–434.

Barsalou, L. W., Santos, A., Simmons, W. K., & Wilson, C. D. (2008). Language and simulation in conceptual processing. In M. De Vega, A. M. Glenberg, & A. C. Graesser, A. (Eds.), *Symbols, embodiment, and meaning* (pp. 245–283). Oxford, UK: Oxford University Press.

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

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract words. *Journal of Cognitive Neuroscience, 17*, 905–917.

Bleasdale, F. A. (1987). Concreteness-dependent associative priming: Separate lexical organization for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 582–594.

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology, 11*, 617–660.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41*, 977–990.

Church, K. W. (1988). A stochastic parts program and noun phrase parser for unrestricted text. In *Proceedings of the second conference on applied natural language processing* (pp. 136–143). Austin, TX: Association for Computational Linguistics.

Clark, J. M., & Paivio, A. (2004). Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods, 36*, 371–383.

Collins, J., Pecher, D., Zeelenberg, R., & Coulson, S. (2011). Modality switching in a property verification task: An ERP study of what happens when candles flicker after high heels click. *Frontiers in Psychology, 2*(10), 1–10.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Connell, L., & Lynott, D. (in preparation). I see what you mean: Semantic activation depends on perceptual attention.

Connell, L., & Lynott, D. (submitted for publication). Principles of representation: Why you can't represent the same concept twice.

Connell, L., & Lynott, D. (2010). Look but don't touch: Tactile disadvantage in processing modality-specific words. *Cognition, 115*, 1–9.

Connell, L., & Lynott, D. (2011). Modality switching costs emerge in concept creation as well as retrieval. *Cognitive Science, 35*, 763–778.

de Groot, A. M. B. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 824–845.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777–799.

Fiebach, C. J., & Friederici, A. D. (2004). Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia, 42*, 62–70.

Fliessbach, K., Weis, S., Klaver, P., Elger, C. E., & Weber, B. (2006). The effect of word concreteness on recognition memory. *Neuroimage, 32*, 1413–1421.

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition, 1*, 56–60.

Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behavior Research Methods and Instrumentation, 12*, 395–427.

Hald, L. A., Marshall, J., Janssen, D. P., & Garnham, A. (2011). Switching modalities in a sentence verification task: ERP evidence for embodied language processing. *Frontiers in Psychology, 2*(45), 1–15.

Heidbreder, E. (1945). Toward a dynamic psychology of cognition. *Psychological Review, 52*, l–22.

Howell, J. R., & Bryden, M. P. (1987). The effects of word orientation and imageability on visual half-field presentations with a lexical decision task. *Neuropsychologia, 25*, 527–538.

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin, 136*, 302–329.

James, C. T. (1975). The role of semantic information in lexical decisions. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 130–136.

Kiehl, K. A., Liddle, P. F., Smith, A. M., Mendrek, A., Forster, B. B., & Hare, R. D. (1999). Neural pathways involved in the processing of concrete and abstract words. *Human Brain Mapping, 7*, 225–233.

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*, 14–34.

Kroll, J., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 92–107.

Louwerse, M. M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science, 35*, 381–398.

Lynott, D., & Connell, L. (in press). Modality exclusivity norms for 400 nouns: The relationship between perceptual experience and surface word form. *Behavior Research Methods*.

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

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology, 1*(212), 1–14.

Macoir, J. (2008). Is a plum a memory problem? Longitudinal study of the reversal of concreteness effect in a patient with semantic dementia. *Neuropsychologia, 47*, 518–535.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–195). Cambridge, UK: Cambridge University Press.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*, 402–407.

Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language, 31*, 237–260.

New Oxford American Dictionary application (Version 2.1.2) [computer software]. Cupertino, CA: Apple Inc.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

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

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76*(1, Pt. 2), 1–25.

Paivio, A., Yuille, J. C., & Smythe, P. C. (1966). Stimulus and response abstractness, imagery, and meaningfulness, and reported mediators in paired-associate learning. *Canadian Journal of Psychology, 20*, 362–377.

Papagno, C., Fogliata, A., Catricalà, E., & Miniussi, C. (2009). The lexical processing of abstract and concrete nouns. *Brain Research, 1263*, 78–86.

Pexman, P. M., Hargreaves, I. S., Edwards, J. D., Henry, L. C., & Goodyear, B. G. (2007). Neural correlates of concreteness in semantic categorization. *Journal of Cognitive Neuroscience, 19*, 1407–1419.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. F., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review, 15*, 161–167.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child*. New York: Basic Books.

Reed, H. B., & Dick, R. D. (1968). The learning and generalization of abstract and concrete concepts. *Journal of Verbal Learning and Verbal Behavior, 7*, 486–490.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*, 89–104.

Romani, C., McAlpine, S., & Martin, R. C. (2007). Concreteness effects in different tasks: Implications for models of short-term memory. *Quarterly Journal of Experimental Psychology, 61*, 292–323.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Rubin, D. C. (1980). 51 Properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior, 19*, 736–755.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage, 27*, 188–200.

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*, 499–520.

Schwanenflugel, P. J. (1991). Why are abstract concepts hard to understand? In P. J. Schwanenflugel (Ed.), *The psychology of word meanings* (pp. 223–250). Hillsdale, NJ: Erlbaum.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 82–102.

Schwanenflugel, P. J., & Stowe, R. W. (1989). Context availability and the processing of abstract and concrete words in sentences. *Reading Research Quarterly, 24*, 114–126.

Spreen, O., & Schultz, R. W. (1966). Parameters of abstraction, meaningfulness, and pronunciability for 329 nouns. *Journal of Verbal Learning and Verbal Behavior, 5*, 459–468.

Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Hillsdale, NJ: Erlbaum.

Tsai, P. S., Yu, B. H. Y., Lee, C. Y., Tzeng, O. J. L., Hung, D. L., & Wu, D. H. (2009). An event- related potential study of the concreteness effect between Chinese nouns and verbs. *Brain Research, 1253*, 149–160.

Tyler, L. K., Russell, R., Fadili, J., & Moss, H. E. (2001). The neural representation of nouns and verbs: PET studies. *Brain, 124*, 1619–1634.

van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: Norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods, 43*, 145–154.

Vigliocco, G., Meteyard, L., Andrews, M., & Kousta, S. (2009). Toward a theory of semantic representation. *Language and Cognition, 1*, 219–248.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1256–1271.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain, 107*, 829–854.

Whaley, C. P. (1978). Word-nonword classification times. *Journal of Verbal Learning and Verbal Behavior, 17*, 143–154.

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science, 29*, 719–736.

Wilson, M. D. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.0. *Behavior Research Methods, Instrumentation, and Computers, 20*, 6–10.

Yi, H. A., Moore, P., & Grossman, M. (2007). Reversal of the concreteness effect for verbs in patients with semantic dementia. *Neuropsychology, 21*, 9–19.