COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–27

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12348

Abstract Conceptual Feature Ratings Predict Gaze Within Written Word Arrays: Evidence From a Visual Wor(l)d Paradigm

Silvia Primativo,[a,b] Jamie Reilly,[c,d] Sebastian J. Crutch[a]

[a]*Dementia Research Centre, Department of Neurodegenerative Disease, University College London, Institute of Neurology*
[b]*Neuropsychology Unit, IRCCS Fondazione Santa Lucia*
[c]*Eleanor M. Saffran Center for Cognitive Neuroscience, Temple University*
[d]*Department of Communication Sciences and Disorders, Temple University*

Received 6 October 2014; received in revised form 30 October 2015; accepted 16 November 2015

Abstract

The Abstract Conceptual Feature (ACF) framework predicts that word meaning is represented within a high-dimensional semantic space bounded by weighted contributions of perceptual, affective, and encyclopedic information. The ACF, like latent semantic analysis, is amenable to distance metrics between any two words. We applied predictions of the ACF framework to abstract words using eyetracking via an adaptation of the classical "visual word paradigm" (VWP). Healthy adults (n = 20) selected the lexical item most related to a probe word in a 4-item written word array comprising the target and three distractors. The relation between the probe and each of the four words was determined using the semantic distance metrics derived from ACF ratings. Eye movement data indicated that the word that was most semantically related to the probe received more and longer fixations relative to distractors. Importantly, in sets where participants did not provide an overt behavioral response, the fixation rates were nonetheless significantly higher for targets than distractors, closely resembling trials where an expected response was given. Furthermore, ACF ratings which are based on individual words predicted eye fixation metrics of probe-target similarity at least as well as latent semantic analysis ratings which are based on word co-occurrence. The results provide further validation of Euclidean distance metrics derived from ACF ratings as a measure of one facet of the semantic relatedness of abstract words and suggest that they represent a reasonable approximation of the organization of abstract conceptual space. The data are also compatible with the broad notion that multiple sources of information (not restricted to sensorimotor and emotion information) shape the organization of abstract concepts. While the adapted "VWP" is potentially a more metacognitive task than the classical visual world paradigm, we argue that it offers potential utility for studying abstract word comprehension.

Correspondence should be sent to Silvia Primativo, Dementia Research Centre, UCL Institute of Neurology, Box 16, National Hospital, Queen Square, London WC1N 3BG, UK. E-mail: s.primativo@ucl.ac.uk

Keywords: Abstract knowledge; ACF ratings; LSA cosines; Visual world paradigm; Eye movements

1. Introduction

The nature and organization of abstract concepts have received relatively little attention compared with their more concrete counterparts, although abstract and concrete words occur with comparable frequencies in the English lexicon (Reilly, 2005; Reilly & Kean, 2007). In the neuropsychological literature, studies of abstract and concrete word processing include many reports of double dissociations between patients showing relative preservation of concrete concepts (e.g., Coltheart, Patterson, & Marshall, 1987; Franklin, Howard, & Patterson, 1995; Katz & Goodglass, 1990; Martin & Saffran, 1992; Roeltgen, Sevush, & Heilman, 1983; Warrington, 1975) or abstract concepts (e.g., Breedin, Saffran, & Coslett, 1994; Cipolotti & Warrington, 1995; Marshall, Pring, Chiat, & Robson, 1996; Papagno, Capasso, Zerboni, & Miceli, 2007; Reilly, Peelle, & Grossman, 2007; Sirigu, Duhamel, & Poncet, 1991; Warrington, 1975; Warrington & Shallice, 1984; although see Hoffman, Jones, & Lambon Ralph, 2013a). In the functional neuro-imaging literature, the most common finding is that relative to concrete words, abstract words elicit more extensive activation within left hemisphere brain regions such as inferior frontal gyrus (LIFG; Binder, Westbury, McKiernan, Possing, & Medler, 2005; Fiebach & Friederici, 2004; Jessen et al., 2000; Noppeney & Price, 2004; Perani et al., 1999) and the superior temporolateral cortex (Binder, Desai, Graves, & Conant, 2009; Binder et al., 2005; Kiehl et al., 1999; Mellet, Tzourio, Denis, & Mazoyer, 1998; Wise et al., 2000). In contrast, concrete words tend to evoke more extensive bilateral activation (Binder et al., 2005; Fiebach & Friederici, 2004; Grossman et al., 2002; Sabsevitz, Medler, Seidenberg, & Binder, 2005).

Numerous theoretical models have emerged to account for the word concreteness effect, a collective advantage manifested by concrete relative to abstract words across a wide range of cognitive domains. One dominant approach specifies quantitative differences in the amount of featural support available for concrete relative to abstract words. Dual coding theory, (Paivio, 1971, 2013), for example, holds that concrete words are represented in both a verbal and an imagistic format, while abstract words are represented solely through a verbal code. Plaut and Shallice (1993) similarly argued that concreteness effects emerge at least in part due to a more extensive network of features for concrete words (Plaut & Shallice, 1993). The context availability hypothesis reflects an alternative account emphasizing discrepancies in the capacity of concrete and abstract words to rapidly engage event schemas (Schwanenflugel, 1991; Schwanenflugel & Shoben,1983). That is, concrete words more rapidly and effectively evoke a network of related contextual information (e.g., picnic evokes sun, sandwiches, ants, grass, lemonade, etc.). A different approach, known as the *semantic diversity hypothesis* (Hoffman, Lambon Ralph, & Rogers, 2013b), highlights how abstract word meaning is more variable, subjective, and nuanced relative to concrete words that tend to more often denote fixed entities.

Straddling these positions, the qualitatively different representational (QDR) framework hypothesis suggests that featural and contextual information are important for both types of words but that their relative contribution differs along the concreteness continuum: information about semantic similarity (largely based on concept features) is relatively more important in the organization of concrete concepts, while information about semantic association (largely based on context) has disproportionate weighting in the organization of abstract concepts (Crutch & Jackson, 2011; Crutch & Warrington, 2005; Duñabeitia, Avilés, Afonso, Scheepers, & Carreiras, 2009).

Despite the predominance of semantic association over similarity engendered for abstract words within the QDR hypothesis, many questions remain over how semantic similarity between abstract words is represented and, more broadly, what actually constitutes an abstract word. After all, the great majority of studies and theories on abstract words have considered their relationship with concrete concepts, rather than studying their development and processing in their own right. Mirman and Magnuson (2008) investigated the effect of semantic neighborhood density (SND) on semantic representation. The authors explored facilitative and inhibitory effects played by neighbors and suggested that different dynamics are likely to emerge when taking into account neighbors' number and distance. Results suggested that semantic processing is slower for dense near neighborhoods and faster for dense but distant neighborhoods (Mirman & Magnuson, 2008).

Andrews, Vigliocco, and Vinson (2009) proposed that semantic representation should be conceived as the combination of two types of information: experiential data (derived by the interaction with the physical world) and distributional data (which represent the statistical distribution of words in the language). A further step forward in the understanding of semantic knowledge representation has been the importance attributed to affective information, along with experiential and linguistic information (Andrews et al., 2009; Kousta, Vinson, & Vigliocco, 2009; Kousta, Vigliocco, Vinson, Andrews, & Del Campo, 2011; Newcombe, Campbell, Siakaluk, & Pexman, 2012; see Pecher, Boot, & Van Dantzig, 2011 for a review). As with the QDR model, the distinction between abstract and concrete stimuli is relative and continuous rather than absolute and discrete. It is claimed that both concrete and abstract words convey many different types of information (e.g., sensory-motor, affective, and linguistic), but they differ in terms of which kind of information has the greatest weight: concrete words are characterized by a statistical preponderance of sensorimotor information, while abstract words are statistically marked by affective and linguistic information (Kousta et al., 2009).

Two main reasons are provided to explain why emotion information might have privileged status in the development and processing of abstract words. First, emotion words largely refer to abstract, introspective states. Second, emotion development tends to precede linguistic development and may constitute a crucial prerequisite or rate limiting step in the development of abstract semantic representations (Bloom, 1998). For example, the acquisition of abstract words such as melancholy or malaise requires shades of experience with the more fundamental human emotion of sadness. Yet, it is also clear that many other cognitive systems also demonstrate development prior to language acquisition.

Thus, emotion may be only one of a number of latent factors driving the representation of abstract concepts.

Recently we have introduced a new approach to studying the semantic attributes of single abstract words, referred to as the abstract conceptual feature (ACF) rating method (Crutch, Troche, Reilly, & Ridgway, 2013; Crutch, Williams, Ridgway, & Borgenicht, 2012; Troche, Crutch, & Reilly, 2014). This involves requesting healthy individuals to provide Likert scale ratings of how much different types of cognitive information contribute to their understanding of individual concepts. The premise underlying this approach is that mapping abstract semantic space requires the identification and quantification of not only sensorimotor and emotional information, but also of a broader set of cognitive domains. A further motivation was to measure the content or semantic attributes of abstract words based on different cognitive dimensions in a manner comparable to the feature listing approach used to study the structure of concrete conceptual knowledge (e.g., Cree & McRae, 2003; Garrard, Lambon Ralph, Hodges, & Patterson, 2001). One advantage of this method relative to word co-occurrence metrics like latent semantic analysis (LSA; Landauer & Dumais, 1997) is that all ratings are gathered on individual words, not word pairs, thus eliciting data from which more flexible, context-independent semantic similarity metrics can be derived.

The cognitive dimensions taken into account were: sensation, ease of teaching, ease of modifying, action, time, emotion, morality, polarity, social interaction, thought, space, and quantity. *Sensation, action, emotion* were included owing to their inclusion in weak embodiment models of Vigliocco et al. (Andrews et al., 2009; Kousta et al., 2009, 2011). *Social interaction* and *thought* were included following work by Borghi and Cimatti (2012) and Barsalou (1999) in which the contributions of social interaction and introspection on abstract word acquisition and representation are emphasized. Similarly, Allman and Meck (2012) highlighted the importance of the *time*, whereas Lakoff and Johnson (2008) and Zwaan and Yaxley (2003) describing the organization of geographic concepts, denote the importance of considering *spatial information* as a key attribute of abstract words. The distinction between numerical and non-numerical abstract concepts (Gathercole, 1985) leads to considering *quantity* among the important abstract word's features. *Ease of teaching* refers to the learning style and/or age of acquisition, while *ease of modifying* represents an index of the contextual availability of a word. *Polarity* refers to the positive or negative feelings associated with a word. Finally, *morality* is thought to reflect the cognitive emotions associated with social mores, driver of many human behaviors (Moll, Zahn, de Oliveira-Souza, Krueger, & Grafman, 2005).

These ACF ratings have been used to generate a high-dimensional representation of abstract conceptual space, from which Euclidean distance metrics (how far apart are two abstract concepts within this space) have been derived. In previous studies, these semantic distance metrics have been shown to predict at least as well as LSA the performance of a globally aphasic stroke patient on two verbal comprehension tests (Crutch et al., 2013), suggesting they represent a relevant measure of at least one aspect of semantic relatedness. ACF metrics have also been used to help explain a pattern of superior antonym than synonym comprehension performance in three aphasic patients (Crutch et al., 2012).

Fig. 1 represents a contour plot depicting how abstract and concrete words cluster within the multidimensional semantic space bounded by a series of factors representing emotion, magnitude, and perceptual salience.

In the present paper we examined whether the ACF distance metrics could predict healthy individuals' eye movements when examining the semantic relationship between written words. For this purpose we used an adapted version of the well-established visual *world* paradigm (VWP). Originally Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) developed the methodology to systematically investigate the temporal relationship between eye movements and language processing by verbally presenting sentences with different level of syntactic complexity to participants and recording their visual exploration of a screen containing multiple objects. Since then, many different studies (Cree & McRae, 2003; Huettig & Altmann, 2005; Myung, Blumstein, & Sedivy, 2006; Yee & Sedivy, 2006) provided evidence in favor not only of the reliability of the paradigm but also of the presence of a semantic relatedness rule that mediates between language and eye movements (Altmann, 2011). Moreover, Farris-Trimble and McMurray (2013) have shown that the eye tracking data obtained using the VWP provide a stable measure of individual performance (for a review about the using of VWP for studying language processing, see Huettig, Rommers, & Meyer, 2011). Recently the same paradigm has been used by Duñabeitia et al. (2009) to investigate the differences in representation between abstract and concrete words. In their pictorial visual world paradigm, participants were presented with an abstract or a concrete word target included in a verbally presented sen-

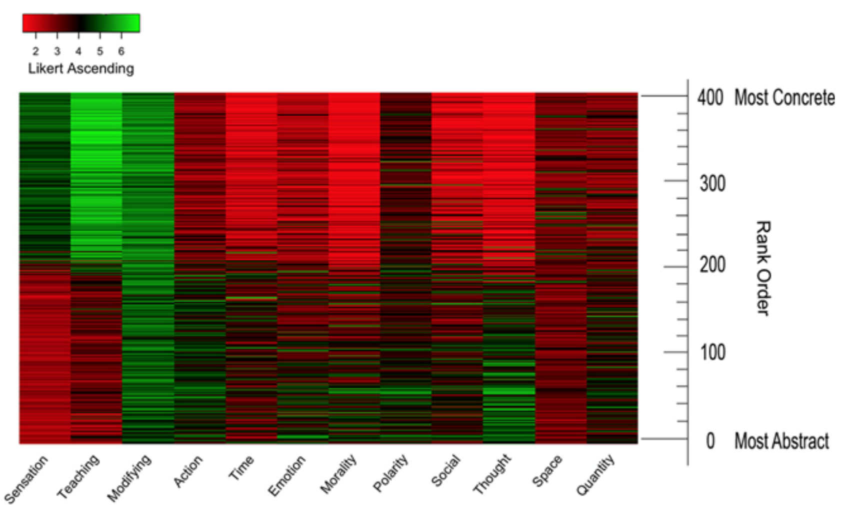

Fig. 1. The topography of abstract and concrete nouns across twelve cognitive dimensions. Each row in the heat map matrix above represents one English noun. The 400 rows are rank ordered on the *Y*-axis by word concreteness from most abstract to most concrete. The twelve cognitive dimensions across the *x*-axis are ordered corresponding to their aggregative properties via factor analysis (for detailed discussion, see Troche et al., 2014). The heat map color range reflects the average Likert scale rating (1–7 range) across participants (n = 365).

tence. Afterward, line drawings of four objects (one target and three distractors) were shown, one in each quadrant of the screen, while participants' eye movements were recorded. The results showed that participants tended to fixate more (and earlier) on depicted objects that were associates of abstract words than associates of concrete words, consistent with the predictions of the QDR framework (Crutch & Warrington, 2005). However, the focus of the current paper is exclusively on abstract words that, by definition, cannot be visually depicted given their intrinsic low image ability. In an effort to study abstract as well as concrete words, a printed word version of the visual world paradigm has emerged (Huettig & McQueen, 2007; Mcqueen & Viebahn, 2007). Participants listened to sentences containing target words that were similar semantically to concepts invoked by concurrently displayed printed words. Here we further adapt the visual world paradigm by substituting orthographically presented words for pictures. As compared to the classical visual world paradigm, where pictures of real objects are used in order to keep the task as ecological as possible, in our case, the focus is on the words' processing (Huettig & McQueen, 2007; Mcqueen & Viebahn, 2007; Salverda & Tanenhaus, 2010).

Our aims here were two-fold. From a theoretical standpoint we evaluate the claim that multiple cognitive systems (not restricted to sensorimotor and emotion information) influence the organization and processing of abstract conceptual knowledge. This hypothesis was based on previous evidence and the premise that the ACF rating procedure provides both a window into the semantic attributes of individual abstract concepts and a quantitative metric for measuring similarity between two or more concepts (as opposed to their paradigmatic association, which is better captured by co-occurrence-based metrics such as LSA). The hypothesis was tested by examining the utility of semantic similarity metrics derived from ACF ratings of individual concepts to predict eye movement behavior within arrays of written abstract words. From a methodological perspective, our goal was to adapt the classical visual world paradigm in order to create a within-modality VWP suited to exploring the semantic attributes of and relationships between abstract words.

2. Methods

2.1. Participants

Twenty healthy adults completed the experiment (Mean [SD] age = 45.9 [21.6]; 12/20 female). This participant group was comprised of 10 young volunteers (Mean [SD] age = 25.6 [4.03]; 5/10 female) and 10 older volunteers (Mean [SD] age = 67.1 [7.4]; 7/10 female). The two groups were recruited in order to examine the impact of age upon performance in the experimental paradigm, with an eye to the future use of the experimental paradigm in both cognitive studies involving a typical young (largely student) participant population and neuropsychological studies of individuals with stroke or degenerative disease-related language dysfunction in which healthy age-matched controls may be required. All participants had normal or corrected to normal vision. Only individuals who had British English as their native language were recruited for the study.

2.2. Stimuli

The stimuli were 40 probe words and 40 four-item written word arrays. For each probe word (e.g. logic), the matching array comprised a target and three distractors (e.g. wisdom, evidence, establishment, darkness; see Appendix A). The list of stimuli used as well as the probe's average relatedness in the ACF matrix (and standard deviations, min and max), and the semantic relatedness of each target and distractor to the probe are also reported.

Probe, target and distractor words were selected from 208 abstract words (concreteness rating <450; MRC Psycholinguistic Database; Coltheart, 1981) which formed part of a larger corpus of 400 nouns on which ACF ratings were previously acquired (Troche et al., 2014). Following Crutch et al. (2012), ratings were gathered by asking participants to rate individual concepts on 12 unique dimensions using 7-point Likert scales. The Likert ratings from 7 (agree) to 1 (disagree) indicated participants' level of agreement with statements concerning the target word's salience in the following 12 cognitive dimensions: sensation, ease of teaching, ease of modifying, action, time, emotion, morality, polarity, social interaction, thought, space, and quantity. Three further rating scales concerning the extent to which a concept was positive or negative (polarity) and the ease with which the concept could be modified or taught were also completed. A description of these parameters as presented to raters can be found in Appendix B (see also Troche et al., 2014). Examples of words with different values (high, medium and low) for each different cognitive dimension are reported in Appendix B.

A symmetric matrix of pairwise semantic similarity ratings was derived for the 208 word set. Values denoted the Euclidean distance between words in a given pair based upon ACF ratings on the 12 dimensions specified above, with low values indicating semantic relatedness (semantically close items) and high values indicate semantic unrelatedness (semantically distant items).

In this study, five words were used in each trial: a probe, a target and three distractors. The similarity relationships between the probe and the target and between the probe and each of the three distractors were defined on the basis of the ACF ratings:

1. Target: closely related words (e.g. logic-wisdom; ACF Euclidean distance-from-probe value: 1.0–2.5)
2. Distractor 1: moderately related words (e.g. logic-evidence; ACF Euclidean distance-from-probe value: 3.0–4.0)
3. Distractor 2: minimally related words (e.g. logic-establishment; ACF Euclidean distance-from-probe value: 4.0–5.0)
4. Distractor 3: unrelated words (e.g. logic-darkness; ACF Euclidean distance-from-probe value: 5.5–8.0)

For all selected probe-target and probe-distractor word pairs, pairwise Latent Semantic Analysis (LSA; www.lsa.colorado.edu) cosines were also obtained (see Table 1). In addition, data showing the mean and standard deviation word frequency (CELEX; Baayen, Piepenbrock, & van Rijn, 1993), word length, concreteness, and age of acquisition are

Table 1
Mean (and standard deviation) ACF Euclidean distance, LSA cosine, frequency, length, concreteness, age of acquisition, orthographic and phonologic neighborhood data for probe, target, and distractor stimuli

	Probe	Target	Distractor 1	Distractor 2	Distractor 3
ACF Euclidean distance to probe	—	1.66 (0.37)	3.45 (0.23)	4.53 (0.24)	6.10 (0.48)
LSA cosine with probe	—	0.26 (0.16)	0.14 (0.09)	0.14 (0.12)	0.07 (0.07)
CELEX frequency	44.8 (49.9)	60.2 (52.7)	33.4 (39.7)	27.9 (52.7)	34.0 (31.1)
Length (number of letters)	8.10 (2.44)	8.43 (2.33)	9.00 (2.17)	9.25 (2.36)	8.23 (2.55)
Concreteness	314 (54)	299 (52)	316 (53)	297 (46)	334 (63)
Age of acquisition	8.9 (2.6)	9.0 (2.0)	10.0 (2.2)	9.6 (1.8)	8.1 (2.6)
Orthographic neighbourhood (N)	0.3 (0.6)	0.4 (0.9)	0.3 (0.9)	0.3 (0.6)	0.4 (1.0)
Phonological neighbourhood (N)	1.5 (2.4)	1.3 (1.5)	1.0 (1.1)	1.0 (1.0)	1.6 (1.6)

shown in Table 1. Words were all nouns and were matched for length, orthographic and phonological neighborhood, and concreteness.[1] Targets were significantly more frequent than the three distractors (all ps < .01). Significant differences in the AoA variable emerged also between target and Distractor 1, Distractors 1 and 3, and Distractors 2 and 3 (all ps < .01). We corrected for all these psycholinguistic variables adding them as covariates in the regression analyses (see below, Statistical analysis section).

All word stimuli were presented in lowercase Courier New font (letter height subtending a maximum of 1.1° visual angle). Probe words were presented at the center of the screen. Target and distractor words in the written word response array were located in the center of each quadrant of the screen (±10.5° horizontally and ±7.3° vertically from the center of the screen). Each word type appeared equally often at each location within the array.

2.3. Procedure

Each trial commenced with a fixation cross subtending 1.5° being presented at the center of the screen. We employed a gaze contingent response trigger to ensure that each participant started the exploration from the center of the screen. This gaze trigger worked by establishing a bounding box area of interest (AOI) around the fixation cross. When the eye tracker accumulated >100 ms of consecutive dwell time in the bounding box AOI, the trial automatically advanced to the next screen which consisted of a 4 word preview array. Following Odekar, Hallowell, Kruse, Moates, and Lee (2009), the written word response array was then previewed for 4 s, followed by the probe word for 2 s, and then the written word response array was re-presented for 4 s. Participants were requested to look at each word in the response array preview, and then to look at the probe word and then decide as quickly as possible which word in the response array was most related to the probe. Participants were asked to indicate the target word by clicking on the item using a standard mouse placed in their dominant hand. The mouse indicator was not visible during the preview and the probe presentation, but it appeared only during the postview period, and its initial position was always centered in the middle of the screen. Eye movements and fixation patterns were recorded throughout each trial, plus the response latencies of participants' word selection.

2.4. Apparatus

Stimuli were presented on a Dell 2120 desktop computer. Eye movements were recorded using a head-mounted infrared video-based eye tracker (Eyelink II; SR Research, Canada). Gaze position was recorded at 250 Hz, and corneal reflection was used for the 20 participants). Fixations and saccades were parsed by the Eyelink system, using standard velocity and acceleration thresholds (30°/s and 8000°/s^2). We used built-in programs provided with the eye tracker for calibration and validation purposes (9 points presented in a random sequence). All the data analyzed were obtained from recordings with an average Cartesian prediction error of <1° during the validation procedures. Participants used a chin rest (wide HeadSpot; University of Houston College of Optometry) to provide stability and maintain viewing distance throughout the experiment.

2.5. Analysis

2.5.1. Data processing:

Eye movement data were processed using EyeLink Data Viewer software (SR Research Ltd., Mississauga, Ontario, Canada). Blinks were identified and removed using Eyelink's automated blink detection. Each trial was split in three different time intervals: preview, probe presentation and postview. For each time interval three eye movement parameters were measured separately for each participant on each stimulus presented on the screen: number of fixations, mean fixation duration, and total fixation duration. We also measured in both the preview and the postview condition the ending point of the first saccade. Technically, there were no correct or incorrect responses, so we labeled those responses when participants chose the target as expected responses and responses when they chose one of the distracters instead as unexpected responses. Following this classification, for each participant we measured the percentage of expected responses (i.e., number of mouse clicks on the target word) and the reaction times (i.e., time from the postview array onset to the participant mouse click).

2.5.2. Statistical analysis

In order to assess the relation between the eye movement pattern and the relatedness of the words to the probes, separate hierarchical regression models were constructed for the following dependent variables: number of fixations, mean fixation duration, and total fixation duration. Although indexing different dimensions of visual attention allocation, the three dependent variables were highly correlated (all ps $< .0001$). For this reason, a single multivariate model including the three dependent variables could not be run. We, therefore, conducted separate hierarchical multiple regression models between the three dependent variables. Each analysis included the following predictors: relatedness of the words to the probe (see below for details), position on screen (top left, top right, bottom left, bottom right), word length, frequency, age of acquisition, concreteness, and trial number, with responses clustered by participant in order to adjust for repeated measures. Wald tests were carried out to explain main effects, pairwise comparisons of different levels of factors, and interactions. The words' relatedness to the probe was initially modeled as a discrete variable ("ITEM TYPE"; target, Distractor 1, Distractor 2, Distractor 3), following the stimuli selection described in the method section. However, we also aimed at obtaining an indirect ACF-LSA comparison in order to further validate ACF metrics as a relevant marker of one facet of semantic relatedness. To this end we conducted an additional analysis excluding ITEM TYPE and including the ACF (continuous measure of Euclidean distance from probe) and LSA values (cosine between probe and response item).

Finally, in order to describe the eye movement pattern associated with different behavioral responses as a post hoc analysis we re-ran the models isolating trials with expected and unexpected responses and omissions (see below).

3. Results

Three metrics of eye movement behavior when viewing the written word arrays both before (preview) and after (postview) presentation of the probe are shown in Fig. 2: number of fixations (Panel A), mean fixation duration (Panel B), and total fixation duration (Panel C).

3.1. Item type

Considering the results of the postview condition in terms of the ITEM TYPE as defined by the semantic relatedness of each word to the probe (left hand side of Fig. 2), targets (low ACF Euclidean distance from the probe) were shown to be fixated significantly more frequently than distractors (high ACF Euclidean distance from the probe; $p < .001$). These fixations were also found to be of significantly greater mean duration ($p < .001$) and the total fixation time was higher ($p < .001$). By contrast, in the preview period, there were no differences in the number, mean duration or total duration of fixations between targets and distractors ($p > .9$, $p = .08$ and $p > .2$). A significantly higher percentage of first saccades were also made toward the target word relative to distractors during the postview (37.6%; chance = 25%), while this was not true for the preview condition (24.2%). During the 2 s probe presentation, a small proportion of fixations (23%) were not on the probe itself but on the position occupied by the four words during the preview and postview periods. Considering these fixations, 27% of them, just above chance level, were directed toward the position occupied by the target although the word was not there at that moment.

3.2. ACF factor

In the postview period, ACF Euclidian distance predicted the number of fixations ($t = -14.06$, $p < .001$), mean fixation duration ($t = -8.95$, $p < 0001$), and total fixation duration ($t = -17.92$, $p < .001$) indicating that words that had a closer semantic relationship with the probe (smaller ACF distance) were fixated more frequently and for

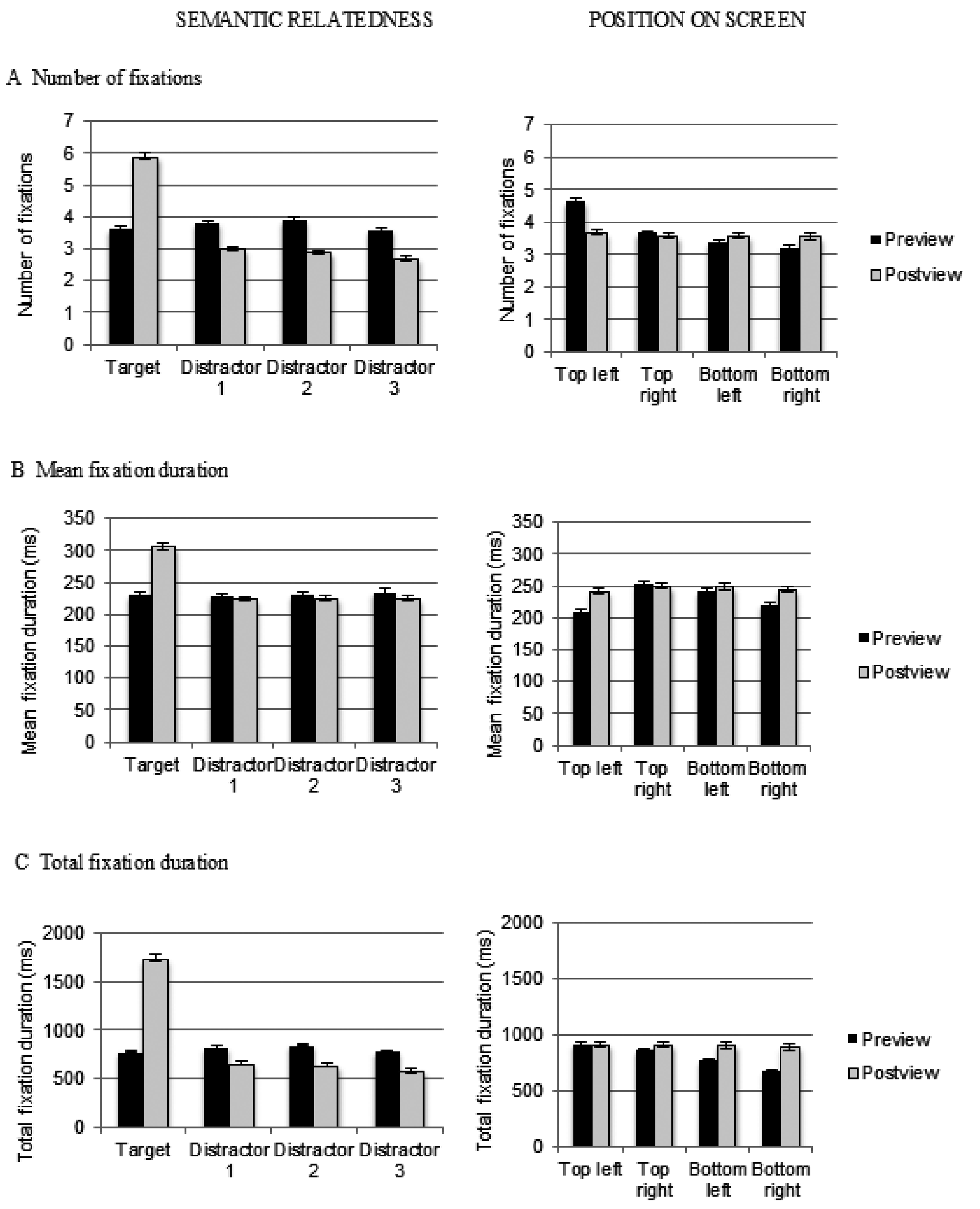

Fig. 2. Mean and standard error (A) number of fixations, (B) mean fixation duration, and (C) total fixation duration, considering results both by semantic relatedness and position on screen.

longer. Words having higher LSA cosines (indicating greater co-occurrence of items) were also associated with significant increases in number of fixations ($t = 6.05$; $p < 0.001$) and total fixation duration ($t = 4.81$; $p < .001$) but not mean fixation duration ($t = 1.38$; $p > 0.2$).

We further explored the relationship between the number of fixations and the ACF ratings of the three distractors. Results indicate the presence of a graded effect, with a smaller number of fixations on distractors having a higher distance from the probe ($r = -.234$, $p = .005$). We believe that the observed effect is an underestimate, due to the task instructions which required participants to click on the most related word, while no explicit judgment was required for the remaining stimuli. This has probably determined a specific search strategy aimed at finding and focusing on the single most related target.

3.3. Position on the screen

The mean and standard error fixation data for words in each position (top left, top right, bottom left, bottom right) in the preview and postview periods are shown in Fig. 2 (right hand side). Word position had a significant effect upon number of fixations ($p < .001$) and total fixation duration ($p < .001$) but not mean duration ($p > .3$) during the preview period. By contrast, word position had no significant effect upon number of fixations, mean duration or total fixation time in the postview period ($p > .3$, $p > .5$ and $p > .5$, respectively).

3.4. Psycholinguistic variables

During the critical postview period, word frequency was associated with all three eye movement metrics (all $p < .01$), AoA was associated with mean fixation duration (both $p = .01$) but not with number of fixations and total fixation duration ($p = .9$ and .7, respectively) and word length with number of fixations ($p < .05$) but neither mean nor total fixation duration ($p > .2$ and $p > .8$, respectively). Concreteness was associated with total fixation duration ($p < .01$) but neither with number of fixations nor with fixation duration ($p = .14$ and .1, respectively).

3.5. Response type

The total number of expected response on the task was 72.4%. Of the remaining responses, 17.3% were unexpected and 10.3% were omissions (reflecting participants either stating they did not know the answer or being unable to respond within the 4000 ms period for which the postview written word array was visible). Scoring of unexpected responses indicated that in 40.2% of responses participants selected the moderately related word, in 30.6% of cases the minimally related and in 29.2% the unrelated word. Mean reaction times were 2842 ms for expected responses and 3473 ms for unexpected responses.

The mean and standard deviation fixation number and duration for each of these response types, is shown in Table 2. Eye movement behavior was re-analyzed according to response types. For items with an expected response, ACF ratings were a significant predictor of fixation number, mean and total fixation duration (t values = −15.8, −11.7 and −18.5, respectively; all $p < .001$). LSA cosines did not predict any eye movement measure (t values = 1.3, 0.8 and 1.8, respectively; p = .2, .4 and .1, respectively). For items with an unexpected response, ACF ratings were again a significant predictor of all three fixation metrics (number of fixations number: $t = 3.41$, $p = .003$; mean fixation duration: $t = 3.17$, $p = .005$; total fixation duration: $t = 5.13$, $p < .001$, respectively)., LSA cosines again did not predict any eye movement measure ($t = 1.59$, $p = .1$; $t = .13$, $p = .8$ and $t = 1.87$, $p = .08$). For these items with an unexpected response, Distractor 1 (moderately related) was fixated most frequently, with longer fixation duration and for the greatest total duration. Most notably, for items with no response, the fixation metrics most closely resembled those of expected responses, with fixation activity focused most upon the target. ACF ratings continued to be a significant predictor of fixation number ($t = -2.3$, $p = .03$) and total duration ($t = -3.3$, $p = .004$). LSA cosines were not a significant predictor of any of the three metrics (fixation number: $t = .97$, $p = .35$; mean duration: $t = -1.9$, $p = .07$; total duration: $t = .2$, $p = .84$).

Unexpected responses were well distributed across trials. Globally 26 trials were affected by unexpected responses, even if most of them (16 trials, 61%) were only minimally affected, with less than 30% of participants choosing the unexpected word as a response.

We looked at the difference in terms of ACF PROBE-TARGET in trials obtaining 100% of correct responses and trials triggering errors. Results suggest a higher similarity between target and probe in those trails obtaining 100% of target responses (1.53) and compared to trials eliciting “errors” (1.72, $p = .058$).

Table 2
Mean (and standard deviation) number of fixations and mean and total fixation duration upon target and distractor words for items where participants gave correct responses ($n = 579$), unexpected responses ($n = 139$) or made no response ($n = 82$)

	Target	Distractor 1	Distractor 2	Distractor 3
Number of fixations				
Expected responses	6.7 (2.7)	2.5 (1.7)	2.5 (1.6)	2.4 (1.6)
Unexpected responses	3.0 (1.8)	4.4 (2.7)	4.0 (2.5)	3.5 (2.5)
No response	5.0 (2.5)	3.9 (2.2)	3.8 (1.9)	3.0 (1.8)
Mean fixation duration (ms)				
Expected responses	327 (143)	211 (91)	218 (97)	212 (90)
Unexpected responses	226 (95)	258 (103)	247 (106)	268 (207)
No response	295 (163)	254 (112)	243 (90)	239 (112)
Total fixation duration (ms)				
Expected responses	2059 (860)	511 (375)	513 (366)	486 (381)
Unexpected responses	639 (422)	1120 (772)	997 (735)	925 (811)
No response	1359 (762)	947 (591)	924 (564)	684 (508)

The large majority of unexpected responses regarded the selection of the moderately related word (Distractor 1). In order to understand the source of the effect we studied the similarity between target and distractors in terms of ACF distance. Target and Distractor 1 were more semantically related (2.8) when Distractor 1 was chosen as the response as compared to trials were the Target was selected (3.3). Moreover the difference in term of ACF distance of the target and Distractor 1 from the probe was smaller in those trials with a high proportion of Distractor 1 words selected as response. The other responses concerning Distractor 2 and Distractor 3 were distributed across trails which prevented us from any formal analysis. These results, in addition to longer RTs for errors as compared to expected responses and to the eye movement data, suggest that participants were prone to select the "expected" most related word but a higher similarity with the Distractor 1 represented a source of confusion. In 57.5% of the trials, (23/40) LSA cosines and ACF ratings agreed with respect to which of the four words presented on the screen was the most semantically related to the probe. In the remaining 17 trials of disagreement, ACF predicted responses in 71.2% of the trials, LSA in 12% and neither of the two predicted responses in 16.8% of the cases.

3.6. Scan path

In order to describe the scan path of participants during the preview and postview conditions, we created the heatmaps shown in Fig. 3. For each participant we calculated, across the 40 trials, the order of exploration of the 4 words on the screen, excluding fixations within 3 degrees of the probe position. We considered only the first four non-consecutive observed quadrants and then we calculated, separately for each participant, whether each quadrant was explored first, second, third, or fourth on average. Each of the 20 participants is represented by a small square within each quadrant and numbered from 1 to 20. In the figure, later explorations are represented by progressively lighter colors. In the preview condition (top panel), the pattern of exploration is quite clear and similar for all participants, with the top-left quadrant tending to be explored first (darker color), followed by top-right, bottom-right, and bottom-left quadrants. Conversely, in the postview condition (bottom panel), there is no clear-cut pattern to the order in which quadrants were explored. This is probably due to the randomization of the target's position, which thus determined a different scan path across trials, suggesting that participants explored the display not according to the physical position of the words on the screen but according to the target position. Indeed, in many cases participants firstly fixated the target (38%) which was in a different screen position on each trial.

3.7. Time course analysis

Results from the time course exploration averaged by participants are reported in Fig. 4. The proportion of fixations on the probe (red line) is shown together with the proportion of fixations on the target and distractors. As shown in the figure, during the preview period (left side of the panel) the proportion of fixations is equally distributed

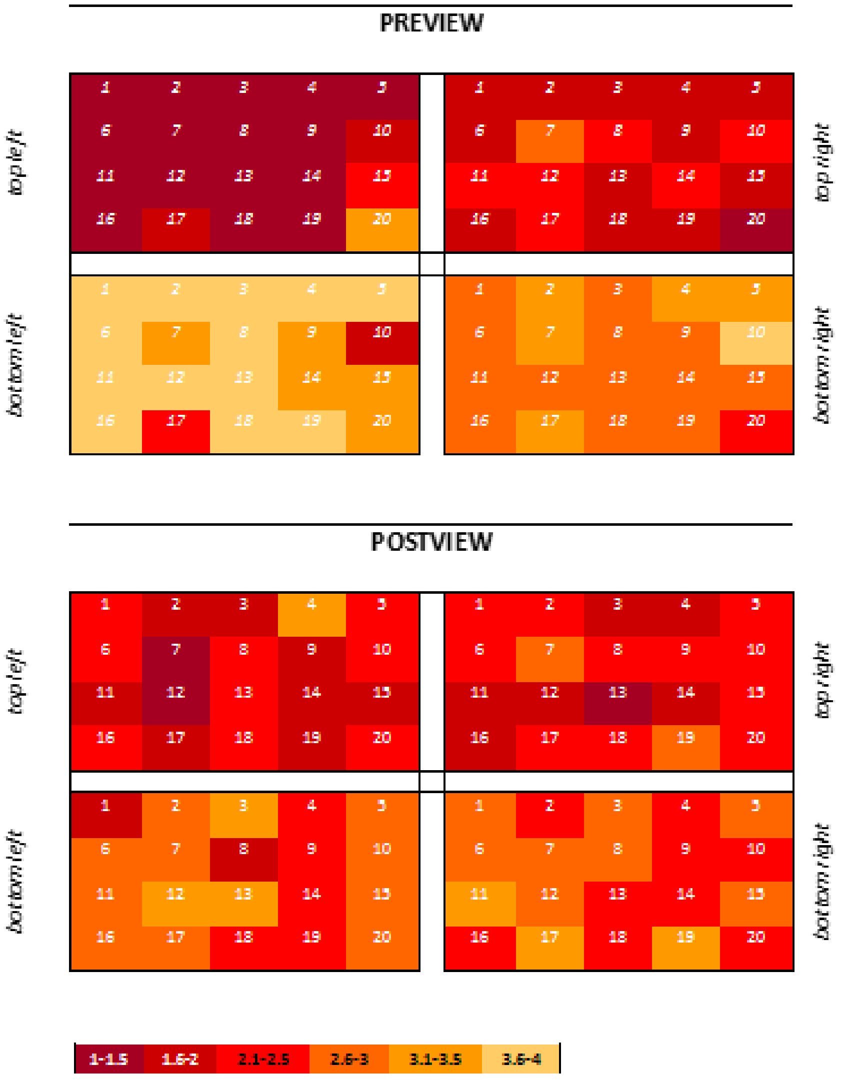

Fig. 3. Scan path for each participant (numbered from 1 to 20) for the four quadrants where words were presented: top left, top right, bottom left, and bottom right. In the top panel is represented the preview scan path while in the bottom panel is presented the postview scan path. For each trial the four quadrants were rank ordered in order to establish in what sequence (1st, 2nd, 3rd, 4th) they were explored. Darker colors indicated an earlier exploration of the area, while progressively lighter colors indicated a later exploration.

among the 4 words and it is stable for the 4s-preview. In the postview period (right side of the panel) there is an increase in proportion of fixations on the target word, which remains stable until the offset of the stimulus. Conversely, a decline in the proportion of fixations for Distractors 1, 2, and 3 as compared to the preview condition is observed.

The fine grained measures derived from the presented time course results strengthen the idea that the eye movement pattern described in the previous analysis is not merely

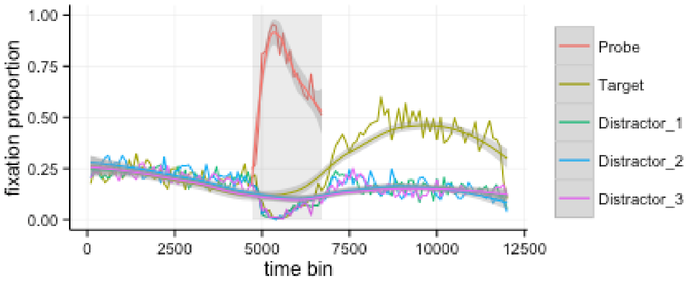

Fig. 4. Proportion of fixation averaged by participants on the probe (red line), on the target and on the three distractors. On the left side of the panel the preview period (4 seconds) is reported, in the middle there is the probe presentation (2 seconds) and on the right side there is the postview condition. Smoothed confidence intervals are reported for each time series.

the outcome of low and high level word-based features, such as psycholinguistic variables. This is specifically shown by the absence of any difference in terms of visual exploration among the 4 words in the preview time. The observed differences between the target and the distractors in the postview condition only suggest that the reported eye tracking data are a reasonable expression of the cognitive task assigned to participants.

3.8. Age group

The same linear regression analyses run and described above were re-run adding age group as a covariate (younger and older adults). The only age group difference of note was that older subjects showed a trend toward making more fixation in the postview condition ($p = .05$), while age group did not predict mean and total fixation duration (both $p > .1$). Analyses of semantic relatedness effects were also re-run independently for each group of participants. The influence of ACF Euclidian distance was comparable in each age group to that reported for the combined group above. All three eye movement metrics—number of fixations, mean, and total fixation duration - were influenced by the ACF distance in the younger ($t = -9.63$, -8.62 and -9.96, respectively; all $p < .0001$) and older groups ($t = -11.27$, -5.32 and -17.23, respectively; all $p < .0001$).

LSA cosines predicted number of fixations in the younger ($t = .383$, $p = .004$) and older subjects ($t = 2.18$, $p = .057$). Mean fixation duration and total fixation duration were significantly influenced by LSA values in older participants ($t = 2.54$ and 4.6, respectively, both $p < .05$) but not in younger participants ($t = .3$ and 1.51, respectively, both $p > .1$).

4. Discussion

The present paper explored the semantic relatedness of abstract words, and in particular tested the idea that multiple sources of information influence the representation of abstract concepts (Crutch et al., 2012, 2013; Troche et al., 2014). Our working hypothesis was that the semantic distance metrics derived from ACF ratings would predict participants' eye movements and response accuracy on a semantic relatedness written word 4-choice task.

Adapted from the visual *world* paradigm, the current VWP involved matching a written word probe to one of four written words (rather than pictures) in a response array. The results indicate that, once the probe had been processed, the word that was most semantically related to the probe (as defined as the probe-item pair separated by the smallest ACF Euclidean distance) received more and longer fixations, and was fixated for a larger amount of time as compared to the distractors. Participants were also inclined to make the first saccade toward the target word position relative to the distractors. In the preview period, word position on the screen significantly predicted the number of fixations and the total fixation duration. This effect is likely to reflect the natural tendency among participants to read words in the array from left to right and from top to bottom. In the postview condition, this was not the case and word position did not influence scan paths. Participants showed an overall tendency of selecting the most semantically related word according to the ACF rating of 73%; the distractors less semantically related to the probe according to the ACF ratings were also the items selected less frequently by participants. Crucially, eye movements were predicted by ACF Euclidian distances even when no response was given within the permitted time; target items were fixated more and for longer than other distractors. LSA cosines were not significant predictors of fixation number, mean duration or total duration when considering expected, unexpected or no response trials separately.

From a theoretical standpoint, the results provide further validation of Euclidean distance metrics derived from ACF ratings as a measure of semantic relatedness of abstract words. By extension, the results also provide further evidence that the high-dimensional space derived from ACF ratings provides a reasonable approximation of the organization of abstract conceptual space. More broadly, the data are consistent with the theoretical account on which ACF measures are grounded: multiple sources of information and a broad set of cognitive domains shape the acquisition and organization of abstract concepts (Crutch et al., 2012, 2013; Troche et al., 2014). The study is also the first to demonstrate that ACF ratings are predictive of semantic task performance in healthy individuals, with previous studies having tested the ability of ACF data to predict or explain cognitive performance in comprehension-impaired stroke patients.

The Dual Coding Theory (DCT, Paivio, 1971, 2013) states that while both concrete and abstract words are verbally represented, only concrete words also have nonverbal imagistic representation, which are modality specific (i.e., visual, auditory, etc). Conversely other studies primarily interested in the nature of the abstract concepts, highlighted the importance of sensory information in the representation of such concepts. For example, Connell and Lynott (2012) highlight the important role played by some perceptual factors, such as sound, taste, touch, smell and vision. Our results, by following the predictions made on the basis of the ACF model, better line up with the latter theoretical hypothesis: Not only linguistic information, as claimed by the DCT, but multiple cognitive systems contribute to the representation of abstract words. Specifically the idea that many sources of information about concepts are derived from specific modalities, converge and are then integrated into a single, coherent representation is compatible with the recently proposed Dynamic Multilevel Reactivation framework (Reilly, Peelle, Garcia, & Crutch, 2015). The latter is, indeed, a multilevel model which integrates fully embodied and fully disembodied theories. It highlights the necessary interconnection between sensorimotor and a modal representation and describes a multidimensional conceptual topography for both abstract and concrete words.

The results coming from the present experiment, which highlight how well the ACF measure of semantic similarity predicts performance with abstract words, might on the surface appear to contradict the qualitatively different representations hypothesis (QDR, Crutch & Warrington, 2005). The QDR position is often summarized as claiming that concrete concepts are organized by semantic similarity, whereas abstract concepts are organized by associative relations. However, the QDR proposal actually claims that both similarity-based and associative information are both relevant for the representation of abstract and concrete words, but their relative relevance for the two is different. Indeed, the concept of "relative rather than absolute" higher importance of similarity than association for concrete concepts, and of association than similarity for abstract concepts, is constantly highlighted in the original proposal and subsequent studies in both aphasic and healthy individuals (e.g. Crutch, 2006; Crutch & Jackson, 2011).

From a methodological standpoint, our results support the utility of an adapted visual world paradigm in which written words rather than pictures and spoken words are employed. Historically, the VWP has been used to study how relevant visual information affects spoken language processing (Cooper, 1974; Tanenhaus et al., 1995). To investigate such a relationship, fixation proportions on the interest areas during each time window or counts of saccades toward the regions of interest initiated during each time window have been systematically investigated (e.g., Altmann, 2004). Although visual world paradigms involving written words have been developed and validated previously (Huettig & McQueen, 2007; Mcqueen & Viebahn, 2007), such studies have tended to continue to use a spoken word or sentence probe and have focused on orthographic processing during speech perception (Salverda & Tanenhaus, 2010) or aimed to disentangle phonological and semantic interactions among stimuli (e.g., Huettig & McQueen, 2007). In the present paper our aims were different, since we sought to investigate the semantic relatedness between the probe and the target abstract words. The target-focused eye movement behavior observed in the current trial even for items to which no response was made suggests that even in cases in which an individual is unable to explicitly indicate or respond to a semantic relationship between two stimuli, it may still be possible to gather meaningful data regarding semantic processing of those items.

Several potential limitations of the study are noteworthy. First, it is important to emphasize that the (indirect) comparison of ACF and LSA ratings in this study does not in any way imply that one metric is superior to another. In the authors' view, these two metrics measure different but complementary facets of abstract concepts; ACF ratings were designed to measure the features or constituents of individual concepts (thus permitting examination of the similarity of two or more concepts), while LSA cosines reflect the paradigmatic association of concepts through the co-occurrence of two words in language usage. The QDR framework (Crutch & Warrington, 2005) and other theories or models of abstract and concrete conceptual representation (e.g. Plaut, 1995) stipulate the importance of and/or potential mechanisms supporting both similarity and association-based information, both of which are necessary for a rich and flexible semantic system. The other reason to caution against a direct comparison of ACF and LSA is that these metrics are generated in different ways: although both relate to distances within machine-generated high-dimensional abstract spaces, ACF metrics are derived from explicit human ratings of individual words on multiple dimensions (Crutch et al., 2013) while LSA values are derived from an indirect corpus-based analyses of documents and the terms they contain (Landauer & Dumais, 1997).

Second, it is also important to bear in mind that eye movement patterns are greatly influenced by the task instructions. Buswell (1935) and Yarbus (1967), but also more recently Castelhano, Mack, and Henderson (2009) used different task instructions, like visual search and memorization on the same set of stimuli with the aim of individuating differences in the eye movement pattern. These studies demonstrated that task instructions influence a number of eye movement measures including the number of fixations and gaze duration on specific objects, although some other measures like average saccade amplitude and individual fixation duration remain constant across the tasks (Castelhano et al., 2009). This sort of bias may partly explain the lack of a more linear relationship between eye movement measures and semantic relatedness across the targets and (progressively less related) distractors [i.e., by asking participants to select the (one) word most related to the probe]. It is plausible that a different task instruction, like asking participants to select all the words related to the probe, or to rank the four items in the response array in terms of their relatedness to the probe may have resulted in a different, more nuanced pattern of results. Third, frequency matching of the target and distractor items was not optimal with target words having marginally higher frequency values. This difference is unlikely to have influenced the results reported as item frequency was a covariate in all analyses. In addition, eye movement literature on word processing has indicated systematically that low—and not high—frequency words receive more fixations, which are also of longer duration (Inhoff & Rayner, 1986; Juhasz, Liversedge, White, & Rayner, 2006; Juhasz & Rayner, 2003, 2006; Kennison & Clifton, 1995; Kliegl, Grabner, Rolfs, & Engbert, 2004; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner & Duffy, 1986; Rayner & Fischer, 1996; Rayner, Sereno, & Raney, 1996; Schilling, Rayner, & Chumbley, 1998; Yan, Tian, Bai, & Rayner, 2006). Such an eye movement pattern is believed to reflect the higher difficulty in processing low frequency words (for a review, see Rayner, 2009). Fourth, in order to allow a similar amount of time for the preview and the postview conditions, eye movements' data have been collected and analyzed even after the participant's motor response, until the 4 s time limit had expired. This may have biased results, by inflating the number of fixations on the selected word. However, the large amount of missed responses within the time limit (10%) and the observed long reaction times (average = 2964 ms), suggest that, globally, a small amount of fixations has been made after the motor response. Moreover, in those cases where it happened, although some participants kept looking at the target word after its selection, others moved around among the four words presented on the display.

In conclusion, the results provide further support for the idea that a high-dimensional semantic space derived from ACF ratings may yield a reasonable, approximation of abstract conceptual space (Crutch et al., 2012, 2013; Troche et al., 2014). Moreover, it has been shown that a within-modality written word version of the visual world paradigm has utility in studying the relationships between and organization of abstract words in both young and older participants.

Acknowledgments

This work was undertaken at UCLH/UCL which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centers funding scheme. The Dementia Research Centre is supported by Alzheimer's Research UK, Brain Research Trust, and The Wolfson Foundation. This work was supported by the NIHR Queen Square Dementia Biomedical Research Unit, US Public Health Service grant DC013063 to JR, and by an Alzheimer's Research UK Senior Research Fellowship and ESRC/NIHR (ES/K006711/1) and EPSRC (EP/M006093/1) grants to SC.

Note

1. Only a marginally significant difference emerged between distractors 2 and 3 ($p = .05$).

References

Allman, M. J., & Meck, W. H. (2012). Pathophysiological distortions in time perception and timed performance. *Brain*, *135*(3), 656–677. doi: 10.1093/brain/awr210.

Altmann, G. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, *93*(2), B79–B87. doi: 10.1016/j.cognition.2004.02.005.

Altmann, G. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*(2), 190–200. doi: 10.1016/j.actpsy.2010.09.009.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, *116*(3), 463. doi: 10.1037/a0016261.

Baayen, R. H., Piepenbrock, R., & van H, R. (1993). The {CELEX} lexical database on {CD-ROM}.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. doi: 10.1093/cercor/bhp055.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917.

Bloom, L. (1998). Language development and emotional expression.*Pediatrics*, *102* (Supplement E1), 1272–1277.

Borghi, A. M., & Cimatti, F. (2012). Words are not just words: The social acquisition of abstract words. *Rivista Italiana di Filosofia del Linguaggio*, *5*, 22–37. doi: 10.4396/20120303.

Breedin, S. D., Saffran, E. M., & Coslett, H. B. (1994). Reversal of the concreteness effect in a patient with semantic dementia. *Cognitive Neuropsychology*, *11*(6), 617–660. doi: 10.1080/02643299408251987.

Buswell, G. (1935). *How people look at pictures*. Oxford, England: University of Chicago Press.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3):6, 1–15. doi: 10.1167/9.3.6.

Cipolotti, L., & Warrington, E. K. (1995). Semantic memory and reading abilities: A case report. *Journal of the International Neuropsychological Society*, *1*(01), 104–110.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497–505.

Coltheart, M., Patterson, K., & Marshall, J. C. (1987). *Deep dyslexia since 1980*. Routledge.

Connell, L., & Lynott, D. (2012). Strength of perceptual experience predicts word processing performance better than concreteness or imageability. *Cognition*, *125*(3), 452–465.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning ofchipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163. doi: 10.1037/0096-3445.132.2.163.

Crutch, S. J. (2006). Qualitatively different semantic representations for abstract and concrete words: Further evidence from the semantic reading errors of deep dyslexic patients. *Neurocase*, *12*, 91–97. doi: 10.1080/13554790500507172.

Crutch, S. J., & Jackson, E. C. (2011). Contrasting graded effects of semantic similarity and association across the concreteness spectrum. *The Quarterly Journal of Experimental Psychology*, *64*(7), 1388–1408.

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*, doi:10.3389/fnhum.2013.00186.

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain*, *128*(3), 615–627. doi: 10.1093/brain/awh349.

Crutch, S. J., Williams, P., Ridgway, G. R., & Borgenicht, L. (2012). The role of polarity in antonym and synonym conceptual knowledge: Evidence from stroke aphasia and multidimensional ratings of abstract words. *Neuropsychologia*,*50*(11), 2636–2644. doi: 10.1016/j.neuropsychologia.2012.07.015.

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition*, *110*(2), 284–292. doi: 10.1016/j.cognition.2008.11.012.

Farris-Trimble, A., & McMurray, B. (2013). Test–retest reliability of eyetracking in the Visual World Paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, *56*(4), 1328–1345. doi: 10.1044/1092-4388(2012/12-0145).

Fiebach, C. J., & Friederici, A. D. (2004). Processing concrete words: fMRI evidence against a specific right-hemisphere involvement. *Neuropsychologia*, *42*(1), 62–70. doi: 10.1016/S0028-3932(03)00145-3.

Franklin, S., Howard, D., & Patterson, K. (1995). Abstract word anomia. *Cognitive Neuropsychology*, *12*(5), 549–566. doi: 10.1080/02643299508252007.

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174. doi: 10.1080/02643290042000053.

Gathercole, V. C. (1985). 'He has too much hard questions': The acquisition of the linguistic mass–count distinction in much and many. *Journal of child language*, *12*(02), 395–415.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., & Gee, J. (2002). The neural basis for category-specific knowledge: An fMRI study. *Neuro Image*, *15*(4), 936–948. doi: 10.1006/nimg.2001.1028.

Hoffman, P., Jones, R. W., & Lambon Ralph, M. A. (2013a). Be concrete to be comprehended: Consistent imageability effects in semantic dementia for nouns, verbs, synonyms and associates. *Cortex*, *49*(5), 1206–1218. doi: 10.1016/j.cortex.2012.05.007.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013b). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–30. doi: 10.3758/s13428-012-0278-x.

Huettig, F., & Altmann, G. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. doi: 10.1016/j.cognition.2004.10.003.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. doi: 10.1016/j.jml.2007.02.001.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. doi: 10.1016/j.actpsy.2010.11.003.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431–439.

Jessen, F., Heun, R., Erb, M., Granath, D. O., Klose, U., Papassotiropoulos, A., & Grodd, W. (2000). The concreteness effect: Evidence for dual coding and context availability. *Brain and Language*, *74*(1), 103–112. doi: 10.1006/brln.2000.2340.

Juhasz, B. J., Liversedge, S. P., White, S. J., & Rayner, K. (2006). Binocular coordination of the eyes during reading: Word frequency and case alternation affect fixation duration but not fixation disparity. *The Quarterly Journal of Experimental Psychology*, *59*(9), 1614–1625. doi: 10.1080/17470210500497722.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312–1318.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition*, *13*(7–8), 846–863. doi: 10.1080/13506280544000075.

Katz, R. B., & Goodglass, H. (1990). Deep dysphasia: Analysis of a rare form of repetition disorder. *Brain and Language*, *39*(1), 153–185. doi: 10.1016/0093-934X(90)90009-6.

Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 68.

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*(4), 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*(1), 14. doi: 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. doi: 10.1016/j.cognition.2009.06.007.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

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.

Marshall, J., Pring, T., Chiat, S., & Robson, J. (1996). Calling a salad a federation: An investigation of semantic jargon. Part 1—Nouns. *Journal of Neurolinguistics*, *9*(4), 237–250. doi: 10.1016/S0911-6044(97)82796-0.

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language*, *43*(2), 240–274. doi: 10.1016/0093-934X(92)90130-7.

Mcqueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661–671. doi: 10.1080/17470210601183890.

Mellet, E., Tzourio, N., Denis, M., & Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *NeuroReport*, *9*(5), 803–808.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 65–79. doi: 10.1037/0278-7393.34.1.65.

Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, *6*(10), 799–809. doi: 10.1038/nrn1768.

Myung, J. Y., Blumstein, S. E., & Sedivy, J. C. (2006). Playing on the typewriter, typing on the piano: Manipulation knowledge of objects. *Cognition*, *98*(3), 223–243. doi: 10.1016/j.cognition.2004.11.010.

Newcombe, P. I., Campbell, C., Siakaluk, P. D., & Pexman, P. M. (2012). Effects of emotional and sensorimotor knowledge in semantic processing of concrete and abstract nouns. *Frontiers in Human Neuroscience*, *6*. doi: 10.3389/fnhum.2012.00275.

Noppeney, U., & Price, C. J. (2004).Retrieval of abstract semantics. *Neuro Image*, *22*(1), 164–170. doi: 10.1016/j.neuroimage.2003.12.010.

Odekar, A., Hallowell, B., Kruse, H., Moates, D., & Lee, C. Y. (2009). Validity of eye movement methods and indices for capturing semantic (associative) priming effects. *Journal of Speech, Language & Hearing Research*, *52*(1), 31–48. doi: 10.1044/1092-4388(2008/07-0100).

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

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta, S., T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). *Journal of Experimental Psychology: General*, *142*(1), 282–287. doi: 10.1037/a0027004

Papagno, C., Capasso, R., Zerboni, H., & Miceli, G. (2007). A reverse concreteness effect in a subject with semantic dementia. *Brain and Language*, *103*(1), 90–91.

Pecher, D., Boot, I., & Van Dantzig, S. (2011). Abstract concepts: Sensory-motor grounding, metaphors, and beyond. In B. Ross (Ed.), *The Psychology of Learning and Motivation*, vol. 54 (pp. 217–248). Burlington: Academic Press.

Perani, D., Cappa, S. F., Schnur, T., Tettamanti, M., Collina, S., Rosa, M. M., & Fazio, F. (1999). The neural correlates of verb and noun processing A PET study. *Brain*, *122*(12), 2337–2344. doi: 10.1093/brain/122.12.2337.

Plaut, D. C. (1995). Semantic and associative priming in a distributed attractor network. In Proceedings of the 17th annual conference of the cognitive science society (Vol. 17, pp. 37–42). Hillsdale, NJ: Lawrence Erlbaum Associates.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10*(5), 377–500. doi: 10.1080/02643299308253469.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. doi: 10.1080/17470210902816461.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the EZ Reader model. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 720–732.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics*, *58*(5), 734–747.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188–1200.

Reilly, J. (2005). A tale of two imageabilities: An interaction of sound and meaning in natural language perception. Doctoral dissertation. Temple University.

Reilly, J., & Kean, J. (2007). Formal distinctiveness of high-and low-imageability nouns: Analyses and theoretical implications. *Cognitive Science*, *31*(1), 157–168. doi: 10.1080/03640210709336988.

Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016, in press). Linking somatic and symbolic representation in semantic memory: The Dynamic Multilevel Reactivation Framework. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-015-0824-5.

Reilly, J., Peelle, J. E., & Grossman, M. (2007). A unitary semantics account of reverse concreteness effects in semantic dementia. *Brain and Language*, *103*(1), 86–87.

Roeltgen, D. P., Sevush, S., & Heilman, K. M. (1983). Phonological agraphia: Writing by the lexical-semantic route. *Neurology*, *33*(6), 755–755.

Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *NeuroImage*, *27*(1), 188–200. doi: 10.1016/j.neuroimage.2005.04.012.

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1108–1117. doi: 10.1037/a0019901.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281.

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* (1), 82–102.

Sirigu, A., Duhamel, J. R., & Poncet, M. (1991). The role of sensorimotor experience in object recognition a case of multimodal agnosia. *Brain*, *114*(6), 2555–2573.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Troche, J., Crutch, S., & Reilly, J. (2014). Clustering, hierarchical organization, and the topography of abstract and concrete nouns. *Frontiers in Psychology*, *5*. doi: 10.3389/fpsyg.2014.00360.

Warrington, E. K. (1975). The selective impairment of semantic memory. *The Quarterly Journal of Experimental Psychology*, *27*(4), 635–657.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–853.

Wise, R. J., Howard, D., Mummery, C. J., Fletcher, P., Leff, A., Büchel, C., & Scott, S. K. (2000). Noun imageability and the temporal lobes. *Neuropsychologia*, *38*(7), 985–994. doi: 10.1016/S0028-3932(99)00152-9.

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, *97*(2), 259–268.

Yarbus, I. A. (1967). *Eye movements and vision*. New York: Plenum Press.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 1–14. doi: 10.1037/0278-7393.32.1.1.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review*, *10*(4), 954–958.

Appendix A

Set	Probe	Probe's mean (sd) semantic relatedness	Range	Target (distance from the probe)	Distractor 1 (distance from the probe)	Distractor 2 (distance from the probe)	Distractor 3 (distance from the probe)
1	Knowledge	4.31 (1.34)	1.12–8.25	Intelligence (1.12)	Consideration (3.18)	Ignorance (4.54)	Announcement (5.52)
2	Confidence	3.24 (1.14)	0.65–6.77	Determination (1.41)	Economy (3.11)	Convergence (4.29)	Midnight (6.77)
3	Context	2.98 (0.96)	0.98–6.52	Situation (1.76)	Benefactor (3.76)	Interaction (4.05)	Laughter (6.51)
4	Variety	3.39 (1.02)	1.14–6.19	Difference (1.56)	Uncertainty (3.51)	Disagreement (4.17)	Honesty (6.12)
5	Topic	3.42 (1.04)	1.06–6.18	Example (1.06)	Benefactor (3.56)	Improvement (4.49)	Freedom (6.04)
6	Wisdom	3.92 (1.35)	1.48–7.64	Knowledge (1.55)	Assistance (3.59)	Impairment (4.63)	Clearance (6.55)
7	Logic	3.54 (1.18)	1.16–7.22	Wisdom (1.59)	Evidence (3.19)	Establishment (4.26)	Darkness (6.57)
8	Magnitude	4.19 (0.99)	1.48–7.34	Capacity (1.49)	Comparison (3.2)	Identification (4.33)	Christmas (6.11)
9	Faith	3.61 (1.36)	0.99–7.46	Belief (1)	Complication (3.4)	Impossibility (4.59)	Unit (6.45)
10	Proportion	4.04 (1.10)	1.38–7.23	Magnitude (1.77)	Adversity (3.55)	Intention (4.69)	Independence (5.53)
11	Behavior	4.05 (1.37)	1.46–8.08	Justice (1.46)	Deceit (3.13)	Cowardice (4.47)	Duration (6.42)
12	Paradigm	3.47 (1.04)	1.38–7.09	Method (2.22)	Error (3.85)	Opinion (4.53)	Truth (6.12)
13	Honesty	4.62 (1.53)	1.16–8.47	Truth (1.74)	Democracy (3.35)	Circumstance (4.74)	Accumulation (5.92)
14	Attitude	3.31 (1.23)	1.01–7.37	Opinion (1.01)	Crisis (3.52)	Dimension (4.63)	Opera (6.94)
15	Independence	3.61 (1.33)	1.32–7.37	Responsibility (1.75)	Availability (3.38)	Description (4.5)	Location (6.14)
16	Episode	3.85 (1.19)	1.49–6.91	Occasion (1.78)	Consistency (3.5)	Charity (4.65)	Kindness (6.91)
17	Character	3.39 (1.13)	1.33–7.49	Attitude (1.77)	Context (3.22)	Hierarchy (4.8)	Autumn (6.94)
18	Category	3.88 (1.25)	1.22–7.17	Combination (1.22)	Event (3.1)	Identity (4.84)	Responsibility (6.36)
19	Idiom	4.24 (1.14)	1.18–7.19	Metaphor (1.18)	Separation (3.48)	Incentive (4.47)	Accomplishment (6.13)
20	Satisfaction	3.61 (1.25)	0.95–7.28	Admiration (1.45)	Difference (3.65)	Moment (4.76)	Amplitude (6.13)
21	Freedom	4.45 (1.49)	1.14–8.10	Justice (1.41)	Expression (3.44)	Exclusion (4.98)	Appointment (6.41)
22	Autumn	5.61 (1.09)	1.99–8.15	Darkness (2.18)	Origin (3.81)	Debt (4.87)	Behavior (7.48)
23	Appointment	3.91 (1.05)	1.52–6.60	Arrangement (1.52)	Paradigm (3.54)	Dilemma (4.24)	Intelligence (6.02)
24	Brevity	3.69 (1.14)	1.21–7.12	Extent (1.21)	Conclusion (3.33)	Recognition (4.25)	Equality (5.59)
25	Accomplishment	3.78 (1.23)	1.47–7.35	Ambition (1.48)	Fantasy (3.83)	Pretense (4.6)	Category (5.89)
26	Kindness	4.72 (1.40)	1.58–8.12	Mercy (1.59)	Criticism (3.32)	Limitation (4.78)	Selection (5.64)
27	Translation	3.39 (0.95)	1.46–5.82	Definition (1.62)	Corporation (3.37)	Consequence (4.27)	Laughter (5.58)
28	Advantage	2.67 (0.85)	1.08–5.35	Opportunity (2.44)	Immortality (3.13)	Honesty (4.81)	Dynasty (5.67)
29	Equality	2.79 (0.77)	1.04–5.62	Unity (1.76)	Availability (3.8)	Coincidence (4.84)	Addition (5.66)

(continued)

Appendix A. (continued)

Set	Probe	Probe's mean (sd) semantic relatedness	Range	Target (distance from the probe)	Distractor 1 (distance from the probe)	Distractor 2 (distance from the probe)	Distractor 3 (distance from the probe)
30	Amplitude	4.32 (1.26)	1.50–7.56	Extent (1.91)	Arrangement (3.63)	Denial (4.16)	Character (5.73)
31	Clearance	4.15 (1.30)	1.65–7.36	Replacement (1.66)	Fallacy (3.61)	Explanation (4.64)	Hatred (5.75)
32	Christmas	5.03 (0.67)	1.90–6.47	Holiday (1.9)	Danger (3.87)	Duty (4.28)	Heresy (5.92)
33	Heresy	3.67 (1.02)	1.29–6.24	Fallacy (1.72)	Brevity (3.6)	Calculation (4.46)	Memory (5.74)
34	Memory	4.02 (0.90)	2.21–6.92	Attention (2.22)	Destiny (3.11)	Exception (4.82)	Gender (5.53)
35	Idea	3.95 (0.95)	1.83–6.97	Information (1.84)	Development (3.23)	Definition (4.15)	Idiom (5.78)
36	Dynasty	4.54 (1.18)	1.85–7.51	Distribution (2.26)	Coincidence (3.18)	Irony (4.5)	Willingness (5.78)
37	Leadership	3.39 (1.09)	1.36–6.74	Democracy (1.74)	Introduction (3.59)	Majority (4.12)	Item (6.36)
38	Decision	3.78 (1.18)	1.85–7.50	Logic (1.89)	Instinct (3.46)	Phenomenon (4.6)	Reduction (5.51)
39	Unit	4.71 (1.20)	1.85–7.69	Episode (2.45)	Disaster (3.42)	Stimulus (4.6)	Perception (6.05)
40	Responsibility	4.01 (1.45)	1.17–8.03	Integrity (1.51)	Legality (3.66)	Vocabulary (4.68)	Topic (5.59)

Appendix B

Parameter	Definition
Polarity	I relate this word to positive or negative feelings in myself
Sensation	I relate this word to physical feelings like vision, hearing, smelling, etc
Action	I relate this word to actions, doing, performing, and influencing
Thought	I relate this word to mental activity, ideas, opinions, and judgments
Emotion	I relate this word with human emotion
Social interaction	I relate this word with relationships between people
Time	I relate this word with time, order, or duration
Space	I relate this word to position, place, or direction
Quantity	I relate this word to size, amount, or scope
Morality	I relate this word to morality, rules, or anything that governs my behavior
Ease of modifying	I can easily choose an adjective for this word (the ugly truth, whole truth, etc.)
Ease of teaching	This word could be easily taught to a person who does not speak English

	High values words (>5)		Medium values words (3.5–5)		Low values words (<3)	
Polarity	Deceit	Honesty	Fact	Attitude	Item	Unit
Sensation	Opera	Autumn	Gender	Disaster	Impossibility	Legality
Action	Interaction	Production	Response	Skill	Dynasty	Debt
Thought	Knowledge	Idea	Capacity	Danger	Opera	Autumn
Emotion	Kindness	Laughter	Leadership	Duty	Reduction	Proportion
Social interaction	Unity	Equality	Ambition	Permission	Darkness	Dimension
Time	Midnight	Duration	Possibility	Situation	Hatred	Cowardice
Space	Location	—	Origin	Adversity	Deceit	Crisis
Quantity	Abundance	Magnitude	Recognition	Intention	Character	Gender
Morality	Honesty	Truth	Admiration	Identity	Addition	Holiday
Ease of modifying*	Truth	Responsibility	Preparation	Situation	Idiom	Gender
Ease of teaching*	Intelligence	Knowledge	Ignorance	Theory	Myth	Irony

*For these scores it should be noted that lower scores indicate greater ease of teaching and modifying.