Concreteness and word production

**J. Richard Hanley · Rebecka P. Hunt ·
Deborah A. Steed · Shannon Jackman**

Abstract Two experiments are reported that investigated the effect of concreteness on the ability to generate words to fit sentence contexts. When participants attempted to retrieve words from dictionary definitions in Experiment 1, abstract words were associated with more omissions and more alternates than were concrete words. These findings are consistent with the view that the semantic–lexical weights in the word production system are weaker for abstract than for concrete words. We found no evidence that greater competition from semantic neighbors was an additional reason why abstract words were harder to produce. Participants also reported more positive tip-of-the-tongue states (TOTs) when attempting to produce abstract words from their definitions, consistent with more phonological retrieval problems for abstract than for concrete words. In Experiment 2, participants attempted to generate words to fit into a sentence that described a specific event. The difference between the numbers of abstract and concrete words recalled was significantly smaller in the event condition than in the definition condition, and evidence no longer emerged of greater phonological retrieval failure for abstract words. Overall, the results are consistent with the view that the semantic–lexical weights, but not the lexical–phonological weights, are weaker for abstract than for concrete words in the word production system.

Keywords Concreteness effect · Word production ·
Tip-of-the-tongue states

In this study, we investigated the way in which rated concreteness influences the processes that are involved in word

J. R. Hanley (✉) · R. P. Hunt · D. A. Steed · S. Jackman
Department of Psychology, University of Essex,
Colchester CO4 3SQ, UK
e-mail: rhanley@essex.ac.uk

production. Experiments examining the effects on word production of variables such as frequency and age of acquisition have often employed tests of picture naming (e.g., Johnston, Dent, Humphreys, & Barry, 2010). Because it is not feasible to present abstract words in pictorial form, a few studies have instead used definition tasks to compare concrete and abstract word retrieval (Allen & Hulme, 2006; Hanley & Kay, 1997; Newton & Barry, 1997). Participants are presented with dictionary definitions of concrete and abstract words, and they attempt to recall the word that fits each definition. Allen and Hulme reported significantly better concrete than abstract word production on this task, even though their sets of concrete and abstract words were matched for frequency, morphemic complexity, age of acquisition, familiarity, neighborhood size, and word length.

As Allen and Hulme (2006) pointed out, generating words from dictionary definitions involves several distinct processes. It first requires access to a target word's semantic representation in memory. In terms of a model of word production such as that put forward by Foygel and Dell (2000), activation of the word's lexical representation can then take place from its semantic representation, followed by activation of the word's phonological features. Gollan and Brown (2006) argued that the kinds of errors that participants make during word production make it possible to distinguish problems in phonological access from problems that occur earlier in the word production process. Allen and Hulme only reported the numbers of correct retrievals associated with abstract and concrete word production. In the present study, however, we conducted a detailed analysis of the kinds of errors that participants made when retrieving abstract and concrete words, in an attempt to discover the processing stage(s) that are sensitive to concreteness. The errors comprise failures to respond (omissions), the production of an answer other than the target word (alternates), and tip-of-the-tongue (TOT) states. During a TOT state, a target
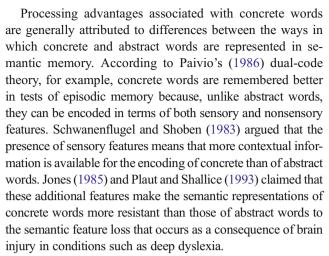
item feels as if it is about to be retrieved, even though it is temporarily inaccessible (see Brown, 2012, and Schwartz & Metcalfe, 2011, for recent reviews). If a participant resolves a TOT state by producing the target item or subsequently recognizes the target item as the word that elicited the TOT state, the experience is classified as a positive TOT. If the participant had a different word in mind, the experience is classified as a negative TOT.

Gollan and Brown (2006) argued that omissions, alternates, and negative TOTs indicate a failure to access either the semantic or the lexical representation of the target word, which they term a *Step 1* retrieval failure. Their claim is compatible with Foygel and Dell's (2000) model of word production. In this model, omissions reflect a failure to activate sufficiently strongly *any* lexical representation (e.g., Dell, Lawler, Harris, & Gordon, 2004; Laine, Tikkala, & Juhola, 1998). Alternates reflect activation of the lexical representation of a different word from the target item, as a result of insufficient activation of the appropriate lexical representation relative to a competitor. It may be the case that some activation of semantic alternates takes place during phonological mapping. However, the available evidence (Goldrick, 2006; Nozari, Dell, & Schwartz, 2011) suggests that such activation is weak and that only a relatively small number of semantic errors occur in Step 2 during phonological access. It is therefore assumed, for the present purposes, that production of a semantically related alternate reflects a failure at Step 1 of the retrieval process.

Consistent with several accounts of the etiology of TOT states (e.g., Meyer & Bock, 1992; Perfect & Hanley, 1992), Gollan and Brown (2006) argued that a positive TOT means that appropriate semantic and lexical information about the target word has been activated, but there has been a failure to access its associated phonological features. When participants report a positive TOT, therefore, they are experiencing a failure at the phonological retrieval stage, which Gollan and Brown term a *Step 2* retrieval failure. In terms of Foygel and Dell's (2000) model of word production, Step 2 failure reflects an inability to fully activate the appropriate output phoneme units, despite successful activation of the word's abstract lexical unit (for further discussion, see Hanley, 2011). A similar account of TOTs is provided by the discrete two-step model of production (e.g., Levelt, Roelofs, & Meyer, 1999), in which activation of phonological candidates only occurs after abstract lexical activation is complete.

Gollan and Brown (2006) developed some relatively simple equations that use the incidence of these different types of errors to distinguish the probability of Step 1 from Step 2 retrieval failure. They defined the probability of Step 1 failure as the sum of the omissions, alternates, and negative TOTs, divided by the total number of trials. The probability of Step 2 failure was defined as the number of positive TOTs divided by the sum of the positive TOTs and correct recalls.

Processing advantages associated with concrete words are generally attributed to differences between the ways in which concrete and abstract words are represented in semantic memory. According to Paivio's (1986) dual-code theory, for example, concrete words are remembered better in tests of episodic memory because, unlike abstract words, they can be encoded in terms of both sensory and nonsensory features. Schwanenflugel and Shoben (1983) argued that the presence of sensory features means that more contextual information is available for the encoding of concrete than of abstract words. Jones (1985) and Plaut and Shallice (1993) claimed that these additional features make the semantic representations of concrete words more resistant than those of abstract words to the semantic feature loss that occurs as a consequence of brain injury in conditions such as deep dyslexia.

If concrete words have richer semantic representations than abstract words do, it follows that the concreteness effect in the definition task reflects, to some extent at least, the greater ease with which the semantic representations of concrete target words are successfully accessed in response to their definitions. Hanley and Kay's (1997) study of a neuropsychological patient with speech output problems suggested that concreteness can also exert an effect on later stages of word production. This individual's word production, on tasks including both naming to definition and auditory repetition, was significantly worse when he was asked to produce words of low rather than high imageability. These word production problems were not the consequence of a semantic impairment involving the loss of semantic features, because this individual showed unimpaired auditory comprehension of low-imageability words. In the absence of a semantic impairment, it appears that he was suffering from a problem that disrupted the mapping of preserved semantic information onto lexical information during spoken word production. In an attempt to explain this pattern of performance, Hanley, Dell, Kay, and Baron (2004) used Foygel and Dell's (2000) model of speech production to simulate differences in imageability by varying the strength of the semantic–lexical connections in the model. Words of high imageability were assumed to have relatively strong semantic–lexical weights, and words of low imageability to have relatively weak semantic–lexical weights. In Foygel and Dell's model, this manipulation leads to weaker activation of lexical units and to the production of more alternates and omissions for words of low than of high imageability.

Newton and Barry (1997) put forward a slightly different explanation of why abstract words are hard to retrieve in word production tasks. They also argued that problems with abstract word production occur after the semantic representation of the target word has been successfully activated. Newton and Barry believed that problems arise because abstract words face particularly strong competition from

semantically similar words. Because the semantic representation of an abstract word generally contains relatively few features, it will typically share a high proportion of its semantic features with other lexical representations. As a consequence, the phonological representations of several competitors will become strongly activated, so a competitor of an abstract word will be more likely to reach threshold levels of activation than is the case with competitors of concrete words. Therefore, more semantic errors will occur for abstract than for concrete words. Such an account is consistent with recent research by Mirman and Magnuson (2008) showing that close semantic neighbors can have an inhibitory effect on a target word during some lexical-processing tasks. Of course, activation between the rich semantic representations of concrete words and their competitors will also take place. Nevertheless, Newton and Barry argued that a concrete target will contain a higher proportion of distinctive semantic features and, as a consequence, should be more easily activated to threshold on the definition task than should any of its competitors (see Cree & McRae, 2003, for evidence that distinctive semantic features can improve performance in some word production tasks).

Newton and Barry (1997) therefore believed that the problems with abstract words on the definition task occur during lexical access, because more semantic competitors are activated for abstract words. In a simulation using Foygel and Dell's (2000) model, Dell et al. (2004) examined the effect of varying the number of semantic competitors on word production. The simulation showed that an additional semantic competitor made the model produce more semantic errors and fewer omissions. Such an outcome is consistent with evidence from aphasic picture naming (Bormann, Kulke, Wallesch, & Blanken, 2008) that dense semantic neighborhoods elicit more semantic alternates but fewer omissions. So, contrary to Hanley et al. (2004), it follows from Newton and Barry's account that more semantic errors but not more omissions should arise during attempts to produce abstract words from dictionary definitions.

Our Experiment 1 provided an additional test of Newton and Barry's (1997) account. Half of the participants were provided with information about the number of letters that each target word contained, alongside its definition. It is likely that a competitor would often be a word of different length from the target word. If so, then informing participants about letter length should be beneficial for word retrieval, because such knowledge should make it possible to eliminate alternates whenever they contained a different number of letters from the target word. It follows from Newton and Barry's account that such information should be particularly beneficial for the number of abstract words retrieved, because according to their model, widespread activation of alternates of abstract words is the reason why abstract words are harder to produce.

Another important issue is whether concreteness affects the phonological (Step 2) as well as the semantic and lexical stages of the word production process. Weak semantic–lexical weights in an interactive activation model such as that proposed by Foygel and Dell (2000) could easily lead to retrieval problems at Step 2 as well as at Step 1. Because activation between the semantic, lexical, and phonemic levels occurs in a cascade, there will sometimes be sufficient activation of the lexical representation of an abstract word to partially, but not fully, activate its phonological representation (see Hanley & Turner, 2000, for a simulation of the way in which weak activation at one level in an interactive activation model can have knock-on effects at subsequent processing stages). Partial activation of phonological units should be particularly common for abstract words if their semantic–lexical weights are relatively weak. It therefore follows that positive TOT experiences should occur more often for abstract than for concrete words, with a greater probability of Step 2 retrieval failure.

In summary, in Experiment 1 we used Allen and Hulme's (2006) word definition task, in which half of the targets were concrete words and half were abstract, to investigate the effects of concreteness on word production. Participants were asked to indicate whether or not they were experiencing a TOT state when they failed to recall a word from its definition. This procedure made it possible to determine whether concreteness has its effect at semantic, lexical (Step 1), and phonological (Step 2) stages of the retrieval process (Gollan & Brown, 2006). It also made it possible to determine whether the greater probability of Step 1 failure comes about because definitions of abstract words elicit more failures to respond, more alternates, or more examples of both of these types of error than do concrete words. Finally, half the participants received information about the number of letters in each target word, to investigate whether this information would be particularly helpful for the retrieval of abstract words.

Experiment 1

Method

Participants A group of 56 undergraduate students at the University of Essex took part and were randomly assigned to one of two experimental groups. Half of the participants received information about the letter length of each target word, and half of the participants did not.

Materials The 68 target words and their definitions were generated by Allen and Hulme (2006). These included 34 concrete and 34 abstract words, which were matched for rated age of acquisition, rated word frequency, Kučera–Francis word frequency, familiarity, neighborhood size, number of phonemes, and number of letters. The definitions were taken from several different dictionaries. Allen and Hulme had used the ratings of six judges on a scale of 1–5

(high scores = *high suitability*) to closely match the suitability of the definitions for the concrete (mean = 4.74/5) and abstract (mean = 4.73/5) target words.

Procedure Participants were given a booklet containing 68 definitions arranged in a random order (half of the participants received a different booklet in which the order of the items was reversed). For example, the definition for *dare* (abstract) read, "To challenge someone into doing something." The definition for *lake* (concrete) read, "A large body of water surrounded on all sides by land." If participants knew the word that fit the definition, they wrote it down immediately below the definition. If they could not recall the word, they were asked to indicate whether or not they were in a TOT state, which was defined as follows: "If you know the word and feel that you are about to retrieve it but are unable to bring it to mind at the moment, then please indicate that you are in a tip-of-the-tongue state." The participants were tested individually and worked through the booklet at their own pace. They were not allowed to go back and answer earlier questions once they had moved on to another definition. If they had indicated a TOT that resolved spontaneously before they had started to read the next question, they were allowed to write down the word that had come to mind. At the end of the experiment, participants were shown the correct answer for every question for which they had indicated that they were in a TOT state. If the participants stated that the target item was the item for which they had experienced a TOT, the TOT was considered to be a positive TOT. If not, it was considered a negative TOT.

Any response that was not the target word was considered to be an alternate. In some situations, this meant that very close synonyms such as *insane* as a response to the definition of *mad* were counted as alternates. We return to this issue in Experiment 2, which included an attempt to assess how closely the alternates for concrete and abstract words matched the sentences that were used.

Half of the participants also received information about the letter length of each target word, in the form of a series of dashes that indicated the number of letters that each target contained. Participants were told that the number of dashes was the same as the number of letters in each target word, and that this information might help them recall the target word. These dashes were presented immediately below each definition.

Results

Table 1 shows the mean performance in Experiment 1 as a function of concreteness and information about letter length. Because the data were categorical (see Jaeger, 2008), they were converted to proportions and underwent an arcsine transformation before the analyses of variance (ANOVAs) were carried out.

Correct responses Significantly more concrete than abstract words were correctly named in response to the definitions, $F(1, 54) = 510.78$, $MSE = .03$, $p = .001$.

Step 1 errors We found a significantly higher overall probability of Step 1 failure (alternates, omissions, and negative TOTS) for abstract than for concrete words, $F(1, 54) = 384.84$, $MSE = .04$, $p = .001$ (see Table 2). Abstract words were associated with the production of significantly more alternative answers than concrete words, $F(1, 54) = 67.23$, $MSE = .03$, $p = .001$. A total of 87.5 % of the alternates were semantically related to the target word, and the remainder were phonologically related. Significantly more failures to respond occurred to definitions of abstract than of concrete words, $F(1, 54) = 69.55$, $MSE = .03$, $p = .001$, and significantly more negative TOTs occurred for abstract than for concrete words, $F(1, 54) = 26.85$, $MSE = .01$, $p = .001$.

Step 2 errors We found significantly more positive TOTs, $F(1, 54) = 14.89$, $MSE = .01$, $p = .001$, for abstract words. There was also a significantly higher probability of Step 2 failure [positive TOTs/(correct responses plus positive TOTs)] for abstract than for concrete words, $F(1, 54) = 23.07$, $MSE = .03$, $p = .001$.

Table 1 Mean proportions of different types of responses to the definitions for abstract and concrete target words in Experiment 1

Condition	Correctly Named	Positive TOTs	Negative TOTs	No Response	Alternate
No letter length information provided					
Abstract words	.29 (.14)	.06 (.07)	.06 (.07)	.18 (.16)	.41 (.21)
Concrete words	.68 (.14)	.02 (.04)	.02 (.04)	.08 (.11)	.20 (.08)
Letter length information provided					
Abstract words	.40 (.16)	.04 (.06)	.07 (.08)	.29 (.17)	.20 (.11)
Concrete words	.77 (.12)	.03 (.04)	.02 (.03)	.10 (.08)	.08 (.05)

*SD*s are in parentheses.

Table 2 Probabilities of failure at Steps 1 and 2 (Gollan & Brown, 2006) for abstract and concrete target words in Experiment 1

Condition	No Letter Length Information	Letter Length Information
Probability of Step 1 failure		
Abstract words	.65 (.13)	.56 (.16)
Concrete words	.30 (.14)	.20 (.11)
Probability of Step 2 failure		
Abstract words	.17 (.22)	.08 (.13)
Concrete words	.03 (.05)	.03 (.05)

*SD*s are in parentheses.

Effects of letter length cues The number of words correctly recalled was significantly improved by cues about letter length, $F(1, 54) = 8.77$, $MSE = .03$, $p = .01$. The Cue × Concreteness interaction was not significant, $F < 1$, so no evidence was apparent that letter length cues helped the retrieval of abstract more than of concrete words.

Providing a letter length significantly reduced the number of alternate responses, $F(1, 54) = 31.87$, $MSE = .03$, $p = .001$. The Concreteness × Letter Length interaction was also significant, $F(1, 54) = 7.71$, $MSE = .01$, $p = .008$. Tests of simple main effects revealed that cues significantly reduced alternates for both abstract words, $F(1, 108) = 38.62$, $MSE = .02$, $p = .001$, and concrete words, $F(1, 108) = 9.10$, $MSE = .02$, $p = .003$. The interaction appears to have occurred because letter length information reduced the number of alternates for abstract words more than for concrete words (see Table 1). In order to demonstrate this point formally, an analysis compared the size of the difference between abstract and concrete words in the number of alternates generated as a function of the provision of letter length information. The analysis revealed a significantly larger reduction in the number of alternates for abstract than for concrete words when cues were provided, $F(1, 54) = 6.93$, $MSE = 22.77$, $p = .01$.

Cues had no significant effect on the number of negative TOTs that were reported ($F < 1$). Letter length information increased the number of failures to respond to definitions of abstract words, $F(1, 108) = 8.14$, $MSE = .02$, $p = .005$, but not of concrete words, $F < 1$.

A significant Concreteness × Cues interaction emerged for the probability of Step 2 failure: Cues significantly reduced Step 2 failure for abstract words, $F(1, 108) = 25.99$, $MSE = .01$, $p = .001$, but not for concrete words, $F < 1$.

Discussion

The results of this experiment replicated the findings of Allen and Hulme (2006) in that more concrete than abstract words were correctly named from their dictionary definitions. More importantly, the responses to definitions of

abstract words contained a significantly larger number of failures to respond, a significantly larger number of semantically related alternates, and significantly more positive and negative TOT states. In terms of Gollan and Brown's (2006) argument, abstract words were more likely to be associated with failures at both Step 1 (semantic and lexical access) and Step 2 (access to phonology) of the retrieval process.

The finding that definitions of abstract words elicited significantly more alternates than did definitions of concrete words is consistent with Hanley et al.'s (2004) claim that the semantic–lexical weights are weaker for abstract than for concrete words. This finding is also consistent with Newton and Barry's (1997) claim that there are more semantic competitors for abstract than for concrete words. Nevertheless, the experiment provided two pieces of evidence that it is more difficult for participants to activate to threshold any semantic or lexical representation in response to abstract word definitions.

First, we obtained significantly more failures to respond to definitions of abstract words. This finding is inconsistent with the account put forward by Newton and Barry, because dense semantic neighborhoods containing many semantic competitors should elicit fewer rather than more omissions (Bormann et al., 2008; Dell et al., 2004). Second, when participants were provided with cues as to the number of letters in a word, the number of correct retrievals increased significantly, and the probability of failure at Step 1 was significantly reduced. However, letter length information did not interact significantly with concreteness on either of these two dependent variables, suggesting that letter length information improved performance equally for abstract and concrete words. The significant interaction between these factors on the number of alternates that participants generated indicated that letter length information eliminated more competitors of abstract than of concrete words. So, if abstract target words were hard to generate solely because more semantically related alternates were brought to mind by definitions of abstract words, then letter length information should have significantly reduced the effect of concreteness on the number of target words recalled.

Table 1 suggests an explanation of why letter length information did not help recall of abstract more than of concrete items: It appears that provision of letter length information turned many of the alternate responses for abstract words into omissions rather than correct retrievals. On these trials, it would appear that participants could not access another semantic or lexical representation in response to the definition. Conversely, letter length cues seemed to turn alternates of concrete words into correct answers, suggesting that an alternative semantic or lexical representation for a concrete word (in this case, the target word) could be accessed relatively easily.

In conclusion, as Newton and Barry (1997) claimed, semantic competitors are produced on many trials when

participants attempt to produce abstract words from definitions. However, the fundamental problem appears to be that the abstract word definitions often failed to activate the appropriate lexical representation particularly strongly, consistent with weak semantic–lexical weights for abstract words (Hanley et al., 2004). The consequence of weak activation was significantly more omissions and alternates for abstract than for concrete words. Contrary to the account put forward by Newton and Barry, therefore, the large number of semantic alternates elicited by the definitions of abstract words seems more likely to be a consequence of weak lexical access than evidence of its cause. We will return to this issue in Experiment 2, in which we attempted to directly assess the strength of semantic competition by examining whether the alternates that participants generated fit the definitions of the abstract words better than those of the concrete words.

The increased probability of positive TOTs and of failure at Step 2 of the retrieval process for abstract words suggests that even when the correct semantic and lexical representation has been activated, the phonological representation of an abstract target word is relatively difficult to retrieve. As was explained in the introduction, in a cascade model such as that of Foygel and Dell (2000), it might be the case that the Step 2 problems with abstract words are an inevitable consequence of weaker activation of the lexical representations for abstract target words during Step 1 (Hanley et al., 2004). An alternative explanation of reduced phonological activation is that abstract words are associated with independent problems at both Steps 1 and 2 of the word production process, as appears to be the case for words of low frequency (see Kittredge, Dell, Verkuilen, & Schwartz, 2008). In a model such as that put forward by Burke and Shafto (2004), for example, abstract words might suffer from a phonological transmission deficit whereby lexical-to-phonological weights are significantly weaker for abstract than for concrete words, over and above the problems with lexical access for abstract words. Experiment 2 was designed to investigate this issue further.

Probability of Step 2 failure for abstract words was significantly reduced by letter length information. In previous research, Heine, Ober, and Shenaut (1999) showed that when combined with initial letters, letter length information resolved 34 % of the reported TOTs elicited by definitions. It appears that letter length information facilitated phonological access in Experiment 1 also, although the precise mechanism by which this occurred is still an open question. We found no similar effect with definitions of concrete words, presumably because they elicited so few Step 2 failures, even when letter length information was not provided. It is also noteworthy that letter length information did not reduce the number of negative TOTs reported for abstract words. Presumably, letter length information is not useful in

resolving a TOT state in these circumstances because the word that elicits a negative TOT often contains a different number of letters from the target word. This finding also suggests that participants do not have access to the number of letters in the word that is eliciting a negative TOT; if they did, they would realize that it could not be the target word.

Experiment 2

In Experiment 2, we attempted to design a word production task in which it would be easier to access the semantic and lexical representations of abstract words than it had been from their definitions in Experiment 1. Crutch and Warrington (2005) argued that concrete words tend to be represented in memory by a particular set of semantic features. Conversely, they claimed that abstract words are more likely to be represented in semantic memory in terms of their associative connections with other words. Consistent with these claims, a series of neuropsychological investigations and a number of studies with unimpaired participants have shown that abstract words are more efficiently processed in associative contexts, whereas concrete words are better processed in the context of semantically similar items such as members of the same semantic category (e.g., Crutch, Connell, & Warrington, 2009; Crutch & Jackson, 2011).

If it is assumed that retrieval from definitions depends heavily on an attempt to recall a word on the basis of its semantic features, it follows that the naming-to-definition task would likely prove particularly difficult for abstract words. The way in which words are retrieved during normal language production, however, is in the context of a sentence in which there are often rich associative connections between each word and the sentence context in which it appears. In Experiment 2, therefore, we gave participants sentences that described events in which an abstract word was missing (e.g., *In court, Joan entered a _____ of "not guilty" to the crime she was accused of*) and sentences in which a concrete word was missing (e.g., *Jason played a game of _____ on the famous links*). Participants were asked to try and generate the word that best fit into that sentence (*plea* and *golf*, respectively, in the sentences above).

We assume that the retrieval of a word's lexical representation from an event context such as these will reflect a greater influence of associative processing and a weaker influence of semantic similarity than is the case for the definition task. If so, it then follows from Crutch and Warrington (2005) that the problems associated with the production of abstract words (fewer correct retrievals and increased probability of Step 1 failure relative to concrete words) may be reduced or eliminated when they have to be retrieved from an event context. Should the differences between the probabilities of Step 1 failure be reduced

significantly, a key question would then be whether the probability of phonological retrieval failure (Step 2) would remain substantially greater for abstract than for concrete words. If the connection between the lexical and phonological levels in the production system is a primary problem for abstract words, then significant phonological retrieval problems would remain even if the Step 1 problems with abstract words were resolved. Conversely, if the phonological access problems for abstract words occur because the lexical representations of such words are activated strongly enough to elicit a TOT but not the target word, then higher levels of Step 2 retrieval failure for abstract words should no longer be observed.

Method

Participants A group of 58 undergraduate students at the University of Essex took part in the experiment. None of them had been participants in Experiment 1. Half were randomly allocated to the definition condition and half to the event condition.

Materials For participants in the definition condition, the presentation was the same as in Experiment 1, except that none of the participants received any letter length information. For participants in the event condition, the definitions used in Experiment 1 were replaced by new sentences that described a simple event. These sentences were generated so as to create a context in which the target word was highly predictable. However, no attempt was made to match the degree of lexical association between the target and any of the words in the concrete and abstract event sentences. The sentences were given to six new participants with the target word underlined to ensure that the concrete and abstract words were equally congruent with their sentence contexts. These participants were therefore asked to indicate on a scale of 1–5 how well each underlined target word fit into the sentence in which it appeared (1 = *very poorly*, 5 = *very well*). The overall mean (with *SD*) for the concrete sentences was 4.75 (0.23), and the overall mean for the abstract

sentences was 4.73 (0.26). These means did not differ significantly, $t(66) = 0.33$ $p = .742$. These suitability ratings were very similar to those provided for the definitions by Allen and Hulme's (2006) participants (see Exp. 1 for details). The Appendix provides details of all of the event sentences that were used in Experiment 2. Three new words that were similar in meaning to each target word were generated specifically for use as recognition test foils in this experiment.

Procedure Participants were given a set of 68 sentences in which a word was missing. They were told that their task was to produce the word that best fit into that sentence context. If they were unable to recall the word, they were asked to indicate whether they were experiencing a TOT state, exactly as in Experiment 1. At the end of the experiment, the participants undertook a recognition test for those items for which they had experienced a TOT. They heard the sentence and were shown four words on a computer screen. For example, the foils for *dare* were *provoke*, *bully*, and *goad*, and the foils for *lake* were *pool*, *lagoon*, and *spring*. They were asked to indicate whether any of the four words was the item that had previously elicited their TOT state. It was emphasized that their task was not to indicate which was the item that they now believed best fit the definition. If the participants selected the target item, the TOT was considered to be a positive TOT. If they selected a foil or stated that none of the four words had elicited their TOT, it was considered a negative TOT.

Results

Number correct Performance in Experiment 2 is summarized in Tables 3 and 4. There was a significant effect of concreteness, $F(1, 58) = 127.31$, $MSE = .05$, $p = .00001$, and significantly more words were correctly recalled from event than from definition sentences, $F(1, 58) = 50.51$, $MSE = .05$, $p = .001$. The interaction was also highly significant, $F(1, 58) = 16.01$, $MSE = .05$, $p = .001$. Tests of simple main effects showed a significant effect of concreteness in both the

Table 3 Mean proportions of different types of responses for abstract and concrete target words in Experiment 2

Condition	Correctly Named	Positive TOTs	Negative TOTs	No Response	Alternate
Definition sentences					
Abstract words	.28 (.11)	.06 (.05)	.11 (.12)	.21 (.12)	.35 (.10)
Concrete words	.55 (.17)	.06 (.07)	.05 (.07)	.10 (.10)	.23 (.08)
Event sentences					
Abstract words	.60 (.14)	.03 (.04)	.02 (.03)	.05 (.08)	.30 (.08)
Concrete words	.71 (.16)	.03 (.04)	.02 (.04)	.06 (.09)	.19 (.09)

*SD*s are in parentheses.

Table 4 Probabilities of failure at Steps 1 and 2 (Gollan & Brown, 2006) for abstract and concrete target words in Experiment 2

Condition	Definition Sentences	Event Sentences
Probability of Step 1 failure		
Abstract words	.67 (.10)	.37 (.12)
Concrete words	.39 (.13)	.26 (.15)
Probability of Step 2 failure		
Abstract words	.19 (.19)	.06 (.10)
Concrete words	.10 (.14)	.04 (.06)

*SD*s are in parentheses.

definition, $F(1, 16) = 116.81$, $MSE = .05$, $p = .001$, and event, $F(1, 16) = 26.51$, $MSE = .05$, $p = .001$, conditions. The interaction seems to have come about because the concreteness effect was larger in the definition than in the event condition. To demonstrate this point formally, an analysis compared the sizes of the advantage for concrete words in the two conditions. This revealed a significantly larger advantage for concrete words in the definition than in the event condition, $F(1, 58) = 28.95$, $MSE = 17.63$, $p = .001$.

Step 1 failure We found a significantly higher probability of Step 1 failure for abstract than for concrete words, $F(1, 58) = 176.53$, $MSE = .04$, $p = .001$, and a significantly higher probability of Step 1 failure in the definition than in the event condition, $F(1, 58) = 49.63$, $MSE = .04$, $p = .001$. There was also a significant interaction, $F(1, 58) = 45.10$, $MSE = .04$, $p = .001$. As with the number of correct responses, the interaction seems to reflect a stronger effect of concreteness in the definition than in the event condition.

Significantly fewer alternates were reported in the event than in the definition condition, $F(1, 58) = 7.70$, $MSE = .01$, $p = .007$. As in Experiment 1, more alternates were produced in response to sentences about abstract than about concrete words, $F(1, 58) = 72.84$, $MSE = .01$, $p = .001$. The interaction between these factors was not significant.

Significantly fewer failures to respond occurred in the event than in the definition condition, $F(1, 58) = 17.87$, $MSE = .02$, $p = .001$, and significantly more failures to respond occurred for abstract than for concrete words, $F(1, 58) = 23.73$, $MSE = .02$, $p = .001$. The interaction was significant, $F(1, 58) = 29.93$, $MSE = .02$, $p = .001$, with tests of simple main effects revealing a concreteness effect in the definition condition, $F(1, 116) = 53.48$, $MSE = .01$, $p = .001$, but not in the event condition, $F < 1$.

We found significantly more negative TOTs in the definition than in the event condition, $F(1, 58) = 14.11$, $MSE = .01$, $p = .001$. The effect of concreteness on negative TOTs was also significant, $F(1, 58) = 9.20$, $MSE = .01$, $p = .004$, as was the interaction, $F(1, 58) = 8.05$, $MSE = .01$, $p = .006$.

Tests of simple main effects revealed a significant effect of concreteness in the definition, $F(1, 116) = 17.23$, $MSE = .01$, $p = .001$, but not in the event, $F < 1$, sentences.

Step 2 failure Most importantly, we found a significantly higher probability of Step 2 failure for abstract than for concrete words $F(1, 58) = 9.43$, $MSE = .03$, $p = .003$, and significantly higher probability of Step 2 failure in the definition than in the event condition, $F(1, 58) = 9.59$, $MSE = .04$, $p = .003$. The interaction was also significant, $F(1, 58) = 4.63$, $MSE = .03$, $p = .04$: Tests of simple main effects showed a significant effect of concreteness with definitions, $F(1, 116) = 13.64$, $MSE = .02$, $p = .001$, but not with event sentences. Significantly more positive TOTs were reported in the definition condition than in the event condition, $F(1, 58) = 6.82$, $MSE = .02$, $p = .01$, but no effect of concreteness and no significant interaction appeared, both Fs < 1.

Word category Although Allen and Hulme's (2006) set of abstract and concrete words were matched on a number of important variables, they were not matched for word category. Whereas all of the concrete words were nouns, the abstract words comprised nine adjectives, 11 nouns, and 14 verbs. We therefore investigated whether we could find any evidence in Experiment 2 that the observed differences in performance were caused by word category rather than concreteness.

In the definition condition, the probabilities of Step 1 failure were .78 for abstract adjectives, .72 for abstract nouns, and .57 for abstract verbs. These values are all much higher than the probability of Step 1 failure for concrete words (.39). The probabilities of Step 2 failure were .17 for abstract adjectives, .22 for abstract nouns, and .10 for abstract verbs, so the value for Step 2 failure for abstract nouns was descriptively higher than that for concrete words (.10).

In the event condition, the probabilities of Step 1 failure were .36 for abstract adjectives, .32 for abstract nouns, and .41 for abstract verbs, as compared to .26 for the concrete words. Although these results suggest that the event condition may have been particularly beneficial for abstract adjectives and nouns, evidence remains of an advantage for concrete relative to abstract nouns, verbs, and adjectives. The probabilities of Step 2 failure in the abstract condition were .02 for adjectives, .06 for nouns, and .03 for verbs, as compared to .04 in the concrete word condition.

Strength of semantic competitors An attempt was made to investigate whether the abstract and concrete target words faced equally strong semantic competitors during word production in Experiment 2. Additional data were therefore collected to determine how well the most frequent alternates to the target words fit into the event and definition sentences. A

group of 16 new participants, drawn from the same population as those who took part in the main experiment, were shown each of the definition sentences together with three words. One of the words was the target word, and the other two words were the most frequent alternates produced by participants in the definition condition in the main experiment. The participants were asked to indicate how well each of these words fit into its corresponding sentence on a scale of 1–5, where 5 = *very well* and 1 = *very badly*. A further 15 new participants were shown the target word and the two most frequent alternates produced by participants for the event sentences in the main experiment and rated how well each of these words fit into its corresponding sentence, on the same 1–5 scale.

The mean ratings are shown in Table 5. In the event condition, a significant effect of type of word emerged, $F(2, 66) = 162.2$, $MSE = .54$, $p = .001$, but no effect of concreteness, $F < 1$, and no significant interaction, $F < 1$. Newman–Keuls tests revealed that the target words received significantly higher ratings than did the highest-rated alternate, which in turn received higher ratings than the second alternate (both $ps < .01$). In the definition condition, we found a significant effect of concreteness, $F(1, 66) = 14.65$, $MSE = .57$, $p = .001$, a significant effect of type of word, $F(2, 66) = 150.81$, $MSE = .57$, $p = .001$, and a significant interaction, $F(2, 66) = 15.71$, $MSE = .54$, $p = .001$. Newman–Keuls tests revealed no significant effect of concreteness on the ratings given to the concrete and abstract target words. However, the alternate words in the abstract sentences were given significantly higher ratings than were the alternate words in the concrete condition ($p < .01$).

What these findings reveal is that the abstract and concrete target words fit equally well into the definition sentences and into the event sentences. Such an outcome is consistent with the normative data collected for the event sentences prior to Experiment 2, as well as with the normative data collected for the definitions by Allen and Hulme (2006). No evidence emerged from the analyses that the abstract target words faced greater competition from alternates than did the concrete words in the event sentences. The situation was somewhat different for the definition sentences, however. Here, the alternates for abstract target

Table 5 Participants' ratings of how well the target word and alternates fit into the definition and event sentences (1 = *very badly*, 5 = *very well*)

Condition	Target	1st Alternate	2nd Alternate
Definitions			
Abstract words	4.24	3.62	2.80
Concrete words	4.59	2.91	1.94
Events			
Abstract words	4.71	3.25	2.15
Concrete words	4.70	3.25	2.44

words were seen as fitting the definitions better than the alternates to the concrete words did . It therefore appears that abstract words in the definition condition may face stronger competition from alternates than do the concrete target words.

Discussion

In Experiment 2, the effect of concreteness on the proportion of words correctly retrieved was significantly smaller when participants were asked to produce words from event sentences than when they were asked to produce words from dictionary definitions. The use of event sentences seemed, therefore, to have had the desired effect of substantially improving the probability that the semantic and lexical representations of abstract words would be successfully activated relative to concrete words. The critical issue was whether differences would remain between abstract and concrete words in terms of Step 2 (phonological) retrieval.

As in Experiment 1, we again found a significantly greater probability of Step 2 failure for abstract than for concrete words in the definition task. One possibility is that this effect arose because of weaker connections between the lexical and phonological levels in the word production system for abstract than for concrete words. If this was indeed the case, concreteness should have exerted an effect on the probability of Step 2 failure when the task was to retrieve words from event scenarios. However, the results of Experiment 2 revealed no significant effect of concreteness on the probability of Step 2 failure in the event condition. It therefore appears that the increased probability of Step 2 failure for abstract words in the definition condition was a direct consequence of the weaker activation of an abstract word's lexical representation from its dictionary definition. When the semantic and lexical activation of abstract words was increased by presenting event sentences instead of definitions, abstract words were no longer associated with a greater probability of Step 2 retrieval failure.

It might be argued that this interaction came about because of a floor effect in the number of TOTs in the event condition. However, Table 4 shows that the probability of Step 1 failure for abstract words in the event condition was similar to that of Step 1 failure for concrete words in the definition condition. This suggests approximately equivalent levels of lexical access in these two conditions. If there were an independent phonological retrieval problem for abstract words, then a greater probability of Step 2 failures for abstract than for concrete words should have been observed in these two conditions. As Table 4 shows, this was clearly not the case. Experiment 2, therefore, provided no evidence for an independent problem with abstract words at Step 2 of the word production process, as appears to be the situation for words of low frequency (Kittredge et al., 2008).

Nevertheless, it must be acknowledged that for the present study we used only binary measures. One cannot rule out the possibility that more continuous measures, such as reaction times, might reveal more subtle effects of concreteness on Step 2 retrieval.

The difference that remained between the correct retrieval of abstract and concrete words in the event condition was directly related to the larger number of alternates that participants generated for abstract words. A possible explanation for the large number of remaining alternates is that the abstract words may have had more synonyms than the concrete words. However, the additional normative data collected at the end of Experiment 2 provided no evidence that the most common alternates for abstract words in the event condition were more closely related to the meaning of the event sentences than were the most common alternates for the concrete words. Because we found no evidence that abstract words faced more competition from competitors in the event condition, it therefore appears that the greater number of alternates produced in the abstract word condition is a consequence of the failure to access the correct lexical representation as a result of weak semantic–lexical weights (Hanley et al., 2004).

In the definition condition, however, participants did rate the most common alternatives to the abstract target words as being more compatible with the definitions than the most common alternatives to the concrete target words. This observation is consistent with Newton and Barry's (1997) claim that abstract words are difficult to retrieve in the definition task because they face stronger competition from semantic neighbors. It seems unlikely, however, that this is one of the reasons why abstract words were worse recalled. If this had been the case, a larger effect of concreteness on the number of alternates produced should have been observed in the definition condition than in the event condition. In fact, though, this interaction failed to approach significance. The poor retrieval of abstract relative to concrete words in Experiment 2 is therefore explicable solely in terms of weaker semantic–lexical weights for abstract words (Hanley et al., 2004); there is no need to suggest that interference from competitors of abstract words plays an important additional role. As in Experiment 1, the evidence suggests that the increased number of alternates is a consequence rather than a cause of poor lexical retrieval of abstract words. Nevertheless, the finding that the alternates of abstract words were considered to fit the definitions so well merits further investigation. No independent formal evidence has been found that abstract words have more synonyms, although the use of latent semantic analysis (Hoffman, Rogers, & Lambon Ralph, 2011) has clearly established that abstract words have more senses, and therefore are more ambiguous than concrete words. This would be an interesting issue for future research on latent semantic analysis to address.

Although we do not have any direct measure of the degree of association between the target words and the event sentences, it seems reasonable to assume that recalling words in response to specific events is likely to require more associative processing and less featural processing than recalling words from their definitions. Consequently, the significantly greater improvement in the retrieval of abstract than of concrete words in the event condition provides some support for the views of Crutch and Warrington (2005) that abstract words tend to be represented in semantic memory in terms of their associative connections with other words rather than of their semantic features. Nevertheless, recall of concrete words did improve to some extent in the event condition. So, as Crutch and Warrington maintained, any differential dependence of concrete and abstract words on associative and featural information appears to be relative rather than absolute. An alternative explanation of the greater improvement for abstract words would be the presence of a ceiling effect in performance for concrete words in the event condition. Although it is impossible to entirely dismiss this possibility, the similarity of the standard deviations for concrete words correctly recalled from the definition and event sentences suggests that performance had not reached ceiling in the event condition (see Table 3).

Finally, the Experiment 2 Results section provided some evidence that word category might have had an influence on performance in Experiment 2. For example, although Step 1 retrieval in the event condition was consistently more difficult for abstract than for concrete words regardless of word category, the effect was descriptively smaller for nouns and adjectives than for verbs, and it appeared that Step 1 access might be more difficult for abstract verbs than for abstract nouns or adjectives. This outcome must be treated with caution, as the items from these different word categories were not equated on variables such as frequency or age of acquisition. Nevertheless, the effects of word category on lexical retrieval in the event and definition tasks would be an interesting issue for future research to address.

Conclusion

In conclusion, the results of this study have shown that poor performance during attempts to retrieve abstract words from their dictionary definitions (Allen & Hulme, 2006) is associated with more omissions, more alternates, and more TOTs than is the case for concrete words. These findings indicate that problems are associated with abstract words at both Step 1 (semantic–lexical) and Step 2 (phonological) (Gollan & Brown, 2006) of the word production process during performance of this task. These results can all be explained by the assumption (Hanley et al., 2004) that semantic–lexical weights in the word production system

are stronger for concrete than for abstract words. When semantic and lexical access was improved for abstract words in Experiment 2 by the use of event scenarios rather than definitions, the phonological problems disappeared. This finding was interpreted as evidence that lexical–phonological weights are as strong for abstract as for concrete words. Consistent with the views of Newton and Barry (1997), we found evidence that abstract words face stronger competitors than do concrete words on the definition task. No evidence was apparent, however, that the greater activation of competitors is a cause of poor abstract word retrieval.

Appendix

Event sentences used in Experiment 2

Concrete words:

(ash)	The city was covered in _____ after the volcano started to erupt.
(bar)	Stacey used a _____ of soap to wash her hands before eating her lunch.
(bomb)	Disposal experts were asked to detonate a _____ from World War 2.
(branch)	James hung the swing from a _____ of the tree in his back garden.
(card)	John signed a _____ for his girlfriend Lisa's birthday.
(coat)	When it got cold, Phil buttoned up his _____ to keep himself warm.
(cream)	After their dinner, Frank and Laura had dessert of strawberries with _____.
(deck)	In the emergency, the ship's captain called for all passengers to assemble on the _____.
(drug)	The doctor prescribed Sarah a _____ that had just passed its clinical trials.
(edge)	When Emma reached the top, she could stand at the _____ of the cliff.
(gate)	Henry came up the path and opened the_____ to enter his front garden.
(gin)	Fred poured chilled tonic water into his glass of ____.
(golf)	Jason played a game of _____ on the famous links.
(hen)	Anne's _____ laid a large number of eggs for her to sell.
(horn)	While trying to escape the angry bull, the tip of it's _____ caught on Sean's shirt.
(jail)	Will went to _____ after being caught breaking the law.
(jet)	The fighter aircraft was powered by the latest _____ engine.
(key)	Jane used the _____ to open the door to Wendy's house.
(knee)	The footballer tore the cartilage in his _____, which ended his career.

(lake)	Nicola stayed in a log cabin which was by a _____ where she would go swimming.
(lift)	Peter entered the _____ to go up to the 10th floor to his hotel room.
(lock)	Frank put a _____ on the door to secure the room from his houseguests.
(map)	After taking a wrong turning, Paul used a _____ to find his way home.
(moon)	When walking home at night, Alan saw the _____ and stars come out.
(net)	The fisherman was pleased when he saw the number of fish caught in his _____.
(pipe)	The water gushed down the _____ to fill up the drain.
(plate)	Katie put too much food on her _____ and could not finish it all.
(plug)	The TV did not work because the _____ was not in the socket.
(rake)	The gardener used a _____ to gather up the leaves and tidy the garden.
(ranch)	Hank kept all his cattle on a _____ in Texas.
(tape)	In 1950, the speech was recorded onto a _____ so that it could be heard many times in the future.
(tent)	Julie and Ben slept in a _____ when they went camping.
(van)	Patrick had to hire a _____ to move some large furniture to his new flat.
(wave)	A _____ swept Dan overboard into the sea.

Abstract words:

(aim)	Bill had to _____ the loaded gun at the target.
(bid)	Rick placed a _____ for a painting at the auction.
(blame)	Jenny did not _____ Mike for the accident, even though she had been injured.
(brave)	For being _____ while getting an injection, Sammy's mum bought him sweets.
(calm)	Throughout the crisis, Sue remained _____ and clear-headed.
(choose)	Harry's tough decision was to _____ which girl to take to the cinema.
(cope)	Mary was unable to _____ with her children when they were being naughty.
(dare)	As a _____, Jim tried to walk on the electric train track.
(drop)	The fielder was embarrassed to _____ the cricket ball.
(ease)	Kenny took a soothing tablet to _____ the pain in his head.
(glad)	Dawn was _____ to be home from her holiday because she missed her friends and family.
(grade)	Charlie was awarded a _____ of B in his Maths A-Level.
(halt)	The car came to a_____ at the stop sign.

(harm)	The dog did not mean them any _____, but it scared them when it barked.
(hate)	Tina was angry and felt she would always_____ the man who ruined her life.
(joy)	Sally shed tears of _____ when she saw her baby smile for the first time.
(lose)	During an argument, Kerry did not want to _____ her temper so she walked away.
(mad)	Hattie thought she was going _____ when she saw the chaos all around her.
(mood)	Eric was feeling sad but some good news changed his _____.
(neat)	When sewing, Ayesha always took her time to make sure all her stitches were _____ and tidy.
(nice)	George's mother was pleased that his new girlfriend seemed to be a_____ girl
(pause)	There was a ____ for reflection before the decision was taken.
(plea)	In court, Joan entered a _____ of "not guilty" to the crime she was accused of.
(pull)	In the old house, David had to _____ the chain to flush the toilet.
(rule)	There was a _____ in Tammy's house that everyone had to do their daily chores.
(rush)	The traffic on the roads was very heavy in _____ hour.
(save)	Lucy had to jump from the window in order to _____ herself from the fire.
(slow)	Kevin had to _____ his car to get round the bend.
(sold)	Amy and Joe ____ their house for a big profit.
(teach)	Jeff wanted to _____ his son a lesson so he grounded him for his bad behaviour.
(trust)	Michelle felt she could not _____ Nick again after he was caught cheating.
(vague)	The police had only a _____ description of the suspect.
(wise)	In a fable, the forest animals thought the owl was a _____ bird, so they asked it for advice.
(worth)	Margaret took her watch to be valued because she wanted to know what it was _____.

References

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

Bormann, T., Kulke, F., Wallesch, C. W., & Blanken, G. (2008). Omissions and semantic errors in aphasic naming; Is there a link? *Brain and Language, 104,* 24–32.

Brown, A. S. (2012). *The tip of the tongue state.* Hove, U.K.: Psychology Press.

Burke, D. M., & Shafto, M. A. (2004). Aging and language production. *Current Directions in Psychological Science, 13,* 21–24.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology. General, 132,* 163–201. doi:10.1037/0096-3445.132.2.163

Crutch, S. J., Connell, S., & Warrington, E. K. (2009). The different representational frameworks underpinning abstract and conceptual knowledge: Evidence from odd-one-out judgements. *Quarterly Journal of Experimental Psychology, 62,* 1377–1390. doi:10.1080/17470210802483834

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

Crutch, S. J., & Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain, 128,* 615–627.

Dell, G., Lawler, E., Harris, H., & Gordon, J. (2004). Models of errors of omission in aphasic naming. *Cognitive Neuropsychology, 21,* 125–145. doi:10.1080/02643290342000320

Foygel, D., & Dell, G. S. (2000). Models of impaired lexical access in speech production. *Journal of Memory and Language, 43,* 182–216.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language & Cognitive Processes, 21,* 817–855.

Gollan, T. H., & Brown, A. S. (2006). From tip-of-the-tongue (TOT) data to theoretical implications in two steps: When more TOTs means better retrieval. *Journal of Experimental Psychology. General, 135,* 462–483. doi:10.1037/0096-3445.135.3.462

Hanley, J. R. (2011). Why are names of people associated with so many phonological retrieval failures? *Psychonomic Bulletin & Review, 18,* 612–617. doi:10.3758/s13423-011-0082-0

Hanley, J. R., Dell, G. A., Kay, J., & Baron, R. (2004). Evidence for the involvement of a nonlexical route in the repetition of familiar words; A comparison of single and dual route models of auditory repetition. *Cognitive Neuropsychology, 21,* 147–158.

Hanley, J. R., & Kay, J. (1997). An effect of imageability on the production of phonological errors in auditory repetition. *Cognitive Neuropsychology, 14,* 1065–1084.

Hanley, J. R., & Turner, J. M. (2000). Why are familiar-only experiences more frequent for voices than for faces? *Quarterly Journal of Experimental Psychology, 53A,* 1105–1116.

Heine, M. K., Ober, B. A., & Shenaut, G. K. (1999). Naturally occurring and experimenter induced tip-of-the-tongue experiences in three adult age groups. *Psychology and Aging, 14,* 445–457.

Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience, 23,* 2432–2446. doi:10.1162/jocn.2011.21614

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007

Johnston, R. A., Dent, K., Humphreys, G. W., & Barry, C. (2010). British-English norms and naming times for a set of 539 pictures: The role of age of acquisition. *Behavior Research Methods, 42,* 461–469. doi:10.3758/BRM.42.2.461

Jones, G. V. (1985). Deep dyslexia, imageability, and ease of predication. *Brain and Language, 24,* 1–19.

Kittredge, A. K., Dell, G. S., Verkuilen, J., & Schwartz, M. F. (2008). Where is the frequency effect in word production? Insights from aphasic picture naming errors. *Cognitive Neuropsychology, 25,* 463–492.

Laine, M., Tikkala, A., & Juhola, M. (1998). Modelling anomia by the discrete two-stage word production architecture. *Journal of Neurolinguistics, 11,* 275–294.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences, 22,* 1–38. disc. 38–75.

Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory & Cognition, 20,* 715–726.

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,* 65–79.

Newton, P. K., & Barry, C. (1997). Concreteness effects in word production but not word comprehension in deep dyslexia. *Cognitive Neuropsychology, 14,* 481–509.

Nozari, N., Dell, G. S., & Schwartz, M. F. (2011). Is comprehension necessary for error detection? A conflict-based account

of monitoring in speech production. *Cognitive Psychology, 63,* 1–33.

Paivio, A. (1986). *Mental representations*. New York, NY: Oxford University Press.

Perfect, T. J., & Hanley, J. R. (1992). The tip-of-the-tongue phenomenon: Do experimenter-presented interlopers have any effect? *Cognition, 45,* 55–75. doi:10.1016/0010-0277(92)90023-B

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

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 82–102. doi:10.1037/0278-7393.9.1.82

Schwartz, B. L., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. *Memory & Cognition, 39,* 737–749. doi:10.3758/s13421-010-0066-8