Are word association responses really the first words that come to mind?

PLAYFOOT, David <http://orcid.org/0000-0003-0855-334X>, BALINT, Teodor, PANDYA, Vibhuti, PARKES, Averil, PETERS, Mollie and RICHARDS, Samantha

Published version

PLAYFOOT, David, BALINT, Teodor, PANDYA, Vibhuti, PARKES, Averil, PETERS, Mollie and RICHARDS, Samantha (2016). Are word association responses really the first words that come to mind? Applied Linguistics.

Are word association responses really the first words that come to mind?

Journal:	*Applied Linguistics*
Manuscript ID	APPLING-15-08-246.R1
Manuscript Type:	Article
Keyword:	psycholinguistics, Methodology, Word association , Working memory, Lexical organisation

Are word association responses *really* the first words that come to mind?

Abstract

Word association has been a popular tool for research in linguistics and psychology over the last century. The paradigm presents participants with a cue word and asks them to respond with the first associated word that comes to mind. Inferences about the structure and organisation of the lexicon have been made on the basis of the findings of word association tasks, and on the assumption that responses reflect the strongest link between words in the participants' vocabulary. The procedure adopted in traditional word association tasks does not guarantee that this is the case. This paper presents two experiments that aimed to determine whether or not participants make deliberate and strategic responses in word association tasks. Findings indicate that word association responses are likely to reflect the first word that participants activate in their lexicon.

Word association has long been used as a tool for assessing the organization of the lexicon in a variety of populations, including monolingual speakers (e.g. Fitzpatrick, Playfoot, Wray & Wright, 2013; Hirsh & Tree, 2001; Playfoot & Izura, 2013), bilinguals (e.g. Fitzpatrick, 2007; Meara, 2009) and in clinical presentations (e.g. Gewirth, Shundler & Hier, 1984; Gollan. Salmon & Paxton, 2006; Merten, 1993). **To use word association data in this way**, a number of assumptions have been made about the nature of the *responses* that participants give in these types of task. In what follows, we discuss perhaps the most important issue in this type of research - whether or not participant responses are likely to represent the first word that is activated by a word association cue. Although this assumption has gained some (often indirect) empirical support, to our knowledge it has not yet been tested directly.

Estimates of the number of words that an adult knows in their native language vary considerably from study to study, and vocabulary sizes of anywhere between 14,418 (Nusbaum, Pisoni & Davis, 1984) and more than 200,000 words (Hartmann, 1941 - **though it should be noted that this is far higher than most estimates**) have been reported. Given that vocabulary size is often calculated relative to printed corpora (which typically under-represent proper nouns, slang, acronyms etc, see Brysbaert & New, 2009) it is likely that actual vocabulary size exceeds the published estimates. In spite of having such a large number of words to choose from, skilled readers are able to find and produce the appropriate word quickly when they are asked to name an object or in fluent speech. The prevailing opinion is that this is because word knowledge is stored as nodes in an interconnected semantic network (e.g. Collins & Quillian, 1969; Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005). In Steyvers and Tenenbaum's (2005) model, for example, each word node is connected to any number of other words by links that vary in strength. When a word is encountered (or *activated*) some activation is also passed along each intra-lexical link that stems from the stimulus. The amount of activation that passes (or *spreads*) to each connected word is determined by the strength of the link, which in turn is determined by personal experience. The more times a particular link between two words is traversed, the stronger it becomes. The consequence of this spreading activation is that the presentation of one word can lead to increased likelihood of producing a related word soon afterward. For example, in a classic experiment, Meyer and Schvaneveldt (1971) presented participants with targets for lexical decision (i.e. does this combination of letters represent an existing word) that were immediately preceded by a prime word. In some trials, the prime word was related in meaning to the target (e.g. doctor-NURSE). In other trials the prime and target were unrelated (e.g. doctor-BREAD). Meyer and Schvaneveldt (1971) reported that participants were significantly faster to respond to related targets than unrelated targets. In relation to spreading activation theory, the presentation of "doctor" passed some activation along the connection to NURSE. As a consequence, when NURSE was itself presented, it was already partially activated and took less additional effort to reach a recognition threshold. The same is not true for BREAD, which is not connected to "doctor" in the lexicon. It has since been demonstrated that the presentation of a prime word automatically activates all the connected words in the lexicon, whether the connection is semantic, associative (e.g. Ferrand & New 2003), or formal (e.g. Davis & Lupker, 2006) even if exposure to the prime is short enough to prevent conscious processing. The key principle here is that strong links between words can be accessed quickly and automatically.

In discrete word association tasks, participants are presented with a single cue word and required to say or write down the first word that comes to mind. It has been argued that the first word that comes to mind ought to reflect access of the strongest intra-lexical link (e.g. Playfoot & Izura, 2013). That is, the word association cue acts as a prime for the response - once the cue has been activated in the lexicon of the participant, activation will spread to surrounding nodes according to the weights of the connections. The word the participant produces will be the first node to reach a criterion level of activation, and this will be accrued more quickly along strong than weak links. Indeed, much of the word association research in the literature proceeds from this assumption. **One notable exception to this is the work of Wettler, Rapp and Sedlmeier (2005), who argue that contiguities in the**

presentation of words in the speaker's language are the key determinant of word association responses. They examined the responses elicited by the presentation of 100 cue words (those used by Kent & Rosanoff, 1910) and compared these to the probability that the cue and response co-occurred in sentences (using the British National Corpus, BNC, as the source of this information). They argued that the probability of co-occurrence of a pair of words corresponded well with the responses that were given by human participants in a word association task (though this correspondence was far from perfect), and suggested that word association responses could be explained by paired associative learning processes. They acknowledged, however, that their findings did not disprove the theory that there were semantic structures underpinning word association behaviour. In fact, it could be that the contiguities observed in the BNC are a crude measure of the links between words in the lexical network - the co-occurrence of two words strengthens the intra-lexical link between them in Steyvers and Tenenbaum's (2005) model, so words that are more likely to co-occur in the BNC are also likely to have strong links in the lexicon.

Published norms lists (e.g. Fitzpatrick et al., 2013; Postman & Keppel, 1970; Nelson, McEvoy & Schreiber, 1998) present the same word association cues to large numbers of participants and organise the responses according to the frequency with which they occur within the sample population. This is often converted to a metric called *associative strength.* To do so, the frequency of a particular response is divided by the total number of responses to that cue to create a proportion of participants who produce the same word. For example, if 58 people out of 100 say *white* when presented with the cue word BLACK, the associative strength is 0.58.

This metric is essentially an approximation of the how strong the intra-lexical link between black and white is across participants. **It is important to remember that associative strength is a metric of the connections that are made between words at a group level, and that the lexical structure of any given individual is unlikely precisely match the idealised picture provided by associative strength. Nevertheless, Canas (1990) demonstrated that** associative strength is a good predictor of the size of priming effects - prime-target pairs with greater associative strength elicit quicker responses than pairs with lower estimates of associative strength. This finding provides indirect support for the notion that word association responses are the first word that comes to mind following the cue, in that both the word association response and the priming effect are supposed to rely on activation spreading along the same intra-lexical link.

Further, weaker, indirect support for the assumption that word association responses are the first word that comes to the mind of the participant has been provided by a handful of studies that have presented the same cue words to a group of participants on two separate occasions (e.g. Fitzpatrick, 2007) or in two different languages (Fitzpatrick, 2007; 2009; Fitzpatrick & Izura, 2011). In such studies, the types of responses that are given by an individual participant are fairly stable over time and across presentations in their first and second languages - certainly more consistent than would be the case if there was not a common mechanism underpinning performance on the task each time it was performed. However, it is far from certain whether this commonality of responses is because of the automatic retrieval of a particularly strong, and stable, intra-lexical link between words or because the responses that they offer are governed by the application of a consistent strategy for performing the task.

Word association tasks typically allow participants plenty of time in which to generate their responses, and are commonly presented as pen and paper measures so reaction time data is not available. This raises a potential concern - researchers have no way of preventing participants from deliberately selecting a response from a number of possible options, and no measure by which to examine whether this is likely to be happening. To explain further, let us start by assuming that the response that a participant gives in a word association task *is not* the first word that comes to mind. Presumably the decision as to which word to choose must take some time. This is because a) you must allow time for multiple associates to be activated and b) you must then apply some kind of decision-making heuristic to determine which word is the most appropriate for the task at hand. As an analogy, consider buying milk in the shop. Your hand does not have to go as far to reach the bottle at the front of the shelf as it does to reach the bottle at the back of the shelf. There is therefore a difference in the time it takes to pick up the two bottles. In addition, to reach the bottle at the back of the shelf you need to move the first bottle out of the way. This adds a step to your milk-buying process and there is a time cost. Therefore responses *other than* the first word that come to mind are likely to be offered at longer response latencies - the traditional lack of RT data does not allow this to be assessed.[1]

Another factor that has been implicated in both performance on language tasks and in the ability to implement response strategies is working memory (Baddeley & Hitch, 1974; Baddeley, 2000). Working memory is a short term memory

[1] It should be noted that the few word association studies that have collected RTs have shown that the speed with which a response is generated can be affected by the characteristics of the cue (de Groot, 1989; Ernest & Paivio, 1971; Playfoot & Izura, 2013) or the language proficiency of the participant (Fitzpatrick & Izura, 2011). In the case of the current paper we have selected the stimuli and the participant sample to try to limit the effects of these variables on the RT.

space used for manipulating and integrating information from external stimulus and from long term memory stores. It is a limited capacity system, and the amount of working memory resources that an individual has available to them has been linked to performance in a number of language tasks (see Baddeley, 2003). Participants with low working memory capacity consistently respond more slowly and less accurately than participants with high working memory capacity across all these tasks. The explanation for this is that they are ill-equipped to deal with large amounts of information at once. When considering the requirements of a word association task, it is easy to suggest how working memory may play a part although to our knowledge there is no published study explicitly examining this issue. The "new" information from the cue word that is presented to the participant may be integrated with the associated response word that has been accessed in long-term memory within the working memory system. Under circumstances where the response generated is indeed the first one to be activated in the lexicon, the load on the working memory system is fairly light. However, if multiple potential associates for the cue are being compared and a response is being deliberately chosen from among these options, the participant must use their working memory to temporarily store them prior to output. **If we assume that the activation of potential associates is a function of lexical structure and the dynamics of spreading activation, as discussed above, it is likely that a similar number of candidate responses are activated in the lexicon of any respondent from a given population (though which precise words are activated will be unique to each individual). Working memory capacity, therefore, is only a factor in deciding which of the potential associates will be chosen for output. Under these circumstances, individuals with greater working memory resources at their**

disposal will presumably perform better because they will be better able to weigh up the response options than someone with low working memory capacity.

In the current paper we report two experiments that attempt to determine whether the responses generated by participants in word association tasks reflect the first word that is activated by the cue. Experiment 1 required that participants responded to the same cue words twice, under two different task instructions. In one condition, the participants performed a standard word association task. In the second condition, they were asked to respond to each cue with a word that was associated to the cue but that the participant thought would not be given by other people (this is referred to as the **creative** association task hereafter). Essentially, this condition asked them to try hard not to give stereotypical responses. The rationale for this manipulation is that success in the **creative** association task necessarily requires that the strongest intra-lexical link is inhibited or ignored to allow uncommon associates to become activated. Doing so will incur a time cost, and rely on working memory resources. Thus, if responses under standard word association conditions are the first words to be activated, there ought to be a difference in RT between the two versions of the task in the current experiment. There is also likely to be an effect of working memory capacity on responses only in the **creative** association task.

Experiment 1 - Task instructions

Participants

Sixty-eight undergraduate participants (17 male, 51 female, mean age = 20.8, SD 2.34) were recruited for this experiment. All participants were native speakers of English with normal or corrected-to-normal vision. In addition, none of the participants had been diagnosed as dyslexic.

Materials and Design

Participants gave word association responses to the 98 cue words **(see appendix 1)** from Fitzpatrick et al (2013) under two different task requirements. One condition was a standard word association task. The second condition required that participants generated associated words that they thought would be infrequent among respondents. All participants also completed the Operation Span (OSPAN) task as described by Kane, Hambrick, Tuholski, Wilhelm, Payne and Engle (2004). This is a test of working memory. Each trial proceeds as follows. A stimulus is presented in the format "Is 6 + 7 = 13? ball." The participant has to read aloud the sum (is 6 plus 7 equal to 13?), vocally answer the question (yes or no) and then read the word aloud. After a series of 2-5 such items, the participant is asked to recall the words. This means they have to keep words in mind while manipulating and processing the stimuli in front of them. There are 12 groups of operations in total. The working memory scores were used to separate participants into high and low capacity groups prior to analysis. Overall the study was a mixed 2 (standard versus **creative**) x 2 (high versus low working memory) design. The order of the presentation of the repeated measures variable was counterbalanced across participants. Within each condition, cue words were presented in a random order.

Stimulus presentation and response recording was controlled using E-Prime (Schneider, Eschmann & Zuccolotto, 2002).

Procedure

The presentation procedure was the same as in Playfoot and Izura (2013) - a cue was presented onscreen and the participant was instructed to say aloud the first associated word that came to mind. A microphone detected their response and the programme moved to the next screen, on which the participant typed the word they had just said. Typing the response was not time limited in either condition. Once the participant had completed typing, pressing the enter key triggered the presentation of the next cue. Reaction times were recorded from the onset of the cue to the detection of a response by the microphone. After the completion of the first iteration of the word association task, the participants were presented with the OSPAN task. Finally, the participants went through the word association task again under the second **set of task instructions**. Participants were asked to say "pass" in instances where they could not generate an acceptable response.

Results

Creating the norms list

In accordance with the recommendations of Fitzpatrick et al (2013), we created a norms list specific to the population and cue words applicable to the study at hand.

A full discussion of the rationale for doing so is provided in the above paper. In relation to the current work, though, the key issue was that the norms lists published by Fitzpatrick et al (2013) were drawn from the responses of participant groups in Australia, and who were demographically different from our participants. These differences result in patterns of response that may be tied to geographical context - as an example, a popular response to the cue "terrace" in the Australian sample was "school" as it was the name of an educational institution in the local area. None of the participants in this study provided that response as the two words are not inherently related.

The norms list was created using the responses offered by our participants during the standard word association task. The first step was to clean the participants' responses, first by deleting false starts and passes, and then by trimming on the basis of RT. For each participant in turn, a mean and SD of RT were calculated. Any responses recorded above 3 SD from a participant's own mean RT were deleted. By this method we ensured that the responses incorporated into the norms list were an accurate reflection of word association behaviour. The procedure for creating the norms list was identical to that of Fitzpatrick et al (2013), and interested readers are directed to that paper for a comprehensive overview. Briefly, responses to each cue for all 68 participants were collated. Any occasion where the participant's response was a word was assumed to reflect the word that the participant had intended, even if it appeared to be erratic. Spelling mistakes were corrected only when it was clear that the intended word had been mistyped (because there was no other possibility e.g. *controle*). If the participant had typed a non-word response that was equally close to two words, it was treated as an omission to avoid the subjective interpretation of the research team from confounding results. Following this, responses were lemmatised according to Level 2 of the classification system proposed by Bauer and Nation (1993). Finally, the number of instances of each response for each cue was counted, and lists were organised according to response frequency.

Creating high and low working memory groups

The OSPAN task was scored as follows. The proportion of words in each group of stimuli correctly recalled was computed. For example, if the participant remembered 2 words from a series of 2 operations they scored 1.0, 2 out of 3 would score 0.66, 2 out of 4 would score 0.5, and 2 out of 5 would score 0.4. Scoring in this manner was preferred because a) it does not disproportionately reward successful recall of groups of operations of a particular length and b) it created a decent spread of scores across our participant group. An average proportion across the 12 groups of operations was calculated for each participant. As a whole, participants remembered an average of 71.4% of the words in the OSPAN task. Participants were split into high and low working memory groups at this mean.

Participant responses in the standard word association task were trimmed on the basis of RT. For each participant in turn, a mean and SD of RT were calculated. Any responses recorded above 3 SD from a participant's own mean RT were deleted. This process was repeated for the **creative** association task.

Inferential analyses

Participant responses in both the standard and **creative** association tasks were scored for stereotypy according to the norms list created from the data. One point was scored for giving the associate most commonly elicited by a given cue word. For cues with two responses that were equally popular on the norms list, a point was scored for giving *either* of the words[2]. To create a measure of task performance in the **creative** association task we awarded a point for any response that was not represented on the norms list, provided that it was clear to at least one member of our research team that the response was related to the cue in some way. These scores, along with mean RTs, are presented below in Table 1.

[Table 1 about here]

A 2 (standard vs **creative**) x 2 (working memory group) mixed ANOVA was computed with stereotypy as the dependent variable. This was largely a check that the participants had understood and performed the task as instructed. A main effect of instruction was observed [F (1, 66) = 451.249, MSE = 27.264, p < .001]. Stereotypy scores were significantly higher when participants were asked to provide stereotypical answers (29.887) than when required to give unusual answers (10.790). No main effect of working memory was observed, but the interaction between working memory and instruction was significant [F (1, 66) = 5.440, MSE = 27.264, p

[2] **It could be suggested that 0.5 points should be allocated in the event of a tie in the stereotypical response. However, consider a hypothetical situation in which a cue elicits 2 equally popular responses and where these are the only two associates offered by a sample of 100 people. Each response has an associative strength of .50. Consider that, amongst the same 100 participants, another cue elicits one response from 50 people, and another 50 responses from 1 person each. The associative strength of the most popular answer to this second cue is also .50. A participant who agreed with the top answer for the cue with a single strongest response has agreed with 50 people. A participant who agreed with either of the strongest responses to the equally strong response cue has *also* agreed with 50 people. Clearly all three of these potential scoring responses are equally popular in the normative population - allocating a different amount of credit to the answers would be unjustified. While the example given above is hypothetical, the scoring system implemented should be able to deal with such situations fairly in case they do arise in practice.**

< .05]. Post hoc t-tests with Bonferroni correction applied showed that participants with high working memory capacity scored more points for stereotypy (31.774) than those with low working memory capacity (28) in the standard association task, but that no difference was observed in the **creative** association task (10.581 vs 11).

The number of appropriate responses on the creative association (i.e. responses that were legitimate associates and unique to one participant) was compared with stereotypy scores in the standard association task using a second 2 x 2 ANOVA. Here, a main effect of instruction was observed [F (1, 66) = 13.723, MSE = 122.216, p < .001] such that scores on the **creative** association task were significantly higher than scores on the standard association task. This is not surprising because there are fewer possible scoring responses in the standard task (98 top answers) than on the **creative** task. A significant main effect of working memory was also observed [F (1, 66) = 4.423, MSE = 62.044, p < .05]. High working memory participants did better overall than participants with low working memory capacity. There was no interaction.

A final 2 x 2 ANOVA was computed with mean RT as the dependent variable. A main effect of instruction was observed [F (1, 66) = 116.933, MSE = 922327.390, p < .001], such that mean RT was significantly shorter for standard association (2527ms, SD = 748) than **creative** association (4296ms, SD = 1679). No main effect of working memory was observed, nor was there a significant interaction. A significant positive correlation (r = .603) was observed between RT in the standard and **creative** conditions. Slow responders were likely to be slow in both tasks. It should also be noted that there was a significant positive correlation between RT on the **creative** task and scores for generating idiosyncratic responses under these instructions (r = .341). Those who took longer to offer a response were more likely to score highly on this task. Additionally, those participants who scored high for stereotypy in the standard task were also likely to score highly for stereotypy in the creative task (r = .244).

Discussion

The key findings here are as follows. Firstly, there was a significant difference in RT between responses in the normal and **creative** tasks. This suggests that participants required fewer, or less effortful, processes in generating common responses than uncommon responses. This would be expected if standard association responses were indeed reflections of the strongest intra-lexical links. Secondly, participants gave significantly fewer stereotypical associates in the **creative** association task. This indicates that participants were altering their responses according to the demands of the task.

An influence of working memory was observed in relation to task-specific performance scores, and there was an interaction between working memory and task instructions in the analysis of stereotypy. This was contrary to predictions if word association responses are not affected by any response strategy. We will return to discuss this issue following the findings of Experiment 2.

Experiment 2 - Word association under time pressure

Experiment 2 contrasts the associations of high and low working memory participants to cues presented in two different response deadlines in relation to a measure of word association behaviour known as *stereotypy*. In scoring stereotypy, the participant's responses are compared with published norms and a point is awarded for every occasion on which the participant produces the word on norms list with the highest associative strength. In one condition, our participants performed the word association task with no time limit, in common with previous word association studies. In the second condition, participants were forced to respond within 1200 milliseconds. The implementation of a response deadline was designed to preclude the use of any deliberate response strategy. **While the imposition of response deadlines has not, to our knowledge, been applied to word association tasks in the past, there is precedent for varying the speed with which a response must be offered in order to assess other language processes. A particularly good example of this comes from Balota and Chumbley (1985). They conducted a study on the effect of printed word frequency on reading aloud, the typical finding in such studies being that a more commonly encountered word takes less time to read out than a less common word. When presented with a written word, the participant must access its representation in the lexicon and produce its phonological form. Balota and Chumbley argued that word frequency could effect a) lexical access, b) production or c) both. To explore this, participants were presented a series of words onscreen and, after a delay, the participants were given a cue to pronounce the word. Delays ranged from 150ms to 1400ms in 250ms increments, and RT was measured from the presentation of the response cue to the detection of the participant's oral response. At shorter delays (< 900ms), Balota and Chumbley (1985) found a significant frequency effect on response latencies, such that high frequency words were faster to elicit response than low frequency words. At delays beyond 1150ms, the frequency effect**

disappeared. Balota and Chumbley explained this by arguing that the frequency effect was influenced by production processes at shorter delays that were eliminated at longer delays because the participant had time to subvocally rehearse (which is, incidentally, a working memory process) the output between written word presentation and pronunciation cue. That is, processes that affected participant responses at longer stimulus-response intervals could not occur when a tight processing deadline was enforced. The imposition of a response deadline in the current experiment is predicated on findings such as this.

The cut-off for allowing responses was placed at 1200ms based on the mean and SD of reaction times in the standard association condition of Experiment 1. Seventy percent of participants in that experiment responded within 2700ms of the presentation of the cue. The deadline was placed 2SD below that figure - in this way it was intended that most people would be required to respond considerably faster than they would have done without the deadline imposed without preventing any participants from being unable to respond to any of the cues in time. If word association responses are the first word that is activated by the cue, then a) imposing a time limit will not significantly alter the stereotypy of participants in the two conditions and b) will not be affected by working memory capacity. **As stated in the predictions for Experiment 1, we assume that working memory capacity only comes in to play if participants are juggling multiple possible associates in order to choose the best candidate for output. If the response that is offered reflects the first word activated in the participant's lexicon, then working memory is not involved (no alternative**

responses are being assessed). Under circumstances where the participant is precluded from using a strategy that requires them to weigh up several potential responses, as they are in the deadline condition, working memory cannot be involved. Therefore the associate offered by a participant would not be influenced by working memory capacity in either condition.

Method

A group of 28 **undergraduate** participants (5 male, 23 female, mean age = 20.7, SD 1.77) completed this experiment. Participants were not dyslexic, were native speakers of English and had normal or corrected-to-normal vision. None of them had participated in Experiment 1. Participants were asked to offer word association responses to the 98 cues from Fitzpatrick et al (2013) under two different conditions. One condition was a standard word association task. In the second condition, cues were presented for 1200ms, and only responses produced in this window were recorded. Trials proceeded as in Experiment 1. The OSPAN task was also administered.

Results

Stereotypy and OSPAN scores were calculated in the same way as in Experiment 1. The mean OSPAN score was 71.1%, and this was used to split the participants into high and low working memory groups. **Across all participants, the average proportion of trials in which a response was recorded before the deadline**

imposed was 77%. Table 2 presents the relevant descriptive statistics for this experiment.

[Table 2 about here]

A 2 (deadline vs no deadline) x 2 (working memory group) mixed ANOVA was conducted on stereotypy scores. A main effect of deadline was observed [F (1, 26) = 12.260, MSE = 17.252, p < .05] such that stereotypy scores were significantly lower when a response deadline was imposed (25.635 vs 29.563). There was no main effect of working memory group, and no interaction between the factors.

Given that considerably fewer responses were offered under speeded conditions overall, participants were also given a score for *proportion of stereotypical responses*. To do so, their stereotypy score for each condition was divided by the number of valid responses that they recorded. A second 2 x 2 ANOVA was conducted using this proportion stereotypy score. Again, a main effect of deadline was observed [F (1, 26) = 10.908, MSE = 0.001, p < .05], but this time the proportion of stereotypical responses was significantly higher under time pressure (35.7% vs 32.6%). No main effect of working memory was observed, and neither was an interaction between the factors.

It should also be noted that there were significant positive correlations between stereotypy scores under standard and speeded conditions (r = .578) and between the proportion stereotypy scores under standard and speeded conditions (r = .824). That is, participants who gave a greater number of stereotypical responses in one condition tended also to score more stereotypy points in the other condition.

Discussion

The results from Experiment 2 suggest that the imposition of a response deadline alters word association stereotypy. However, it appears that this is not due to working memory, and that it is not a change in word association behaviour per se. When pressed for a quick response, participants offer significantly fewer stereotypical responses. On the face of it, it might appear that this is an indication that participants are not always choosing the first response that comes to mind, occasionally choosing a response that they consider to be more common. However, the fact that the *proportion* of stereotypical responses increases when a deadline is imposed suggests that a much more likely explanation is that relatively weaker intra-lexical links can be employed to arrive at a response given sufficient time to respond. It may be that some of the responses to *slightly* weaker links (those that would have been given just outside the deadline) will be stereotypical. Thus the number of stereotypical responses is greater in standard versus speeded conditions. Particularly weak intra-lexical links are likely to result in idiosyncratic responses. This means that idiosyncratic responses are less likely to be offered when a deadline is imposed because there simply is not time to activate them. Hence the proportion of stereotypical responses in speeded conditions increases by virtue of the fact that it is the responses that are *not* stereotypical that cannot be offered before the deadline. The strong positive correlation between scores in normal and speeded conditions is also supportive of this - participants are scoring consistently, perhaps because the actual response they offer is the same irrespective of condition.

General discussion

We set out to assess whether the responses that are given by participants in word association tasks were likely to reflect the first word that was elicited by the cues, or whether participants were able to deliberately implement some form of response strategy. In what follows, we will argue that our findings suggest that word association responses *are* indeed the first word that comes to mind.

In Experiment 1, we manipulated the task instructions so that one condition compelled participants to choose a response other than the strongest link between two words in their lexicon. By doing so, we intended to measure responses that required several potential options to compete, and for the production of a response to rely on working memory processes. The first key finding from this experiment is that, on average, responses in the standard association task took significantly less time to generate than in the **creative** association condition. That is, when participants were required to produce an uncommon response to a cue, the task demanded that several potential words were considered and were weighed against the criterion for scoring points. Therefore the search for a response took longer than under standard word association instructions, because, we argue, participants were not considering more than one possible response before output. This is further supported by the significant positive correlation between RT in the **creative** association task and success in choosing uncommon responses - participants were more likely to score points for responses that took longer to generate (i.e. that took longer to activate in the lexicon).

Experiment 1 also demonstrated that participants in the high working memory group were better able to choose unusual responses in the **creative** association task, as they possessed the ability to juggle multiple options before deciding on a response. This was as expected. Interestingly though, we also found that participants with high working memory scores performed better than their low working memory counterparts in the standard association task. This was contrary to our predictions - we expected no influence of working memory on word association behaviour if responses reflect the first word to be activated in the participant's lexicon. On the face of it, this finding seems to undermine word association tasks as tapping into the strongest link between two nodes in the participant's lexicon. However, one possible explanation for this finding (an explanation which does not refute the underlying assumptions of word association) is that participants who score highly on measures of working memory do so not because they have a greater capacity available to them, but because they make more efficient use of the resources they have. As an example, consider the bank balance of two people just before payday. They may both have £200 remaining, but what that £200 represents may well differ. One of these people may be paid £2000 per month (i.e. they have a larger financial capacity); the other may get paid £1000 per month but spend it grudgingly (i.e. they are efficient within the confines of the capacity that they have). In our view it is possible that a high score on a working memory test could be achieved if the participant was able to use well-travelled links with long term memory for some parts of the task in order to keep space in the working memory itself available. In other words, efficient use of the connections between input and long-term memory will result in stereotypical responses in a word association task *and* may contribute to a high score in working memory tests. Indeed, there has been some empirical evidence that suggests that performance on tasks of working memory can be improved considerably by making use of long-term memory strategies. For example, Chase and Ericsson (1981) described the performance of **an individual referred to as** SF who had an exceptionally large digit span. SF was able to retain long strings of numbers by converting them to meaningful running times, making use of long-term memory to improve performance on a working memory task. **To SF, a string of 4 digits might reflect the number of minutes and seconds taken to complete a race of a given distance, turning 4 relatively meaningless pieces of information into 1 meaningful chunk.**

We acknowledge that the aim of the study described here is only concerned with the processes involved in lexical retrieval and word selection, but the use of RT as a dependent variable in word association tasks also measures the time taken to perceive the cue. Thus there are factors that potentially influence the response latency that are not attributable to the processes we are interested in. However, the cue words used for both the standard and the creative association tasks were the same, and the same participants took part in both conditions. As a consequence, any influence that perception processes had on RT in one condition are likely to have had a roughly equivalent impact on the other condition as well, because in essence each participant acted as their own control. Thus differences between RT in the standard versus the creative word association task are attributable to processes occurring that are *not* the same in both versions of the task.

It is plausible, of course, to argue that the responses made in either task reflecting the conclusion of some strategic decision-making process, and that the slower reaction times observed in the creative association task are simply because it is a more complex task that requires a greater processing effort before completion. For example, it may be that a number of potential responses are activated in both tasks, but that the process of discarding inappropriate responses in order to rest on a response that is likely to score

takes a greater number of iterations in the creative task. In fact if this is the case, then our RT data do not provide any convincing support for the conclusion that word association responses in standard association tasks are the first words that are elicited by the cue. However, we consider that this is unlikely to be the case. Our basis for this argument is two-fold.

Firstly, research has shown that completing a complex task requires more extensive use of working memory resources and is more difficult for those individuals with low working memory capacity, hence these participants are slower and less accurate in completing the task (e.g. Just & Carpenter, 1992). If the difference between the tasks in our Experiment 1 is simply that one is harder than the other, this would imply that those participants in the low working memory group would score significantly lower in the creative association task than the high working memory group. They did not. Low working memory participants would also be significantly slower to complete the creative association task than their high working memory counterparts. Again, they were not.

The second part of our argument rests on the pattern of stereotypical responses observed in the *creative* association task. Stereotypical responses in the creative task are essentially errors. If a strategic decision is being made in both the standard and the creative task, there ought to be no systematic relationship between appropriate stereotypy scores in the standard task and erroneous stereotypy scores on the creative task. That is, if the mechanism by which a response is generated is the same in both tasks then the likelihood of selecting a commonly-associated word is tied to the task demands and not to the cognitive processes of the individual respondent. In our data, however, we

observed a significant positive correlation between stereotypy scores in the standard task and stereotypical errors in the creative task. Furthermore, there ought to be no difference between the RT for stereotypical errors and for scoring creative responses - if it is the case that the complexity of the task demands are driving the average response latency up then this complexity should influence qualitatively different responses equally. Again, our data do not match this prediction. RT in the creative word association task was negatively correlated with stereotypy scores in that task, indicating that responses that were quicker also tended to be errors given the instructions for the task. The above patterns in the data are not readily reconciled with the notion that the creative association task is completed using the same method as the standard association task and that the former task is simply more difficult than the latter. It does, however, match with an account that the stereotypical response is activated more quickly and has to be inhibited when the required response is to offer a valid, but uncommon, associate. Errors (i.e. offering a stereotypical response when asked for an uncommon response) in the creative association task reflect trials on which the participant has failed to inhibit the automatic response. Thus such responses are more likely to be observed a) in participants who are skilled in accessing strong intra-lexical links and b) in trials where the response was offered quickly, as additional time has not elapsed to allow for other options to be activated and considered.

Our position that word association responses reflect the strongest intra-lexical link for the participant is further corroborated by the findings of Experiment 2, in which we manipulated working memory and imposed a response deadline so that participants did not have sufficient opportunity to implement strategic responses. We observed no effect of working memory capacity in this experiment. Also of note here is that a) fewer stereotypical responses and b) a greater proportion of responses scored a point for stereotypy under time pressure. This matches the predictions of semantic network models (Collins & Quillian, 1969; Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005) in which activation spreads from word to word as a function of the strength of the link between them. Strong links allow activation to pass quickly, and are therefore reflected in word association cue-response pairings that are high in associative strength. These cue-response pairs are likely to be the stereotypical responses in a word association task. As the strength of the intra-lexical link decreases so too does the speed with which activation can be passed from node to node. These cue-response pairs may, in some cases, be weaker for one participant than for the population as a whole. Though the activation required for response is accrued more slowly, the output that is ultimately generated by the participant will still be stereotypical in a proportion of trials. When a response deadline is imposed, there is no longer time for the participant to fully activate a relatively weak cue-response pairing. In some instances this will result in **a participant** being unable to generate the stereotypical response for a given cue, hence fewer stereotypy points will be scored on average. However, the majority of the responses that are slow and effortful will reflect uncommon cue-associate pairs that *would not* have been stereotypical in any case. Thus a greater number of idiosyncratic versus stereotypical responses are omitted overall, and the proportion of participant responses that score points increases. **It would be worthwhile conducting research in the future which specifically examines whether this prediction is borne out by the data. This could be accomplished by systematically changing**

the deadline such that a smaller proportion of responses time-out in each iteration of the tasks - if our interpretation is correct, there may be a point where *only* idiosyncratic responses are omitted. It might also be interesting to determine whether there are predictable characteristics of a) the cues that a given participant responds to particularly slowly and b) the types of responses that are elicited at longer latencies. This would be of interest not only in relation to the allocation of stereotypy points (as we have in the current Experiment 2) but also with regard to the effects of word frequency, concreteness, word class, and the oft-considered category of response (paradigmatic versus syntagmatic, for example) in order to provide a greater depth of understanding regarding the structure and dynamics of the lexicon.

It would appear, therefore, that the assumptions on which word association research has been based are supported by the current study. By and large the responses that are given by participants *do* reflect the first word that comes to mind.

References

Baddeley, A. D. 2000. 'The episodic buffer: A new component of working memory?' *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A.D. 2003. 'Working memory and language: an overview.' *Journal of Communication Disorders, 36 (3),* 189-208.

Baddeley, A. D., and **G.J. Hitch,** 1974. 'Working memory'. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Balota, D.A. and **Chumbley, J.I.** 1985. 'The locus of word frequency effects in the pronunciation task: Lexical access and/or production.' *Journal of Memory and Language, 24,* 89-106.

Bauer, L. M. and **I. S. P. Nation.** 1993. 'Word families,' *International Journal of Lexicography* 6: 253–79.

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

Canas, J. J. 1990. 'Associative strength effects in the lexical decision task.' *Quarterly Journal of Experimental Psychology,* 42A, 121-145.

Chase, W. G., and **K.A.** Ericsson, 1981. Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Hillsdale, N. J.: Erlbaum.

Collins, A., and **M., Quillian,** 1969. 'Retrieval time from semantic memory.'*Journal of verbal learning and verbal behaviour, 8 (2),* 240–248.

Collins, A., and **E. Loftus,** 1975. 'A spreading-activation theory of semantic processing.' *Psychological Review, 82 (6)*, 407–428

Davis, C. J. and **S.J. Lupker,** 2006. 'Masked inhibitory priming in English: Evidence for lexical inhibition.' *Journal of Experimental Psychology: Human Perception & Performance, 32*, 668-687.

Ferrand L. and **B. New.** 2003 'Semantic and associative priming in the mental lexicon.' *Mental lexicon: some words to talk about words* (ed. Bonin P.), pp. 25–43. Hauppauge, NY: Nova Science Publisher

Fitzpatrick, T. 2007. 'Word association patterns: Unpacking the assumptions,' *International Journal of Applied Linguistics* 17: 319–31.

Fitzpatrick, T. 2009. 'Word association profiles in a first and second language: puzzles and problems' in T. Fitzpatrick and A. Barfield (eds): *Lexical Processing in Second Language Learners*. Multilingual Matters, pp. 38–52.

Fitzpatrick, T. and **C. Izura.** 2011. 'Word association in L1 and L2: An exploratory study of response types, response times and inter-language mediation,' *Studies in Second Language Acquisition 33*: 373–98.

Fitzpatrick, T., D. Playfoot, A. Wray, and **M.J. Wright,** 2013. 'Establishing the reliability of word association data for investigating individual and group differences.' *Applied Linguistics,34(5) ,* 1-29.

Gewirth, L. R., A. G. Shindler, and **D. B. Hier**. 1984. 'Altered patterns of word associations in dementia and aphasia.' *Brain and Language, 21*(2), 307-317.

Gollan, T. H., D. P. Salmon, and **J. L. Paxton**. 2006. 'Word association in early Alzheimer's disease.' *Brain and Language, 99*, 289-303.

Hartmann, G. W. 1941. 'A critique of the common method of estimating vocabulary size, together with some data on the absolute word knowledge of educated adults.' *Journal of Educational Psychology, 32,* 351-358.

Hirsh, K. W., and **J. T. Tree**. 2001. 'Word association norms for two cohorts of British adults.' *Journal of Neurolinguistics, 14*(1), 1-44.

Just, M.A. and **Carpenter, P.A.** 1992. 'A capacity theory of comprehension: Individual differences in working memory.' *Psychological Review, 99,* 122-149.

Kane, M.J., D.Z. Hambrick, S.W. Tuholski, O. Wilhelm, T.W. Payne, and **R.W. Engle,** 2004. 'The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning.' *Journal of Experimental Psychology: General, 133(2),* 189-217.

Kent, G.H. and **Rosanoff, A.J.** 1910. 'A study of association in insanity' *American Journal of Insanity, 67,* 37-39, 317-390.

Meara, P.M., 2009. Connected Words: word associations and second language lexical acquisition. *Amsterdam: John Benjamins.*

Merten, T. 1993. 'Word association responses and psychoticism.' *Personality and Individual Differences, 14,* 837-839.

Meyer, D., and **R. Schvaneveldt,** 1971. 'Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations.' *Journal of Experimental Psychology,* 90, 227-234.

Nelson, D. L., C. L.McEvoy, and **T. A. Schreiber**. 1998. 'The University of South Floridaword association, rhyme, and word fragment norms' http://www.usf.edu/FreeAssociation/.

Nusbaum, H.C., D.B. Pisoni, and **C.K. Davis,** 1984. 'Sizing up the Hooiser mental lexicon: Measuring the familiarity of 20,000 words.' In *Research on Speech Perception, Progress Report 10,* (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.

Playfoot, D. and **C. Izura,** 2013. 'Imageability, age of acquisition and frequency factors in acronym comprehension.' *Quarterly Journal of Experimental Psychology, 66(6),* 1131-1145.

Postman L. J. and **G. Keppel** (eds). 1970. Norms of Word Association. Academic Press.

Schneider, W., A. Eschman, and **A. Zuccolotto,** 2002. *E-Prime 1.0*. Pittsburgh, PA: Psychological Software Tools.

Steyvers, M., and **J. Tenenbaum,** 2005. 'The large-scale structure of semantic networks: statistical analyses and a model of semantic growth' *Cognitive Science, 29 (1),* 41-78.

Wettler, M., Rapp, R., and **Sedlmeier, P** 2005. 'Free word associations correspond to contiguities between words in texts.' *Journal of Quantitative Linguistics, 12*, 111-122.

Appendix 1

Word association cues (from Fitzpatrick et al., 2013)			
abbey	delay	landlord	sand
abuse	devote	liquid	script
agenda	diet	loss	session
alley	domestic	manual	shove
annoy	dominate	mathematics	sin
astonish	echo	miracle	snap
attack	effort	multiple	source
basket	establish	nail	spite
bean	expose	nuclear	stiff
blame	extension	nurse	store
bond	fence	overtake	suicide
bread	fined	owe	swear
bucket	foster	peak	symbol
canal	fraction	permit	terrace
candidate	gentle	plug	thick
certificate	gold	poison	torch
cheese	greed	prevent	tour
click	hay	pride	tumble
cloud	heaven	pudding	vandal
concentrate	hood	rack	variety
concert	ideal	reflect	wander
cope	indulge	repair	weak
corridor	irony	rescue	wolf
cupboard	joint	rock	
curious	ladder	routine	

Table 1 - Mean stereotypy scores, task-specific scores and RTs (SD) for each task according to working memory group.

	Stereotypy		Task-specific score		RT	
	High WM	Low WM	High WM	Low WM	High WM	Low WM
Standard	31.77 (7.36)	28.00 (6.80)	31.77 (7.36)	28.00 (6.80)	2421.26 (745.61)	2616.69 (749.28)
Creative	10.58 (4.70)	11.00 (4.73)	37.97 (12.84)	35.97 (10.44)	4423.54 (1628.07)	4190.43 (1756.97)

Table 2 - Mean stereotypy scores, plus mean proportion stereotypy scores for each task according to working memory group (numbers in parentheses denote SD).

	Stereotypy		Proportion stereotypy	
	High WM	Low WM	High WM	Low WM
Deadline	27.44 (6.51)	23.82 (5.54)	0.341 (0.081)	0.372 (0.098)
No Deadline	29.63 (7.47)	29.50 (6.13)	0.327 (0.071)	0.324 (0.082)