

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

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/12205/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

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*.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>



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

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

SCHOLARONE™
Manuscripts

View

1
2
3 **Are word association responses *really* the first words that come to mind?**
4
5
6
7

8 **Abstract**
9

10
11 Word association has been a popular tool for research in linguistics and psychology
12 over the last century. The paradigm presents participants with a cue word and asks
13 them to respond with the first associated word that comes to mind. Inferences about
14 the structure and organisation of the lexicon have been made on the basis of the
15 findings of word association tasks, and on the assumption that responses reflect the
16 strongest link between words in the participants' vocabulary. The procedure adopted
17 in traditional word association tasks does not guarantee that this is the case. This
18 paper presents two experiments that aimed to determine whether or not participants
19 make deliberate and strategic responses in word association tasks. Findings indicate
20 that word association responses are likely to reflect the first word that participants
21 activate in their lexicon.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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

1
2
3 activated by a word association cue. Although this assumption has gained some
4
5 (often indirect) empirical support, to our knowledge it has not yet been tested directly.
6
7

8
9 Estimates of the number of words that an adult knows in their native language
10 vary considerably from study to study, and vocabulary sizes of anywhere between
11 14,418 (Nusbaum, Pisoni & Davis, 1984) and more than 200,000 words (Hartmann,
12 1941 - **though it should be noted that this is far higher than most estimates**)
13 have been reported. Given that vocabulary size is often calculated relative to printed
14 corpora (which typically under-represent proper nouns, slang, acronyms etc, see
15 Brysbaert & New, 2009) it is likely that actual vocabulary size exceeds the published
16 estimates. In spite of having such a large number of words to choose from, skilled
17 readers are able to find and produce the appropriate word quickly when they are
18 asked to name an object or in fluent speech. The prevailing opinion is that this is
19 because word knowledge is stored as nodes in an interconnected semantic network
20 (e.g. Collins & Quillian, 1969; Collins & Loftus, 1975; Steyvers & Tenenbaum, 2005).
21 In Steyvers and Tenenbaum's (2005) model, for example, each word node is
22 connected to any number of other words by links that vary in strength. When a word
23 is encountered (or *activated*) some activation is also passed along each intra-lexical
24 link that stems from the stimulus. The amount of activation that passes (or *spreads*)
25 to each connected word is determined by the strength of the link, which in turn is
26 determined by personal experience. The more times a particular link between two
27 words is traversed, the stronger it becomes. The consequence of this spreading
28 activation is that the presentation of one word can lead to increased likelihood of
29 producing a related word soon afterward. For example, in a classic experiment,
30 Meyer and Schvaneveldt (1971) presented participants with targets for lexical
31 decision (i.e. does this combination of letters represent an existing word) that were
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 immediately preceded by a prime word. In some trials, the prime word was related in
4 meaning to the target (e.g. doctor-NURSE). In other trials the prime and target were
5 unrelated (e.g. doctor-BREAD). Meyer and Schvaneveldt (1971) reported that
6 participants were significantly faster to respond to related targets than unrelated
7 targets. In relation to spreading activation theory, the presentation of "doctor"
8 passed some activation along the connection to NURSE. As a consequence, when
9 NURSE was itself presented, it was already partially activated and took less
10 additional effort to reach a recognition threshold. The same is not true for BREAD,
11 which is not connected to "doctor" in the lexicon. It has since been demonstrated
12 that the presentation of a prime word automatically activates all the connected words
13 in the lexicon, whether the connection is semantic, associative (e.g. Ferrand & New
14 2003), or formal (e.g. Davis & Lupker, 2006) even if exposure to the prime is short
15 enough to prevent conscious processing. The key principle here is that strong links
16 between words can be accessed quickly and automatically.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 In discrete word association tasks, participants are presented with a single
36 cue word and required to say or write down the first word that comes to mind. It has
37 been argued that the first word that comes to mind ought to reflect access of the
38 strongest intra-lexical link (e.g. Playfoot & Izura, 2013). That is, the word association
39 cue acts as a prime for the response - once the cue has been activated in the lexicon
40 of the participant, activation will spread to surrounding nodes according to the
41 weights of the connections. The word the participant produces will be the first node
42 to reach a criterion level of activation, and this will be accrued more quickly along
43 strong than weak links. Indeed, much of the word association research in the
44 literature proceeds from this assumption. **One notable exception to this is the**
45 **work of Wettler, Rapp and Sedlmeier (2005), who argue that contiguities in the**
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 presentation of words in the speaker's language are the key determinant of
4 word association responses. They examined the responses elicited by the
5 presentation of 100 cue words (those used by Kent & Rosanoff, 1910) and
6 compared these to the probability that the cue and response co-occurred in
7 sentences (using the British National Corpus, BNC, as the source of this
8 information). They argued that the probability of co-occurrence of a pair of
9 words corresponded well with the responses that were given by human
10 participants in a word association task (though this correspondence was far
11 from perfect), and suggested that word association responses could be
12 explained by paired associative learning processes. They acknowledged,
13 however, that their findings did not disprove the theory that there were
14 semantic structures underpinning word association behaviour. In fact, it could
15 be that the contiguities observed in the BNC are a crude measure of the links
16 between words in the lexical network - the co-occurrence of two words
17 strengthens the intra-lexical link between them in Steyvers and Tenenbaum's
18 (2005) model, so words that are more likely to co-occur in the BNC are also
19 likely to have strong links in the lexicon.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40

41
42 Published norms lists (e.g. Fitzpatrick et al., 2013; Postman & Keppel, 1970;
43 Nelson, McEvoy & Schreiber, 1998) present the same word association cues to large
44 numbers of participants and organise the responses according to the frequency with
45 which they occur within the sample population. This is often converted to a metric
46 called *associative strength*. To do so, the frequency of a particular response is
47 divided by the total number of responses to that cue to create a proportion of
48 participants who produce the same word. For example, if 58 people out of 100 say
49 *white* when presented with the cue word BLACK, the associative strength is 0.58.
50
51
52
53
54
55
56
57
58
59
60

1
2
3 This metric is essentially an approximation of the how strong the intra-lexical link
4 between black and white is across participants. **It is important to remember that**
5 **associative strength is a metric of the connections that are made between**
6 **words at a group level, and that the lexical structure of any given individual is**
7 **unlikely precisely match the idealised picture provided by associative strength.**
8 **Nevertheless, Canas (1990) demonstrated that** associative strength is a good
9 predictor of the size of priming effects - prime-target pairs with greater associative
10 strength elicit quicker responses than pairs with lower estimates of associative
11 strength. This finding provides indirect support for the notion that word association
12 responses are the first word that comes to mind following the cue, in that both the
13 word association response and the priming effect are supposed to rely on activation
14 spreading along the same intra-lexical link.
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

30 Further, weaker, indirect support for the assumption that word association
31 responses are the first word that comes to the mind of the participant has been
32 provided by a handful of studies that have presented the same cue words to a group
33 of participants on two separate occasions (e.g. Fitzpatrick, 2007) or in two different
34 languages (Fitzpatrick, 2007; 2009; Fitzpatrick & Izura, 2011). In such studies, the
35 types of responses that are given by an individual participant are fairly stable over
36 time and across presentations in their first and second languages - certainly more
37 consistent than would be the case if there was not a common mechanism
38 underpinning performance on the task each time it was performed. However, it is far
39 from certain whether this commonality of responses is because of the automatic
40 retrieval of a particularly strong, and stable, intra-lexical link between words or
41 because the responses that they offer are governed by the application of a
42 consistent strategy for performing the task.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Word association tasks typically allow participants plenty of time in which to
4 generate their responses, and are commonly presented as pen and paper measures
5 so reaction time data is not available. This raises a potential concern - researchers
6 have no way of preventing participants from deliberately selecting a response from a
7 number of possible options, and no measure by which to examine whether this is
8 likely to be happening. To explain further, let us start by assuming that the response
9 that a participant gives in a word association task *is not* the first word that comes to
10 mind. Presumably the decision as to which word to choose must take some time.
11 This is because a) you must allow time for multiple associates to be activated and b)
12 you must then apply some kind of decision-making heuristic to determine which word
13 is the most appropriate for the task at hand. As an analogy, consider buying milk in
14 the shop. Your hand does not have to go as far to reach the bottle at the front of the
15 shelf as it does to reach the bottle at the back of the shelf. There is therefore a
16 difference in the time it takes to pick up the two bottles. In addition, to reach the
17 bottle at the back of the shelf you need to move the first bottle out of the way. This
18 adds a step to your milk-buying process and there is a time cost. Therefore
19 responses *other than* the first word that come to mind are likely to be offered at
20 longer response latencies - the traditional lack of RT data does not allow this to be
21 assessed.¹

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46 Another factor that has been implicated in both performance on language
47 tasks and in the ability to implement response strategies is working memory
48 (Baddeley & Hitch, 1974; Baddeley, 2000). Working memory is a short term memory
49
50
51
52

53
54
55 ¹ It should be noted that the few word association studies that have collected RTs have shown that
56 the speed with which a response is generated can be affected by the characteristics of the cue (de
57 Groot, 1989; Ernest & Paivio, 1971; Playfoot & Izura, 2013) or the language proficiency of the
58 participant (Fitzpatrick & Izura, 2011). In the case of the current paper we have selected the stimuli
59 and the participant sample to try to limit the effects of these variables on the RT.
60

1
2
3 space used for manipulating and integrating information from external stimulus and
4
5 from long term memory stores. It is a limited capacity system, and the amount of
6
7 working memory resources that an individual has available to them has been linked
8
9 to performance in a number of language tasks (see Baddeley, 2003). Participants
10
11 with low working memory capacity consistently respond more slowly and less
12
13 accurately than participants with high working memory capacity across all these
14
15 tasks. The explanation for this is that they are ill-equipped to deal with large
16
17 amounts of information at once. When considering the requirements of a word
18
19 association task, it is easy to suggest how working memory may play a part although
20
21 to our knowledge there is no published study explicitly examining this issue. The
22
23 "new" information from the cue word that is presented to the participant may be
24
25 integrated with the associated response word that has been accessed in long-term
26
27 memory within the working memory system. Under circumstances where the
28
29 response generated is indeed the first one to be activated in the lexicon, the load on
30
31 the working memory system is fairly light. However, if multiple potential associates
32
33 for the cue are being compared and a response is being deliberately chosen from
34
35 among these options, the participant must use their working memory to temporarily
36
37 store them prior to output. **If we assume that the activation of potential**
38
39 **associates is a function of lexical structure and the dynamics of spreading**
40
41 **activation, as discussed above, it is likely that a similar number of candidate**
42
43 **responses are activated in the lexicon of any respondent from a given**
44
45 **population (though which precise words are activated will be unique to each**
46
47 **individual). Working memory capacity, therefore, is only a factor in deciding**
48
49 **which of the potential associates will be chosen for output. Under these**
50
51 **circumstances, individuals with greater working memory resources at their**
52
53
54
55
56
57
58
59
60

1
2
3 **disposal will presumably perform better because they will be better able to**
4 **weigh up the response options than someone with low working memory**
5 **capacity.**
6
7
8
9

10
11 In the current paper we report two experiments that attempt to determine
12 whether the responses generated by participants in word association tasks reflect
13 the first word that is activated by the cue. Experiment 1 required that participants
14 responded to the same cue words twice, under two different task instructions. In one
15 condition, the participants performed a standard word association task. In the
16 second condition, they were asked to respond to each cue with a word that was
17 associated to the cue but that the participant thought would not be given by other
18 people (this is referred to as the **creative** association task hereafter). Essentially,
19 this condition asked them to try hard not to give stereotypical responses. The
20 rationale for this manipulation is that success in the **creative** association task
21 necessarily requires that the strongest intra-lexical link is inhibited or ignored to allow
22 uncommon associates to become activated. Doing so will incur a time cost, and rely
23 on working memory resources. Thus, if responses under standard word association
24 conditions are the first words to be activated, there ought to be a difference in RT
25 between the two versions of the task in the current experiment. There is also likely
26 to be an effect of working memory capacity on responses only in the **creative**
27 association task.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 *Experiment 1 - Task instructions*
53
54
55
56
57
58
59
60

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.

1
2
3 Stimulus presentation and response recording was controlled using E-Prime
4
5 (Schneider, Eschmann & Zuccolotto, 2002).
6
7
8
9

10 11 *Procedure*

12
13
14 The presentation procedure was the same as in Playfoot and Izura (2013) - a cue
15 was presented onscreen and the participant was instructed to say aloud the first
16 associated word that came to mind. A microphone detected their response and the
17 programme moved to the next screen, on which the participant typed the word they
18 had just said. Typing the response was not time limited in either condition. Once the
19 participant had completed typing, pressing the enter key triggered the presentation of
20 the next cue. Reaction times were recorded from the onset of the cue to the
21 detection of a response by the microphone. After the completion of the first iteration
22 of the word association task, the participants were presented with the OSPAN task.
23 Finally, the participants went through the word association task again under the
24 second **set of task instructions**. Participants were asked to say "pass" in instances
25 where they could not generate an acceptable response.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 *Results*

46 47 48 49 50 51 *Creating the norms list*

52
53
54 In accordance with the recommendations of Fitzpatrick et al (2013), we created a
55 norms list specific to the population and cue words applicable to the study at hand.
56
57
58
59
60

1
2
3 A full discussion of the rationale for doing so is provided in the above paper. In
4
5 relation to the current work, though, the key issue was that the norms lists published
6
7 by Fitzpatrick et al (2013) were drawn from the responses of participant groups in
8
9 Australia, and who were demographically different from our participants. These
10
11 differences result in patterns of response that may be tied to geographical context -
12
13 as an example, a popular response to the cue "terrace" in the Australian sample was
14
15 "school" as it was the name of an educational institution in the local area. None of
16
17 the participants in this study provided that response as the two words are not
18
19 inherently related.
20
21

22
23
24 **The norms list was created using the responses offered by our**
25
26 **participants during the standard word association task. The first step was to**
27
28 **clean the participants' responses, first by deleting false starts and passes, and**
29
30 **then by trimming on the basis of RT. For each participant in turn, a mean and**
31
32 **SD of RT were calculated. Any responses recorded above 3 SD from a**
33
34 **participant's own mean RT were deleted. By this method we ensured that the**
35
36 **responses incorporated into the norms list were an accurate reflection of word**
37
38 **association behaviour.** The procedure for creating the norms list was identical to
39
40 that of Fitzpatrick et al (2013), and interested readers are directed to that paper for a
41
42 comprehensive overview. Briefly, responses to each cue for all 68 participants were
43
44 collated. Any occasion where the participant's response was a word was assumed
45
46 to reflect the word that the participant had intended, even if it appeared to be erratic.
47
48 Spelling mistakes were corrected only when it was clear that the intended word had
49
50 been mistyped (because there was no other possibility e.g. *controle*). If the
51
52 participant had typed a non-word response that was equally close to two words, it
53
54 was treated as an omission to avoid the subjective interpretation of the research
55
56
57
58
59
60

1
2
3 team from confounding results. Following this, responses were lemmatised
4
5 according to Level 2 of the classification system proposed by Bauer and Nation
6
7 (1993). Finally, the number of instances of each response for each cue was counted,
8
9 and lists were organised according to response frequency.
10

11 12 13 14 15 16 *Creating high and low working memory groups*

17
18 The OSPAN task was scored as follows. The proportion of words in each group of
19
20 stimuli correctly recalled was computed. For example, if the participant remembered
21
22 2 words from a series of 2 operations they scored 1.0, 2 out of 3 would score 0.66, 2
23
24 out of 4 would score 0.5, and 2 out of 5 would score 0.4. Scoring in this manner was
25
26 preferred because a) it does not disproportionately reward successful recall of
27
28 groups of operations of a particular length and b) it created a decent spread of
29
30 scores across our participant group. An average proportion across the 12 groups of
31
32 operations was calculated for each participant. As a whole, participants
33
34 remembered an average of 71.4% of the words in the OSPAN task. Participants
35
36 were split into high and low working memory groups at this mean.
37
38
39
40

41
42 Participant responses in the standard word association task were trimmed on
43
44 the basis of RT. For each participant in turn, a mean and SD of RT were calculated.
45
46 Any responses recorded above 3 SD from a participant's own mean RT were deleted.
47
48 This process was repeated for the **creative** association task.
49
50

51 52 53 54 *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 words². 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.

1
2
3 < .05]. Post hoc t-tests with Bonferroni correction applied showed that participants
4 with high working memory capacity scored more points for stereotypy (31.774) than
5 those with low working memory capacity (28) in the standard association task, but
6 that no difference was observed in the **creative** association task (10.581 vs 11).
7
8
9
10

11
12 **The number of appropriate responses on the creative association (i.e.**
13 **responses that were legitimate associates and unique to one participant) was**
14 compared with stereotypy scores in the standard association task using a second 2 x
15 2 ANOVA. Here, a main effect of instruction was observed [$F(1, 66) = 13.723$, MSE
16 $= 122.216$, $p < .001$] such that scores on the **creative** association task were
17 significantly higher than scores on the standard association task. This is not
18 surprising because there are fewer possible scoring responses in the standard task
19 (98 top answers) than on the **creative** task. A significant main effect of working
20 memory was also observed [$F(1, 66) = 4.423$, $MSE = 62.044$, $p < .05$]. High working
21 memory participants did better overall than participants with low working memory
22 capacity. There was no interaction.
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

38 A final 2 x 2 ANOVA was computed with mean RT as the dependent variable.
39 A main effect of instruction was observed [$F(1, 66) = 116.933$, $MSE = 922327.390$, p
40 $< .001$], such that mean RT was significantly shorter for standard association
41 (2527ms, $SD = 748$) than **creative** association (4296ms, $SD = 1679$). No main
42 effect of working memory was observed, nor was there a significant interaction. A
43 significant positive correlation ($r = .603$) was observed between RT in the standard
44 and **creative** conditions. Slow responders were likely to be slow in both tasks. It
45 should also be noted that there was a significant positive correlation between RT on
46 the **creative** task and scores for generating idiosyncratic responses under these
47 instructions ($r = .341$). Those who took longer to offer a response were more likely to
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 score highly on this task. Additionally, those participants who scored high for
4
5 stereotypy in the standard task were also likely to score highly for stereotypy in the
6
7 creative task ($r = .244$).
8
9

10 11 12 13 *Discussion*

14
15
16 The key findings here are as follows. Firstly, there was a significant difference in RT
17
18 between responses in the normal and **creative** tasks. This suggests that
19
20 participants required fewer, or less effortful, processes in generating common
21
22 responses than uncommon responses. This would be expected if standard
23
24 association responses were indeed reflections of the strongest intra-lexical links.
25
26 Secondly, participants gave significantly fewer stereotypical associates in the
27
28 **creative** association task. This indicates that participants were altering their
29
30 responses according to the demands of the task.
31
32
33
34

35 An influence of working memory was observed in relation to task-specific
36
37 performance scores, and there was an interaction between working memory and
38
39 task instructions in the analysis of stereotypy. This was contrary to predictions if
40
41 word association responses are not affected by any response strategy. We will
42
43 return to discuss this issue following the findings of Experiment 2.
44
45
46
47
48
49

50 *Experiment 2 - Word association under time pressure*

51
52
53 Experiment 2 contrasts the associations of high and low working memory
54
55 participants to cues presented in two different response deadlines in relation to a
56
57 measure of word association behaviour known as *stereotypy*. In scoring stereotypy,
58
59
60

1
2
3 the participant's responses are compared with published norms and a point is
4
5 awarded for every occasion on which the participant produces the word on norms list
6
7 with the highest associative strength. In one condition, our participants performed
8
9 the word association task with no time limit, in common with previous word
10
11 association studies. In the second condition, participants were forced to respond
12
13 within 1200 milliseconds. The implementation of a response deadline was designed
14
15 to preclude the use of any deliberate response strategy. **While the imposition of**
16
17 **response deadlines has not, to our knowledge, been applied to word**
18
19 **association tasks in the past, there is precedent for varying the speed with**
20
21 **which a response must be offered in order to assess other language**
22
23 **processes. A particularly good example of this comes from Balota and**
24
25 **Chumbley (1985). They conducted a study on the effect of printed word**
26
27 **frequency on reading aloud, the typical finding in such studies being that a**
28
29 **more commonly encountered word takes less time to read out than a less**
30
31 **common word. When presented with a written word, the participant must**
32
33 **access its representation in the lexicon and produce its phonological form.**
34
35 **Balota and Chumbley argued that word frequency could effect a) lexical**
36
37 **access, b) production or c) both. To explore this, participants were presented**
38
39 **a series of words onscreen and, after a delay, the participants were given a**
40
41 **cue to pronounce the word. Delays ranged from 150ms to 1400ms in 250ms**
42
43 **increments, and RT was measured from the presentation of the response cue**
44
45 **to the detection of the participant's oral response. At shorter delays (< 900ms),**
46
47 **Balota and Chumbley (1985) found a significant frequency effect on response**
48
49 **latencies, such that high frequency words were faster to elicit response than**
50
51 **low frequency words. At delays beyond 1150ms, the frequency effect**
52
53
54
55
56
57
58
59
60

1
2
3 disappeared. Balota and Chumbley explained this by arguing that the
4
5 frequency effect was influenced by production processes at shorter delays
6
7 that were eliminated at longer delays because the participant had time to
8
9 subvocally rehearse (which is, incidentally, a working memory process) the
10
11 output between written word presentation and pronunciation cue. That is,
12
13 processes that affected participant responses at longer stimulus-response
14
15 intervals could not occur when a tight processing deadline was enforced. The
16
17 imposition of a response deadline in the current experiment is predicated on
18
19 findings such as this.
20
21
22
23
24
25
26

27 The cut-off for allowing responses was placed at 1200ms based on the mean
28
29 and SD of reaction times in the standard association condition of Experiment 1.
30
31 Seventy percent of participants in that experiment responded within 2700ms of
32
33 the presentation of the cue. The deadline was placed 2SD below that figure - in
34
35 this way it was intended that most people would be required to respond
36
37 considerably faster than they would have done without the deadline imposed
38
39 without preventing any participants from being unable to respond to any of the
40
41 cues in time. If word association responses are the first word that is activated by
42
43 the cue, then a) imposing a time limit will not significantly alter the stereotypy of
44
45 participants in the two conditions and b) will not be affected by working memory
46
47 capacity. As stated in the predictions for Experiment 1, we assume that
48
49 working memory capacity only comes in to play if participants are juggling
50
51 multiple possible associates in order to choose the best candidate for output.
52
53
54
55 If the response that is offered reflects the first word activated in the
56
57 participant's lexicon, then working memory is not involved (no alternative
58
59
60

1
2
3 responses are being assessed). Under circumstances where the participant is
4 precluded from using a strategy that requires them to weigh up several
5 potential responses, as they are in the deadline condition, working memory
6 cannot be involved. Therefore the associate offered by a participant would not
7 be influenced by working memory capacity in either condition.
8
9
10
11
12
13
14
15
16
17

18 *Method*

19
20 A group of 28 **undergraduate** participants (5 male, 23 female, mean age = 20.7, SD
21 1.77) completed this experiment. Participants were not dyslexic, were native
22 speakers of English and had normal or corrected-to-normal vision. None of them
23 had participated in Experiment 1. Participants were asked to offer word association
24 responses to the 98 cues from Fitzpatrick et al (2013) under two different conditions.
25 One condition was a standard word association task. In the second condition, cues
26 were presented for 1200ms, and only responses produced in this window were
27 recorded. Trials proceeded as in Experiment 1. The OSPAN task was also
28 administered.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45 *Results*

46
47 Stereotypy and OSPAN scores were calculated in the same way as in Experiment 1.
48 The mean OSPAN score was 71.1%, and this was used to split the participants into
49 high and low working memory groups. **Across all participants, the average**
50 **proportion of trials in which a response was recorded before the deadline**
51
52
53
54
55
56
57
58
59
60

1
2
3 **imposed was 77%**. Table 2 presents the relevant descriptive statistics for this
4
5 experiment.
6
7

8 [Table 2 about here]
9

10
11 A 2 (deadline vs no deadline) x 2 (working memory group) mixed ANOVA was
12
13 conducted on stereotypy scores. A main effect of deadline was observed [$F(1, 26) =$
14
15 12.260 , $MSE = 17.252$, $p < .05$] such that stereotypy scores were significantly lower
16
17 when a response deadline was imposed (25.635 vs 29.563). There was no main
18
19 effect of working memory group, and no interaction between the factors.
20
21

22
23 Given that considerably fewer responses were offered under speeded
24
25 conditions overall, participants were also given a score for *proportion of stereotypical*
26
27 *responses*. To do so, their stereotypy score for each condition was divided by the
28
29 number of valid responses that they recorded. A second 2 x 2 ANOVA was
30
31 conducted using this proportion stereotypy score. Again, a main effect of deadline
32
33 was observed [$F(1, 26) = 10.908$, $MSE = 0.001$, $p < .05$], but this time the proportion
34
35 of stereotypical responses was significantly higher under time pressure (35.7% vs
36
37 32.6%). No main effect of working memory was observed, and neither was an
38
39 interaction between the factors.
40
41

42
43
44 It should also be noted that there were significant positive correlations
45
46 between stereotypy scores under standard and speeded conditions ($r = .578$) and
47
48 between the proportion stereotypy scores under standard and speeded conditions (r
49
50 $= .824$). That is, participants who gave a greater number of stereotypical responses
51
52 in one condition tended also to score more stereotypy points in the other condition.
53
54
55
56
57
58
59
60

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

1
2
3 We set out to assess whether the responses that are given by participants in word
4 association tasks were likely to reflect the first word that was elicited by the cues, or
5 whether participants were able to deliberately implement some form of response
6 strategy. In what follows, we will argue that our findings suggest that word
7 association responses *are* indeed the first word that comes to mind.
8
9
10
11
12
13

14
15 In Experiment 1, we manipulated the task instructions so that one condition
16 compelled participants to choose a response other than the strongest link between
17 two words in their lexicon. By doing so, we intended to measure responses that
18 required several potential options to compete, and for the production of a response
19 to rely on working memory processes. The first key finding from this experiment is
20 that, on average, responses in the standard association task took significantly less
21 time to generate than in the **creative** association condition. That is, when
22 participants were required to produce an uncommon response to a cue, the task
23 demanded that several potential words were considered and were weighed against
24 the criterion for scoring points. Therefore the search for a response took longer than
25 under standard word association instructions, because, we argue, participants were
26 not considering more than one possible response before output. This is further
27 supported by the significant positive correlation between RT in the **creative**
28 association task and success in choosing uncommon responses - participants were
29 more likely to score points for responses that took longer to generate (i.e. that took
30 longer to activate in the lexicon).
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50

51 Experiment 1 also demonstrated that participants in the high working memory
52 group were better able to choose unusual responses in the **creative** association task,
53 as they possessed the ability to juggle multiple options before deciding on a
54 response. This was as expected. Interestingly though, we also found that
55
56
57
58
59
60

1
2
3 participants with high working memory scores performed better than their low
4
5 working memory counterparts in the standard association task. This was contrary to
6
7 our predictions - we expected no influence of working memory on word association
8
9 behaviour if responses reflect the first word to be activated in the participant's lexicon.
10
11 On the face of it, this finding seems to undermine word association tasks as tapping
12
13 into the strongest link between two nodes in the participant's lexicon. However, one
14
15 possible explanation for this finding (an explanation which does not refute the
16
17 underlying assumptions of word association) is that participants who score highly on
18
19 measures of working memory do so not because they have a greater capacity
20
21 available to them, but because they make more efficient use of the resources they
22
23 have. As an example, consider the bank balance of two people just before payday.
24
25 They may both have £200 remaining, but what that £200 represents may well differ.
26
27 One of these people may be paid £2000 per month (i.e. they have a larger financial
28
29 capacity); the other may get paid £1000 per month but spend it grudgingly (i.e. they
30
31 are efficient within the confines of the capacity that they have). In our view it is
32
33 possible that a high score on a working memory test could be achieved if the
34
35 participant was able to use well-travelled links with long term memory for some parts
36
37 of the task in order to keep space in the working memory itself available. In other
38
39 words, efficient use of the connections between input and long-term memory will
40
41 result in stereotypical responses in a word association task *and* may contribute to a
42
43 high score in working memory tests. Indeed, there has been some empirical
44
45 evidence that suggests that performance on tasks of working memory can be
46
47 improved considerably by making use of long-term memory strategies. For example,
48
49 Chase and Ericsson (1981) described the performance of **an individual referred to**
50
51 **as SF** who had an exceptionally large digit span. SF was able to retain long strings
52
53
54
55
56
57
58
59
60

1
2
3 of numbers by converting them to meaningful running times, making use of long-term
4 memory to improve performance on a working memory task. **To SF, a string of 4**
5 **digits might reflect the number of minutes and seconds taken to complete a**
6 **race of a given distance, turning 4 relatively meaningless pieces of information**
7 **into 1 meaningful chunk.**

8
9
10
11
12
13
14
15 **We acknowledge that the aim of the study described here is only**
16 **concerned with the processes involved in lexical retrieval and word selection,**
17 **but the use of RT as a dependent variable in word association tasks also**
18 **measures the time taken to perceive the cue. Thus there are factors that**
19 **potentially influence the response latency that are not attributable to the**
20 **processes we are interested in. However, the cue words used for both the**
21 **standard and the creative association tasks were the same, and the same**
22 **participants took part in both conditions. As a consequence, any influence**
23 **that perception processes had on RT in one condition are likely to have had a**
24 **roughly equivalent impact on the other condition as well, because in essence**
25 **each participant acted as their own control. Thus differences between RT in**
26 **the standard versus the creative word association task are attributable to**
27 **processes occurring that are *not* the same in both versions of the task.**

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45 **It is plausible, of course, to argue that the responses made in either task**
46 **reflecting the conclusion of some strategic decision-making process, and that**
47 **the slower reaction times observed in the creative association task are simply**
48 **because it is a more complex task that requires a greater processing effort**
49 **before completion. For example, it may be that a number of potential**
50 **responses are activated in both tasks, but that the process of discarding**
51 **inappropriate responses in order to rest on a response that is likely to score**
52
53
54
55
56
57
58
59
60

1
2
3 takes a greater number of iterations in the creative task. In fact if this is the
4
5 case, then our RT data do not provide any convincing support for the
6
7 conclusion that word association responses in standard association tasks are
8
9 the first words that are elicited by the cue. However, we consider that this is
10
11 unlikely to be the case. Our basis for this argument is two-fold.
12
13

14
15 Firstly, research has shown that completing a complex task requires
16
17 more extensive use of working memory resources and is more difficult for
18
19 those individuals with low working memory capacity, hence these participants
20
21 are slower and less accurate in completing the task (e.g. Just & Carpenter,
22
23 1992). If the difference between the tasks in our Experiment 1 is simply that
24
25 one is harder than the other, this would imply that those participants in the low
26
27 working memory group would score significantly lower in the creative
28
29 association task than the high working memory group. They did not. Low
30
31 working memory participants would also be significantly slower to complete
32
33 the creative association task than their high working memory counterparts.
34
35 Again, they were not.
36
37
38
39

40
41 The second part of our argument rests on the pattern of stereotypical
42
43 responses observed in the *creative* association task. Stereotypical responses
44
45 in the creative task are essentially errors. If a strategic decision is being made
46
47 in both the standard and the creative task, there ought to be no systematic
48
49 relationship between appropriate stereotypy scores in the standard task and
50
51 erroneous stereotypy scores on the creative task. That is, if the mechanism by
52
53 which a response is generated is the same in both tasks then the likelihood of
54
55 selecting a commonly-associated word is tied to the task demands and not to
56
57 the cognitive processes of the individual respondent. In our data, however, we
58
59
60

1
2
3 observed a significant positive correlation between stereotypy scores in the
4 standard task and stereotypical errors in the creative task. Furthermore, there
5 ought to be no difference between the RT for stereotypical errors and for
6 scoring creative responses - if it is the case that the complexity of the task
7 demands are driving the average response latency up then this complexity
8 should influence qualitatively different responses equally. Again, our data do
9 not match this prediction. RT in the creative word association task was
10 negatively correlated with stereotypy scores in that task, indicating that
11 responses that were quicker also tended to be errors given the instructions for
12 the task. The above patterns in the data are not readily reconciled with the
13 notion that the creative association task is completed using the same method
14 as the standard association task and that the former task is simply more
15 difficult than the latter. It does, however, match with an account that the
16 stereotypical response is activated more quickly and has to be inhibited when
17 the required response is to offer a valid, but uncommon, associate. Errors (i.e.
18 offering a stereotypical response when asked for an uncommon response) in
19 the creative association task reflect trials on which the participant has failed to
20 inhibit the automatic response. Thus such responses are more likely to be
21 observed a) in participants who are skilled in accessing strong intra-lexical
22 links and b) in trials where the response was offered quickly, as additional
23 time has not elapsed to allow for other options to be activated and considered.
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52

53 Our position that word association responses reflect the strongest intra-lexical
54 link for the participant is further corroborated by the findings of Experiment 2, in
55 which we manipulated working memory and imposed a response deadline so that
56
57
58
59
60

1
2
3 participants did not have sufficient opportunity to implement strategic responses. We
4
5 observed no effect of working memory capacity in this experiment. Also of note here
6
7 is that a) fewer stereotypical responses and b) a greater proportion of responses
8
9 scored a point for stereotypy under time pressure. This matches the predictions of
10
11 semantic network models (Collins & Quillian, 1969; Collins & Loftus, 1975; Steyvers
12
13 & Tenenbaum, 2005) in which activation spreads from word to word as a function of
14
15 the strength of the link between them. Strong links allow activation to pass quickly,
16
17 and are therefore reflected in word association cue-response pairings that are high in
18
19 associative strength. These cue-response pairs are likely to be the stereotypical
20
21 responses in a word association task. As the strength of the intra-lexical link
22
23 decreases so too does the speed with which activation can be passed from node to
24
25 node. These cue-response pairs may, in some cases, be weaker for one participant
26
27 than for the population as a whole. Though the activation required for response is
28
29 accrued more slowly, the output that is ultimately generated by the participant will
30
31 still be stereotypical in a proportion of trials. When a response deadline is imposed,
32
33 there is no longer time for the participant to fully activate a relatively weak cue-
34
35 response pairing. In some instances this will result in **a participant** being unable to
36
37 generate the stereotypical response for a given cue, hence fewer stereotypy points
38
39 will be scored on average. However, the majority of the responses that are slow and
40
41 effortful will reflect uncommon cue-associate pairs that *would not* have been
42
43 stereotypical in any case. Thus a greater number of idiosyncratic versus
44
45 stereotypical responses are omitted overall, and the proportion of participant
46
47 responses that score points increases. **It would be worthwhile conducting**
48
49 **research in the future which specifically examines whether this prediction is**
50
51 **borne out by the data. This could be accomplished by systematically changing**
52
53
54
55
56
57
58
59
60

1
2
3 the deadline such that a smaller proportion of responses time-out in each
4
5 iteration of the tasks - if our interpretation is correct, there may be a point
6
7 where *only* idiosyncratic responses are omitted. It might also be interesting
8
9 to determine whether there are predictable characteristics of a) the cues that a
10
11 given participant responds to particularly slowly and b) the types of responses
12
13 that are elicited at longer latencies. This would be of interest not only in
14
15 relation to the allocation of stereotypy points (as we have in the current
16
17 Experiment 2) but also with regard to the effects of word frequency,
18
19 concreteness, word class, and the oft-considered category of response
20
21 (paradigmatic versus syntagmatic, for example) in order to provide a greater
22
23 depth of understanding regarding the structure and dynamics of the lexicon.
24
25
26
27

28 It would appear, therefore, that the assumptions on which word association
29
30 research has been based are supported by the current study. By and large the
31
32 responses that are given by participants *do* reflect the first word that comes to mind.
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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.

- 1
2
3
4 **Fitzpatrick, T. and C. Izura.** 2011. 'Word association in L1 and L2: An exploratory study of
5 response types, response times and inter-language
6 mediation,' *Studies in Second Language Acquisition* 33: 373–98.
7
- 8 **Fitzpatrick, T., D. Playfoot, A. Wray, and M.J. Wright,** 2013. 'Establishing the reliability
9 of word association data for investigating individual and group differences.' *Applied*
10 *Linguistics*, 34(5), 1-29.
11
- 12 **Gewirth, L. R., A. G. Shindler, and D. B. Hier.** 1984. 'Altered patterns of word
13 associations in dementia and aphasia.' *Brain and Language*, 21(2), 307-317.
14
- 15 **Gollan, T. H., D. P. Salmon, and J. L. Paxton.** 2006. 'Word association in early
16 Alzheimer's disease.' *Brain and Language*, 99, 289-303.
17
- 18 **Hartmann, G. W.** 1941. 'A critique of the common method of estimating vocabulary size,
19 together with some data on the absolute word knowledge of educated adults.' *Journal of*
20 *Educational Psychology*, 32, 351-358.
21
- 22 **Hirsh, K. W., and J. T. Tree.** 2001. 'Word association norms for two cohorts of British
23 adults.' *Journal of Neurolinguistics*, 14(1), 1-44.
24
- 25 **Just, M.A. and Carpenter, P.A.** 1992. 'A capacity theory of comprehension: Individual
26 differences in working memory.' *Psychological Review*, 99, 122-149.
27
- 28 **Kane, M.J., D.Z. Hambrick, S.W. Tuholski, O. Wilhelm, T.W. Payne, and R.W. Engle,**
29 2004. 'The generality of working memory capacity: a latent-variable approach to verbal and
30 visuospatial memory span and reasoning.' *Journal of Experimental Psychology: General*,
31 133(2), 189-217.
32
- 33 **Kent, G.H. and Rosanoff, A.J.** 1910. 'A study of association in insanity' *American Journal*
34 *of Insanity*, 67, 37-39, 317-390.
35
- 36 **Meara, P.M.,** 2009. *Connected Words: word associations and second language lexical*
37 *acquisition. Amsterdam: John Benjamins.*
38
- 39 **Merten, T.** 1993. 'Word association responses and psychoticism.' *Personality and*
40 *Individual Differences*, 14, 837-839.
41
- 42 **Meyer, D., and R. Schvaneveldt,** 1971. 'Facilitation in recognizing pairs of words: Evidence
43 of a dependence between retrieval operations.' *Journal of Experimental Psychology*, 90, 227-
44 234.
45
- 46 **Nelson, D. L., C. L. McEvoy, and T. A. Schreiber.** 1998. 'The University of South
47 Floridaword association, rhyme, and word fragment norms'
48 <http://www.usf.edu/FreeAssociation/>
49
- 50 **Nusbaum, H.C., D.B. Pisoni, and C.K. Davis,** 1984. 'Sizing up the Hooiser mental lexicon:
51 Measuring the familiarity of 20,000 words.' In *Research on Speech Perception, Progress*
52 *Report 10*, (pp. 357-376). Bloomington, IN: Speech Research Laboratory, Indiana University.
53
54
55
56
57
58
59
60

1
2
3
4 **Playfoot, D.** and **C. Izura**, 2013. 'Imageability, age of acquisition and frequency factors in
5 acronym comprehension.' *Quarterly Journal of Experimental Psychology*, 66(6), 1131-1145.
6

7 **Postman L. J.** and **G. Keppel** (eds). 1970. Norms of Word Association. Academic Press.
8

9 **Schneider, W., A. Eschman,** and **A. Zuccolotto**, 2002. *E-Prime 1.0*. Pittsburgh, PA:
10 Psychological Software Tools.
11

12 **Steyvers, M.,** and **J. Tenenbaum**, 2005. 'The large-scale structure of semantic networks:
13 statistical analyses and a model of semantic growth' *Cognitive Science*, 29 (1), 41-78.
14

15 **Wettler, M., Rapp, R.,** and **Sedlmeier, P** 2005. 'Free word associations correspond to
16 contiguities between words in texts.' *Journal of Quantitative Linguistics*, 12, 111-122.
17
18
19
20

21 Appendix 1

22
23

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

50
51
52
53
54
55
56
57
58
59
60

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)

For Peer Review