

## ORIGINAL RESEARCH REPORT

# Does a Working Memory Load *Really* Influence Semantic Priming? A Self-replication Attempt

Tom Heyman\*, Kirsten Goossens\*, Keith A. Hutchison† and Gert Storms\*

The present paper describes two attempts to replicate a recent study of ours in the semantic priming domain (Heyman, Van Rensbergen, Storms, Hutchison, & De Deyne, 2015). In that study, we observed that semantic priming for forward associates (e.g., *panda-bear*) completely evaporated when participants' working memory was taxed, whereas backward (e.g., *baby-stork*) and symmetric associates (e.g., *cat-dog*) showed no ill-effects of a secondary task. This was the case for relatively long and short stimulus onset asynchronies (i.e., 1,200 ms and 200 ms, respectively). The results thus suggested that prospective target activation is, contrary to what some theories of semantic memory posit, not an automatic process. However, the two replication studies reported here cast serious doubt on this conclusion. A Bayesian analysis of all the available data indicated that there is at least substantial evidence for a priming effect in every condition, except for forward associates in the short SOA condition. The null hypothesis is still supported in the latter condition, though the replication studies weakened the evidence for a null effect. The theoretical implications of these findings are discussed.

**Keywords:** semantic priming; automaticity; working memory; lexical decision

## Introduction

The semantic priming effect is an often studied phenomenon by cognitive psychologists, presumably because it tells us something about the structure of semantic memory (among other things). It refers to the improvement in speed (and accuracy) when responding to a target stimulus that is preceded by a semantically related prime stimulus (see McNamara, 2005, for a review). For instance, people tend to recognize the word *dog* faster when they first see the related word *cat*. Over the years, several accounts of the semantic priming effect have been advanced, many of which were inspired by Collins and Loftus' (1975) spreading activation model of semantic processing. According to this model, conceptual knowledge is stored in a network of interconnected nodes, where each node corresponds to a concept (e.g., *cat*). If one processes a concept, for instance, by reading the word *cat*, the matching node gets activated and activation will spread to connected nodes, which entails that semantically related concepts such as *dog* are (partly) activated. A spreading activation mechanism can readily explain the semantic priming effect if one assumes that

the pre-activation of related concepts can result in a head start. Put differently, people respond faster to the target *dog* when its corresponding node was pre-activated due to the presentation of the word *cat*.

The notion of an automatic spreading activation mechanism resurfaces in several other priming accounts (e.g., Neely & Keefe's, 1989, hybrid three-process theory), however it has drawn some criticism over the years. For instance, Stolz and Besner (1999) considered *automatic* semantic processing a myth. Their stance was based on findings indicating that semantic activation requires attentional control. Indeed, traditional definitions of automaticity (Neely, 1977; Posner & Snyder, 1975) imply that (spatial) attention should play no role. To address this issue, Neely and Kahan (2001) updated the criteria for automaticity. They suggested that a process, such as semantic activation, can be considered automatic if "it is unaffected by the intention for it to occur and by the amount and quality of the attentional resources allocated to it" (Neely & Kahan, 2001, p. 89).

Several recent studies have examined whether semantic activation, and the priming effect it produces, indeed fulfill Neely and Kahan's (2001) automaticity criteria, but the results are mixed (e.g., Augustinova & Ferrand, 2014; Besner & Reynolds, 2017; Heyman, Hutchison, & Storms, 2016; Heyman et al., 2015; White & Besner, 2016). Here, we will focus on Heyman et al.'s study (2015), which tested whether the processes underlying semantic priming are

\* University of Leuven, Tiensestraat 102, 3000 Leuven, BE

† Montana State University, Bozeman, MT 59717-3440, US

Corresponding author: Tom Heyman  
([tom.heyman@ppw.kuleuven.be](mailto:tom.heyman@ppw.kuleuven.be))

capacity free (i.e., the second part of Neely and Kahan's definition) by limiting working memory capacity via a secondary task. Using a lexical decision task, Heyman et al. found that semantic priming for forward associates (i.e., the target is an associate of the prime, but not vice versa, henceforth FA; e.g., *panda-bear*) completely disappeared when participants' working memory was taxed via a dot memory task. In contrast, priming effects for backward associates (i.e., the prime is an associate of the target, but not vice versa, henceforth BA; e.g., *baby-stork*) and symmetric associates (i.e., the prime is an associate of the target and vice versa, henceforth SYM; e.g., *cat-dog*) remained intact under a high load.

The observation that FA pairs yielded no priming effect under a high load poses problems for an automatic spreading activation account of semantic priming. Given that FA priming is presumably the result of prospective processes (i.e., processes initiated upon the presentation of the prime like spreading activation; Thomas, Neely, & O'Connor, 2012), one would have expected a (small) priming effect in the high load condition, *if target activation occurs automatically*. The load manipulation didn't seem to impact BA and SYM priming, suggesting that retrospective processes (i.e., processes initiated upon the presentation of the target) are impervious to constraints on working memory. Put differently, many theories have assumed that prospective target activation occurs automatically and that retrospective priming is the result of strategic, non-automatic processes, yet Heyman et al.'s findings imply the opposite. The somewhat surprising pattern of results and the associated theoretical repercussions arguably merit replication.

### Experiment 1

Heyman et al. (2015) manipulated Type of Association (BA, FA, or SYM), Load (high or low), Relatedness (related or unrelated), and Stimulus Onset Asynchrony (200 or

1,200 ms; henceforth SOA). We found a significant Load  $\times$  Type of Association  $\times$  Relatedness interaction using by-subject and by-item analyses of variance (ANOVAs), whereas none of the other interactions reached statistical significance across both subject and item ANOVAs. Further inspection of the Load  $\times$  Type of Association  $\times$  Relatedness interaction suggested that the high load interfered with FA priming, but not BA or SYM priming, in both SOA conditions (see **Table 1**). It is this pattern of results that we seek to replicate here.

The present study is a direct replication of Heyman et al.'s (2015) experiment, though it was conducted at Montana State University (USA) instead of the University of Leuven (Belgium). Hence, it differs from the original in a number of ways: the language (English instead of Dutch), the participant pool, the testing environment, etc. It should be noted that we were actually planning a version of the experiment in which the primes would be masked. So, the goal of the replication was to merely re-establish the "baseline" findings in a pilot study, thereby using a convenience sample.

### Method

#### Participants

Seventy-eight students from Montana State University (31 men, 47 women, mean age = 20 years) participated for partial completion of a requirement for an introductory psychology course.

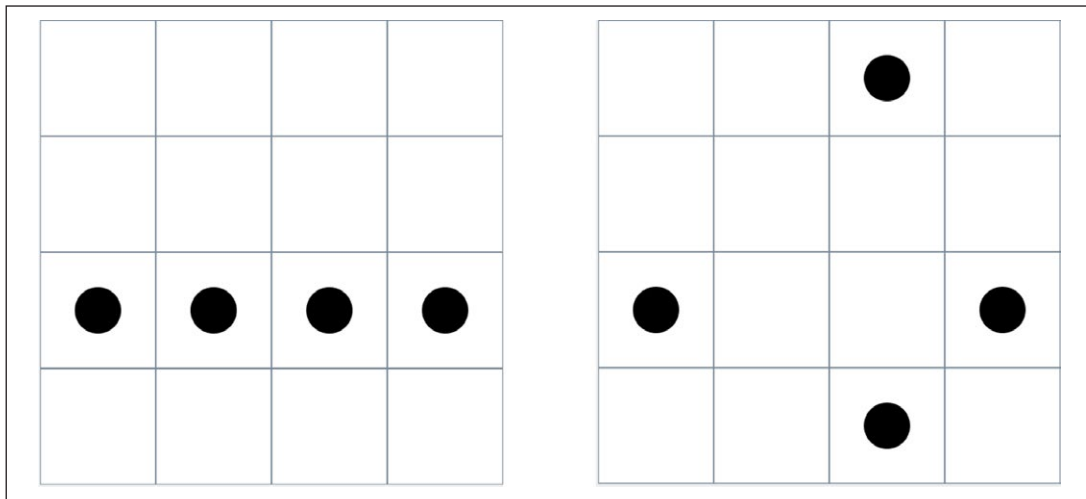
#### Materials

We used the same prime-target pairs as in Hutchison, Heap, Neely, and Thomas' (2014) study, after which Heyman et al.'s (2015) study was modelled. The stimulus set consisted of 120 critical pairs (i.e., 40 BA, 40 FA, and 40 SYM pairs), 80 filler SYM pairs, and 120 filler word-non-word pairs. Critical BA, FA, and SYM targets were matched on length and lexical decision response

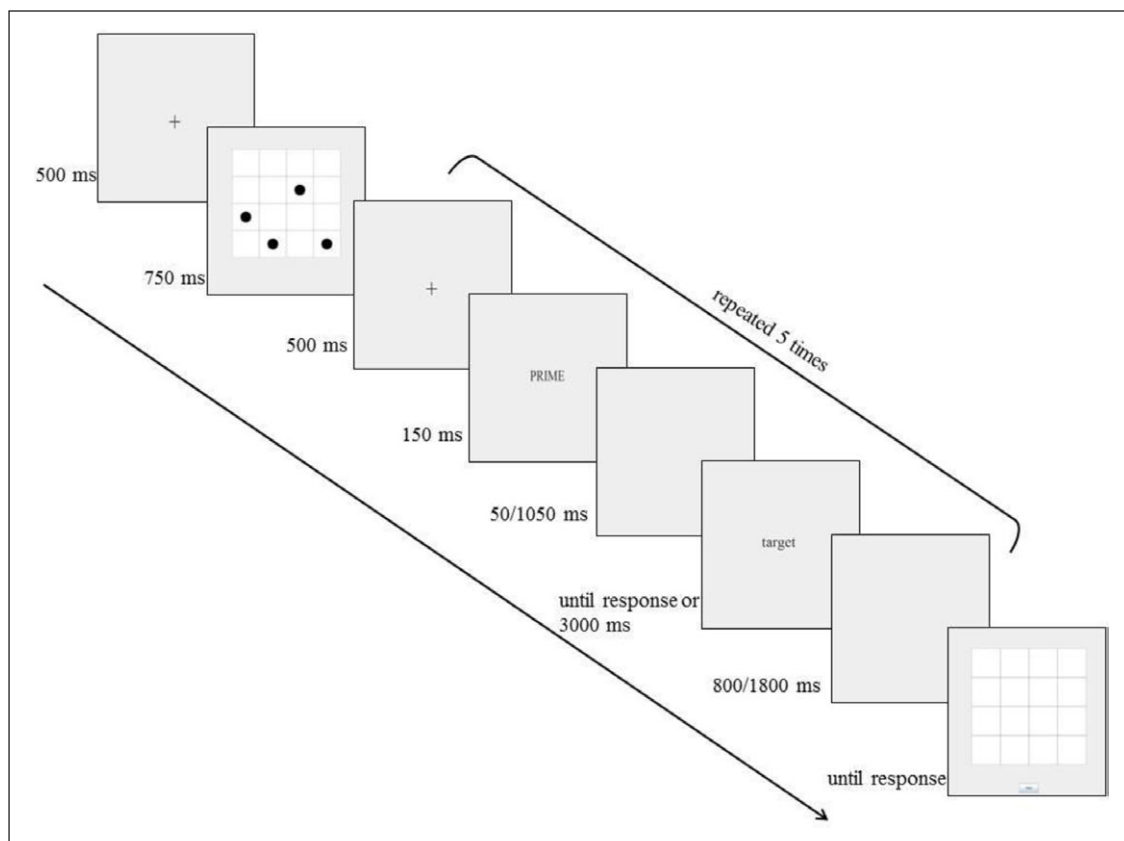
**Table 1:** Mean Response Times to the Critical Targets as a Function of Load, Type of Association, Relatedness, and SOA in Heyman et al. (2015).

	SOA 200		SOA 1200	
	HIGH LOAD	LOW LOAD	HIGH LOAD	LOW LOAD
FA pairs				
Unrelated	837	810	801	778
Related	835	755	799	733
<b>Priming</b>	<b>2 [-41, 45]</b>	<b>55 [24, 85]</b>	<b>2 [-33, 38]</b>	<b>44 [15, 74]</b>
BA pairs				
Unrelated	869	842	857	822
Related	814	792	808	779
<b>Priming</b>	<b>55 [17, 93]</b>	<b>50 [17, 83]</b>	<b>49 [21, 78]</b>	<b>43 [15, 71]</b>
SYM pairs				
Unrelated	851	811	837	810
Related	817	769	761	755
<b>Priming</b>	<b>34 [-3, 71]</b>	<b>43 [10, 75]</b>	<b>76 [44, 108]</b>	<b>55 [26, 83]</b>

*Note:* The by-subject priming effects per condition are printed in bold with the 95% confidence intervals in brackets.



**Figure 1:** Example of a low load dot memory pattern (left panel) and a high load pattern (right panel).



**Figure 2:** Flow of the experiment.

times (RTs). The 40 critical pairs per association type were randomly divided into eight groups of five pairs. Unrelated prime-target pairs were created by recombining primes and targets within every group. These groups were then rotated through all Load  $\times$  Relatedness  $\times$  SOA conditions across participants.

As in Heyman et al. (2015), working memory was taxed using a dot memory task. Participants saw a  $4 \times 4$  matrix containing four dots. The dots either formed a straight line (i.e., the low load condition) or they were semi-randomly distributed across the 16 possible fields (i.e., the high load condition; see **Figure 1** for an illustration).

#### Procedure

The procedure was exactly the same as in Heyman et al. (2015; see **Figure 2**). The entire experiment consisted of 64 cycles, which were all structured as follows. First, participants saw a fixation cross for 500 ms. Then, a dot pattern, either a low load or a high load type, was shown for 750 ms. Next, participants got five typical priming trials, each consisting of five events: a fixation cross for 500 ms, an uppercase prime for 150 ms, a blank screen for 50 or 1,050 ms depending on the SOA condition, a lowercase target that required a response (i.e., a lexical decision), and a blank screen for 800 or 1,800 ms again

depending on the SOA condition. Finally, participants got an empty  $4 \times 4$  matrix and were asked to reproduce the dot pattern by clicking on the appropriate fields. There was also an intercycle interval of 2,000 ms during which a blank screen was shown.

For the lexical decision task, participants were told to respond as fast and accurately as possible by pressing the arrow keys (left for *word*, right for *non-word*). For the dot memory task, only accuracy was stressed. The 64 cycles were divided into two blocks of 32. SOA was held constant within a block and SOA order (i.e., short SOA in block 1 versus long SOA in block 1) was counterbalanced over participants. A random half of the dot patterns per block was of the low load type, the other half of the high load type. Furthermore, the order in which prime-target pairs appeared was randomly determined for each participant separately.

## Results

### Dot memory task

As expected, the mean number of correctly localized dots in the low load condition ( $M = 3.8$ ) was significantly higher than in the high load condition ( $M = 3.1$ ),  $t(77) = 12.41$ ,  $p < .001$ .<sup>1</sup> By-participant one sample  $t$ -tests on the number of correctly localized dots were carried out to test whether everyone performed significantly above chance on the high load patterns (i.e.,  $\mu = 1$ ). This was the case for all but one participant ( $t(31) = 1.91$ ,  $p = .07$ ), who was consequently removed from the analyses.

### Lexical decision task

Eight participants were omitted from the analyses because they made more than 15% errors on the lexical decision task.<sup>2</sup> Two participants had an exceptionally high error rate of 97% suggesting that they had switched

response keys. Rather than discarding their data, we chose to reverse score their responses. Note that the calculation of the error rates was based on all items, but all further analyses are conducted on the 120 crucial items only.

In a next step, errors and outliers were removed employing the same procedure as did Heyman et al. (2015). First, only trials with a response within the 3,000 ms window and an RT above 250 ms were retained. Secondly, response times longer than 3  $SD$ s above the by-participant average were also removed. As a result, 5.3% of the data were excluded from further analysis.

By-subject and by-item ANOVAs were carried out on the remaining data using the `aov_car` function from the `afex` package with the Greenhouse–Geisser correction on the degrees of freedom (Singmann, Bolker, Westfall, & Aust, 2016). Due to missing data, one additional participant had to be omitted from the analyses. Four predictor variables were included in the ANOVAs: Relatedness (related vs. unrelated; manipulated within subjects and within items), Load (high load vs. low load; manipulated within subjects and within items), Type of Association (BA, FA, vs. SYM; manipulated within subjects and between items), and SOA (200 ms vs. 1,200 ms; manipulated within subjects and within items).

The analyses revealed three significant main effects: Load ( $F_s(1, 67) = 39.83$ ,  $MSE = 31,854$ ,  $p < .001$ ,  $\eta_p^2 = .37$ ;  $F_i(1, 117) = 104.79$ ,  $MSE = 7,068$ ,  $p < .001$ ,  $\eta_p^2 = .47$ ), Relatedness ( $F_s(1, 67) = 20.27$ ,  $MSE = 23,211$ ,  $p < .001$ ,  $\eta_p^2 = .23$ ;  $F_i(1, 117) = 31.29$ ,  $MSE = 10,991$ ,  $p < .001$ ,  $\eta_p^2 = .21$ ), and Type of Association ( $F_s(1.96, 131.58) = 21.43$ ,  $MSE = 10,908$ ,  $p < .001$ ,  $\eta_p^2 = .24$ ;  $F_i(2, 117) = 8.06$ ,  $MSE = 21,400$ ,  $p < .001$ ,  $\eta_p^2 = .12$ ). As suggested by the conditional means (see **Table 2**), lexical decision RTs tended to be slower in the unrelated condition (compared

**Table 2:** Mean Response Times to the Critical Targets as a Function of Load, Type of Association, Relatedness, and SOA in Experiment 1.

	SOA 200		SOA 1200	
	HIGH LOAD	LOW LOAD	HIGH LOAD	LOW LOAD
FA pairs				
Unrelated	883	823	891	821
Related	868	788	834	775
<b>Priming</b>	<b>15 [-29, 60]</b>	<b>35 [5, 65]</b>	<b>57 [8, 106]</b>	<b>46 [20, 72]</b>
BA pairs				
Unrelated	915	850	900	869
Related	891	824	858	848
<b>Priming</b>	<b>25 [-36, 86]</b>	<b>25 [-7, 57]</b>	<b>42 [12, 72]</b>	<b>21 [-8, 49]</b>
SYM pairs				
Unrelated	874	817	872	838
Related	852	771	847	789
<b>Priming</b>	<b>22 [-24, 69]</b>	<b>45 [19, 72]</b>	<b>25 [-11, 61]</b>	<b>48 [16, 80]</b>

Note: The by-subject priming effects per condition are printed in bold with the 95% confidence intervals in brackets.

to the related condition), in the high load condition (compared to the low load condition), and for BA targets (compared to FA and SYM targets). These findings mimic Heyman et al.'s (2015), except for the, theoretically less interesting, main effect of SOA, which was statistically significant in Heyman et al.'s analyses.

Crucially, none of the interactions reached significance. This includes the Load  $\times$  Type of Association  $\times$  Relatedness interaction we sought to replicate ( $F_s(1.92, 128.75) = 0.92, MSE = 11,068, p = .40, \eta_p^2 = .01$ ;  $F_i(2, 117) = 1.19, MSE = 7,799, p = .31, \eta_p^2 = .02$ ). Moreover, judging from the point estimates, the largest priming effect (i.e., 57 ms) was found for FA pairs in the high load  $\times$  1,200 ms SOA condition (see **Table 2**). In contrast, the original study found a null effect under those circumstances (i.e., 2 ms, see **Table 1**). FA pairs in the high load  $\times$  200 ms SOA condition did show a non-significant priming effect, but that was also the case in several other conditions (see **Table 2**). Even so, the interpretation of non-significant effects is problematic (an issue that we will revisit in the *Joint analyses* section).

Taken together, Experiment 1 did not succeed in replicating some of the key findings from Heyman et al. (2015). One could make the argument that the discrepancies arose because of differences in the stimuli, the tested language, the participant pool, etc. To be fair, though, we certainly did not anticipate that any of these factors could confound the results. It should be pointed out that the eventual sample size in Experiment 1 was somewhat smaller than in the original study (i.e., 68 and 80, respectively). To address these (potential) concerns, a follow-up experiment was conducted.

## Experiment 2

In Experiment 2, we tried to emulate the original study as closely as possible. The stimuli were exactly the same, participants were recruited from the same pool, the testing environment was the same, etc. Two important differences with the original are that we doubled the number of participants and that the present study was pre-registered on Open Science Framework (see <https://osf.io/sg28r/>).<sup>3</sup>

## Method

### Participants

One hundred fifty-seven first-year psychology students from the University of Leuven participated in return for course credit, three others received 4 euro for their participation (23 men, 137 women, mean age = 19 years).

### Materials

Prime-target pairs were taken from Heyman et al. (2015). BA, FA, and SYM targets (40 each) were matched on baseline RTs, length, contextual diversity, and word frequency (see Heyman et al. for more details). The dot patterns were created as described above.

### Procedure

The procedure was the same as in Experiment 1.

## Results

### Dot memory task

Again, the mean number of correctly localized dots in the low load condition ( $M = 3.8$ ) was significantly higher than in the high load condition ( $M = 3.3$ ),  $t(159) = 13.52, p < .001$ . All participants performed significantly above chance on the high load patterns ( $ps < .05$ ).

### Lexical decision task

Four participants were omitted from the analyses because they made more than 15% errors on the lexical decision task.<sup>4</sup> Subsequently, error responses and outliers were removed from the analyses using the same criteria as outlined above (4.8% of the data).

By-subject and by-item ANOVAs again showed three main effects: Load ( $F_s(1, 155) = 46.58, MSE = 17,936, p < .001, \eta_p^2 = .23$ ;  $F_i(1, 117) = 73.84, MSE = 2,635, p < .001, \eta_p^2 = .39$ ), Relatedness ( $F_s(1, 155) = 123.97, MSE = 11,175, p < .001, \eta_p^2 = .44$ ;  $F_i(1, 117) = 131.14, MSE = 2,888, p < .001, \eta_p^2 = .53$ ), and Type of Association ( $F_s(1.90, 294.39) = 27.06, MSE = 11,719, p < .001, \eta_p^2 = .15$ ;  $F_i(2, 117) = 11.22, MSE = 7,351, p < .001, \eta_p^2 = .16$ ). The results, summarized in **Table 3**, mimic those from Experiment 1 in that lexical decision RTs tended to be slower in the unrelated condition (compared to the related condition), in the high load condition (compared to the low load condition), and for BA targets (compared to FA and SYM targets).

Besides those main effects, only the SOA  $\times$  Type of Association  $\times$  Relatedness interaction reached statistical significance ( $F_s(1.97, 305.35) = 3.49, MSE = 12,499, p = .03, \eta_p^2 = .02$ ;  $F_i(2, 117) = 4.17, MSE = 3,404, p = .02, \eta_p^2 = .07$ ). Thus, the Load  $\times$  Type of Association  $\times$  Relatedness interaction was once again not statistically significant ( $F_s(1.96, 304.52) = 0.18, MSE = 11,895, p = .84, \eta_p^2 = .00$ ;  $F_i(2, 117) = 0.24, MSE = 2,740, p = .79, \eta_p^2 = .00$ ). As in Experiment 1, FA pairs showed a large priming effect in the high load  $\times$  1,200 ms SOA condition (i.e., 47 ms based on the point estimate), but a non-significant priming effect in the high load  $\times$  200 ms SOA condition.

## Joint analyses

Taken together, the crucial Load  $\times$  Type of Association  $\times$  Relatedness interaction was not statistically significant in either experiment. This is especially remarkable for Experiment 2, given that we used the exact same stimuli. Does the observed null effect constitute a non-replication though? After all, the difference between a significant and a non-significant effect is not necessarily statistically significant in its own right (Gelman & Stern, 2006). To address, this issue, we combined the original data with those from the exact replication (i.e., Experiment 2) and conducted similar ANOVAs. The sole difference with the previous analyses is that we added a fifth variable called Experiment (original vs. replication; manipulated between subjects and within items). Besides the main effects of Load, SOA, Relatedness, and Type of Association, only the Experiment  $\times$  Load  $\times$  Type of Association  $\times$  Relatedness interaction was statistically significant in both by-subject and by-item ANOVAs:  $F_s(2, 466.86) = 3.18, MSE = 10,644,$

**Table 3:** Mean Response Times to the Critical Targets as a Function of Load, Type of Association, Relatedness, and SOA in Experiment 2.

	SOA 200		SOA 1200	
	HIGH LOAD	LOW LOAD	HIGH LOAD	LOW LOAD
FA pairs				
Unrelated	794	779	794	759
Related	771	745	747	715
<b>Priming</b>	<b>23 [-2, 48]</b>	<b>34 [11, 57]</b>	<b>47 [24, 70]</b>	<b>44 [22, 66]</b>
BA pairs				
Unrelated	821	813	820	781
Related	784	757	804	763
<b>Priming</b>	<b>37 [12, 62]</b>	<b>56 [30, 82]</b>	<b>16 [-8, 41]</b>	<b>19 [-2, 40]</b>
SYM pairs				
Unrelated	813	785	802	776
Related	778	730	758	725
<b>Priming</b>	<b>34 [7, 61]</b>	<b>56 [35, 76]</b>	<b>44 [21, 68]</b>	<b>51 [29, 73]</b>

Note: The by-subject priming effects per condition are printed in bold with the 95% confidence intervals in brackets.

**Table 4:** Average By-Subject Priming Effects as a Function of Load, Type of Association, and SOA.

	SOA 200		SOA 1200	
	HIGH LOAD	LOW LOAD	HIGH LOAD	LOW LOAD
FA pairs	16 [-4, 35] <i>0.23 (0.12)</i>	40 [24, 56] <i>&gt;100 (37)</i>	38 [19, 56] <i>&gt;100 (0.12)</i>	45 [30, 59] <i>&gt;100 (7)</i>
BA pairs	39 [18, 60] <i>43 (6)</i>	47 [30, 65] <i>&gt;100 (8)</i>	31 [15, 47] <i>68 (24)</i>	26 [11, 40] <i>27 (9)</i>
SYM pairs	31 [12, 51] <i>8 (0.60)</i>	50 [35, 65] <i>&gt;100 (3)</i>	48 [32, 65] <i>&gt;100 (&gt;100)</i>	51 [36, 66] <i>&gt;100 (92)</i>

Note: 95% confidence intervals in brackets, Bayes Factors in italics, and the original Bayes Factors in parentheses.

$p = .04$ ,  $\eta_p^2 = .01$ ;  $F(2, 117) = 3.17$ ,  $MSE = 4,691$ ,  $p = .05$ ,  $\eta_p^2 = .05$ . The latter suggests that the Load  $\times$  Type of Association  $\times$  Relatedness interaction is indeed significantly different in the two experiments. As such, the present study failed to replicate the original, critical findings.

This conclusion is in itself not very satisfying. What did we actually *learn*? In other words, how should we change our beliefs about the priming effects, or lack thereof, in the various conditions? Thus far, we conducted all the analyses within a frequentist statistical framework, but the question of belief revision can be addressed more elegantly when adopting a Bayesian perspective. Therefore, in a final set of analyses, we pooled the data from all experiments – because they can be considered as part of one overarching study – and quantified what we learned about the priming effect in each condition. **Table 4** shows the by-subjects average priming effect per condition with corresponding Bayesian one sample  $t$ -test

using the defaults of BayesFactor’s `ttestBF` function (Morey & Rouder, 2015). The resulting Bayes Factors (henceforth BFs) indicate the relative plausibility of the data under the alternative hypothesis (i.e., there is a priming effect in a certain condition;  $H_A$ ) versus under the null hypothesis (i.e., there is no priming effect in that condition;  $H_0$ ). The higher the BF, the more we ought to believe that there is a priming effect, and vice versa. As a reference, BFs below 1/3 or above 3 indicate substantial evidence for, respectively,  $H_0$  and  $H_A$ , whereas BFs below 1/10 or above 10 provide strong evidence for one hypothesis or the other (see Wetzels et al., 2011).

**Table 4** also contains the BFs based only on the original study’s results (i.e., the values in parentheses). That way, we can evaluate how we should update our beliefs when taking the two replication studies into account. The Bayesian analyses of Heyman et al.’s (2015) data provide substantial support for a null priming effect in two conditions: FA pairs under a high load with an SOA

**Table 5:** Average By-Subject Effect of Load on Priming as a Function of Type of Association and SOA.

	SOA 200	SOA 1200
FA pairs	24 [-1, 49] <i>0.38 (0.72)</i>	7 [-16, 30] <i>0.08 (0.82)</i>
BA pairs	8 [-18, 35] <i>0.08 (0.13)</i>	-5 [-27, 17] <i>0.07 (0.13)</i>
SYM pairs	19 [-5, 42] <i>0.21 (0.13)</i>	3 [-20, 26] <i>0.07 (0.20)</i>

*Note:* positive values indicate a smaller priming effect in the high load condition. 95% confidence intervals in brackets, Bayes Factors in italics, and the original Bayes Factors in parentheses.

of 200 ms or 1,200 ms. There is also anecdotal evidence for a null effect for SYM pairs in the high load × 200 ms SOA condition, but the data in all the other conditions (strongly) point in the direction of a priming effect.

Including the data from Experiments 1 and 2, weakens or completely overturns the evidence for the null effects. Most importantly, there is now decisive evidence *in favor* of a priming effect for FA pairs in the high load × 1,200 ms SOA condition, whereas our belief in a null priming effect for FA pairs in the high load × 200 ms SOA condition has taken a hit. That said, the BF still prefers  $H_0$  over  $H_A$  in the latter condition. Given that there is no reason to discard the original data, we can thus conclude that there is substantial or even stronger evidence for a priming effect in all conditions, except for FA pairs under a high load with a 200 ms SOA, in which case the BF still provides substantial evidence for a null priming effect.

So far, it is still unclear whether the load manipulation differentially affects semantic priming in the various conditions. Most notably, we haven't established that the priming effect for FA pairs in the low load × 200 ms SOA condition differs substantially/significantly from the null effect in the high load condition. Therefore, we conducted paired samples *t*-test on the priming effects in the two load conditions, again using the data from all three experiments (see **Table 5**). The analyses revealed Bayes Factors that always substantially support the null hypothesis (i.e., no differential effect of load), except for FA pairs in the 200 ms SOA condition. In the latter, the Bayes factor is inconclusive, yet leaning towards the null hypothesis as well (BF = 0.38). This finding raises further questions about the original conclusion that the load manipulation interfered with FA priming.

### Discussion

In light of a growing call for self-replication (Cesario, 2014), the present study sought to establish the reliability of our previously reported finding that taxing working memory selectively impacts semantic priming (Heyman et al., 2015). More specifically, we originally found that priming for FA pairs evaporated under a high working memory load, whereas BA and SYM pairs did yield significant priming effects in such circumstances. Two direct replications, one

in a different language with conceptually similar stimuli and one in the same language with identical stimuli, revealed no significant Load × Type of Association × Relatedness interaction, though. Most remarkable were the large priming effects for FA pairs in the high load × long SOA condition, because they originally generated no (significant) priming effect in this condition.

What theoretical implications do these non-replications have? The complete lack of semantic priming for FA pairs in the high load conditions led us to conclude that prospective priming mechanisms require cognitive resources. Hence, target pre-activation is not an automatic process according to Neely and Kahan's (2001) criteria (see Heyman et al., 2015 for more details). The results of the two replication studies seem to discredit this conclusion. However, target pre-activation is presumed to decay rapidly, so only the short SOA condition is actually relevant in this respect. It is therefore noteworthy that a joint analysis, featuring all the available data, still pointed to a null priming effect for FA pairs in the high load × short SOA condition, even though the replication studies weakened our conviction. That is, our belief in the null hypothesis, relative to the alternative, changed from about 8:1 to 4:1 (see **Table 4**). In contrast, there was substantial evidence for priming in all other conditions, including for FA pairs in the low load × short SOA condition. Furthermore, the load manipulation never led to a (significant) decrease (or increase) in semantic priming (see **Table 5**). Taken together, the present findings no longer allow us to make any strong claims about the non-automaticity of target activation or any other prospective priming mechanism for that matter.

The clear non-replication of the null FA priming effect in the high load × long SOA condition also has theoretical implications. More concretely, the complete lack of a load effect on FA priming seems to be at odds with findings from Hutchison and colleagues (2014, see also Hutchison, 2007). They found a significant positive correlation between attentional control and FA priming, meaning that people performing well on attentional control tasks (i.e., OSPAN, antisaccade, and Stroop) showed larger priming effects for FA pairs. It seems therefore evident to predict that depleting attentional resources by imposing a working memory load should reduce FA priming. This was clearly not the case, at least in the long SOA condition of the replication studies (see **Tables 4** and **5**), so how can we explain the apparent discrepancy? There are of course inherent caveats associated with correlational and experimental designs. On the one hand, the link between FA priming and attentional control might be spurious. That is, variables correlated with attentional control such as vocabulary knowledge could have been responsible for the observed relation with FA priming (but see Hutchison et al., 2014 for a discussion of such potential confounds). Because taxing working memory does not affect vocabulary knowledge, one would actually expect no load effect, which would explain the ostensibly inconsistent findings. On the other hand, the experimental manipulation may have (partly) missed the mark. Attentional control, as captured by Hutchison and colleagues (2007, 2014),

likely involves many abilities including verbal fluency, inhibition of distraction, and attention shifting, which could all be responsible for the observed relation with FA priming. As such, the narrower, non-verbal nature of the dot memory task may not (sufficiently) constrain the processes involved in the standard lexical decision priming paradigm, thus producing divergent results. Alternatively, the load manipulation might have been less effective than anticipated. Participants who did not significantly perform above chance were removed from the analyses, but chance in this context equates to correctly remembering the location of one dot. This is arguably not a high working memory load, so the criterion is not very strict. In other words, we assumed that participants would comply and really try to remember all four dots, but this may be naive. Future research might consider addressing these issues by adopting a verbal secondary task and/or more stringent, a priori determined performance criteria.

Sometimes failures to replicate certain (null) effects are attributed to undiscovered moderators. Despite designing Experiment 2 such as to resemble the original study as closely as possible (up to the experimental apparatus, the light and sound conditions, and the geographic location of the test cubicle), there were (of course) subtle differences. For instance, we weren't able to use the same experimenter, testing took place during different seasons, the original experiment was preceded by an unrelated category learning experiment, etc. In theory, it is possible that any of these factors are responsible for the (partly) divergent findings. But if that is the case, a very unlikely scenario in our opinion, further investigation of this phenomenon becomes pointless. When an effect is so fragile that it only occurs under very specific conditions, shouldn't we take a hard look in the mirror and think about the (ecological) validity and relevance of our "findings" (see Simons, 2014)? In any case, we expected that unspecified moderators like the experimenter's behavior or gender would not play a role. In other words, we tacitly assumed that the reported findings in Heyman et al. (2015) would generalize to other contexts. Future research could elucidate whether we underestimated the impact of unknown moderators or whether sampling error merely led us astray, to a certain extent at least.

### Data Accessibility Statement

All materials, data, and analysis scripts can be found on this paper's project page on the Open Science Framework <https://osf.io/jvrrc/>.

### Notes

- <sup>1</sup> All analyses were carried out in R (Version 3.3.1; R Development Core Team, 2016). The analysis scripts as well as the raw data are available at: <https://osf.io/jvrrc/>.
- <sup>2</sup> In Heyman et al. (2015) none of the participants were removed because they all met the 85% accuracy criterion.
- <sup>3</sup> Following Frank and Saxe's (2012) suggestion to train students in replicating recent findings, this study was conducted in the context of KG's master's thesis

project. As such, some documentation on OSF was written in Dutch.

- <sup>4</sup> Note that this criterion was not specified in the pre-registration, yet given the high error rates for some participants (e.g., 59%), we would argue that blindly following the pre-registration is not appropriate in this particular instance as it would introduce noise.

### Acknowledgements

We thank Gabriel Abate, Sarah Freitas, Alexandra Hay, Bethany Johannesson, Owen McGaffigan, and Alish Slade for scheduling and running participants. We also want to thank Pieter Moors for his comments on an earlier version of the manuscript, and for checking the analysis script. T. Heyman is a Postdoctoral Fellow of the Research Foundation-Flanders (FWO-Vlaanderen). Correspondence should be addressed to Tom Heyman, Department of Experimental Psychology, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium. E-mail: [tom.heyman@kuleuven.be](mailto:tom.heyman@kuleuven.be).

### Competing Interests

The authors have no competing interests to declare.

### Author Contributions

T. Heyman and K. A. Hutchison developed the study concept. T. Heyman and K. Goossens performed the data analysis and interpretation. T. Heyman drafted the manuscript, and K. A. Hutchison and G. Storms provided critical revisions. All authors approved the final version of the manuscript for submission.

### References

- Augustinova, M., & Ferrand, L.** (2014). Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science*, 23(5): 343–348. DOI: <https://doi.org/10.1177/0963721414540169>
- Besner, D., & Reynolds, M.** (2017). Is semantic activation from print capacity limited? Evidence from the psychological refractory period paradigm. *Psychonomic Bulletin & Review*, 24(3): 907–913. DOI: <https://doi.org/10.3758/s13423-016-1178-3>
- Cesario, J.** (2014). Priming, replication, and the hardest science. *Perspectives on Psychological Science*, 9(1): 40–48. DOI: <https://doi.org/10.1177/1745691613513470>
- Collins, A. M., & Loftus, E. F.** (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6): 407–428. DOI: <https://doi.org/10.1037/0033-295X.82.6.407>
- Frank, M. C., & Saxe, R.** (2012). Teaching replication. *Perspectives on Psychological Science*, 7(6): 600–604. DOI: <https://doi.org/10.1177/1745691612460686>
- Gelman, A., & Stern, H.** (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician*, 60(4): 328–331. DOI: <https://doi.org/10.1198/000313006X152649>
- Heyman, T., Hutchison, K. A., & Storms, G.** (2016). Is semantic priming (ir)rational? Insights from the speeded word fragment completion task. *Journal*



- of *Experimental Psychology: Learning, Memory, and Cognition*, 42(10): 1657–1663. DOI: <https://doi.org/10.1037/xlm0000260>
- Heyman, T., Van Rensbergen, B., Storms, G., Hutchison, K. A., & De Deyne, S.** (2015). The influence of working memory load on semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3): 911–920. DOI: <https://doi.org/10.1037/xlm0000050>
- Hutchison, K. A.** (2007). Attentional control and the relatedness proportion effect in semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4): 645–662. DOI: <https://doi.org/10.1037/0278-7393.33.4.645>
- Hutchison, K. A., Heap, S. J., Neely, J. H., & Thomas, M. A.** (2014). Attentional control and asymmetric associative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3): 844–856. DOI: <https://doi.org/10.1037/a0035781>
- McNamara, T. P.** (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press. DOI: <https://doi.org/10.4324/9780203338001>
- Morey, R. D., & Rouder, J. N.** (2015). BayesFactor: Computation of Bayes Factors for Common Designs (Version 0.9.12–2). Retrieved from: <https://CRAN.R-project.org/package=BayesFactor>.
- Neely, J. H.** (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106(3): 226–254. DOI: <https://doi.org/10.1037/0096-3445.106.3.226>
- Neely, J. H., & Kahan, T. A.** (2001). Is semantic activation automatic? A critical re-evaluation. In: Roediger, H. L., Nairne, J. S., Neath, I., & Surprenant, A. M. (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*, 69–93. Washington: American Psychological Association. Retrieved from: <http://content.apa.org/books/10394-005>. DOI: <https://doi.org/10.1037/10394-005>
- Neely, J. H., & Keefe, D. E.** (1989). Semantic context effects on visual word processing: A hybrid prospective-retrospective processing theory. In: *Psychology of learning and motivation*, 24: 207–248. New York: Academic Press. DOI: [https://doi.org/10.1016/S0079-7421\(08\)60538-1](https://doi.org/10.1016/S0079-7421(08)60538-1)
- Posner, M. I., & Snyder, C. R. R.** (1975). Attention and cognitive control. In: Solso, R. L. (Ed.), *Information processing and cognition: The Loyola symposium*, 55–86. Hillsdale, NJ: Erlbaum.
- R Development Core Team.** (2016). R: A language and environment for statistical computing (Version 3.3.1). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: <http://www.R-project.org>.
- Simons, D. J.** (2014). The value of direct replication. *Perspectives on Psychological Science*, 9(1): 76–80. DOI: <https://doi.org/10.1177/1745691613514755>
- Singmann, H., Bolker, B., Westfall, J., & Aust, F.** (2016). afex: Analysis of Factorial Experiments (Version 0.16–1). Retrieved from: <https://CRAN.R-project.org/package=afex>.
- Stolz, J. A., & Besner, D.** (1999). On the myth of automatic semantic activation in reading. *Current Directions in Psychological Science*, 8(2): 61–65. DOI: <https://doi.org/10.1111/1467-8721.00015>
- Thomas, M. A., Neely, J. H., & O'Connor, P.** (2012). When word identification gets tough, retrospective semantic processing comes to the rescue. *Journal of Memory and Language*, 66(4): 623–643. DOI: <https://doi.org/10.1016/j.jml.2012.02.002>
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J.** (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 *t*-Tests. *Perspectives on Psychological Science*, 6(3): 291–298. DOI: <https://doi.org/10.1177/1745691611406923>
- White, D., & Besner, D.** (2016). Attentional constraints on semantic activation: Evidence from Stroop's paradigm. *Acta Psychologica*. DOI: <https://doi.org/10.1016/j.actpsy.2016.08.008>

#### Peer review comments

The author(s) of this paper chose the Open Review option, and Streamlined Review option, and all peer review comments are available at: <http://doi.org/10.1525/collabra.96.pr>

**How to cite this article:** Heyman, T., Goossens, K., Hutchison, K. A., & Storms, G. (2017). Does a Working Memory Load *Really* Influence Semantic Priming? A Self-replication Attempt. *Collabra: Psychology*, 3(1): 18, DOI: <https://doi.org/10.1525/collabra.96>

**Submitted:** 07 June 2017

**Accepted:** 29 June 2017

**Published:** 07 August 2017

**Copyright:** © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

