Arousal Affects Short-Term Serial Recall

Éric R. Landry, Dominic Guitard, and Jean Saint-Aubin

Université de Moncton, Moncton, New Brunswick

Authors' note

This research was supported by Discovery grant RGPIN-2015-04416 from the Natural Sciences and Engineering Research Council of Canada to JSA. While working on the manuscript, ÉRL was supported by an Undergraduate Student Research Award from NSERC, and DG was supported by a postdoctoral fellowship from NSERC.

Correspondence concerning this article should be addressed to Éric R. Landry at eel9542@umoncton.ca, Dominic Guitard at dominic.guitard@umoncton.ca or Jean Saint-Aubin at jean.saint-aubin@umoncton.ca

Open practice Statement

The stimuli are provided in the appendix and the data are available on the Open Science Framework project page.

(https://osf.io/6gcen/?view_only=8899134cc338445da9bfde66b5731b02).

Abstract

Arousal affects our lives in a variety of ways; it can direct our attention to what is important in our environment and help us remember it more clearly. However, it remains unclear how arousal impacts short-term memory. Here we addressed this gap in our knowledge by contrasting four hypotheses: the *Arousal Hypothesis*, the *Priority-Binding Hypothesis*, the *Rehearsal Hypothesis* and the *Rapid-Processing Hypothesis*. To distinguish between these competing accounts, we conducted two immediate serial recall experiments in which we manipulated arousal (low-arousal words vs. high-arousal words), list composition (pure vs. mixed), and presentation rate (200ms vs. 1000ms). Overall, participants were better at recalling arousing information, regardless of list type or presentation rate. Our results provide clear evidence in favor of the arousal hypothesis which suggests that arousing information benefits from biologically induced enhancements at encoding.

Keywords: Short-Term Memory, Immediate Serial Recall, Arousal, Emotion.

Public Significance Statement

When studying verbal memory, emotional arousal refers to the level of excitement elicited by a word. However, the effect of arousal on short-term memory is uncertain. The current assessment reveals that arousal enhances performance in a short-term ordered recall task at fast and slower presentation rates, and in both pure and mixed lists.

Arousal Affects Short-Term Serial Recall

Since the early days of psychology, the effect of emotions on memory has been thoroughly debated (e.g., Titchener, 1895). Today, it is well established that emotions impact memory. For instance, emotions can enhance the vividness of episodic memory (e.g., Phelps & Sharot, 2008), or inhibit the encoding of peripheral information (e.g., Easterbrook, 1959). In verbal memory studies, emotional stimuli are usually assessed along three isolable components: valence, arousal, and dominance (e.g., Warriner et al., 2013). In this context, valence refers to the level of pleasantness elicited by a word; it ranges from pleasant words like *sunshine* to unpleasant words like *jail*. Arousal refers to the level of excitement elicited by a stimulus; it ranges from calm words like *librarian* to exciting words like *shotgun*. Dominance refers to the level of control; it ranges from feeling submissive with words like *lobotomy* to feeling in control with words like *successful*. To the best of our knowledge, in short-term memory, valence is the only dimension which has been investigated. Therefore, the influence of arousal and dominance is unknown. This is surprising because in long-term memory studies, arousal has a larger impact on performance than the other two dimensions (Bradley et al., 1992; Cahill & McGaugh, 1998). We attempted to fill this gap by exploring the influence of arousal in immediate serial recall, a canonical short-term memory task.

It is well-known that lexical or long-term memory factors impact performance in short-term memory tasks. For instance, in immediate serial recall, words are better recalled than nonwords and frequent words are better recalled than rare words (e.g., Guérard & Saint-Aubin, 2012; Hulme et al., 2003; Saint-Aubin & Poirier, 2000). In this context, emotional dimensions should be no exception since they are considered to be semantic dimensions (see, e.g., Ishiguro & Saito, 2021; Mejerus & D’Argembeau, 2011). However, recently, Bireta, Guitard, Neath, and

Surprenant (2021) challenged this assumption by showing that, with proper controls, valence does not impact performance in immediate serial recall. Does this mean that with proper controls, none of the emotional dimensions could impact short-term ordered recall tasks?

In the next sections, we briefly review four competing hypotheses accounting for the impact of arousal on memory. As will be seen, all these hypotheses predict better recall of high- than low-arousal words under typical conditions. However, the predictions diverge when list composition and presentation speed vary. More specifically, according to localist accounts like the arousal hypothesis, the locus of the effect is at the item level, while other accounts like the priority binding hypothesis suggest that it depends on list composition. Therefore, in addition to the typical pure list condition in which all words are high- or low-arousal, a mixed list condition was included in which high- and low-arousal words systematically alternate. If the arousal effect occurs at the item level, it should be observed in both conditions as previously observed with the lexicality effect (Hulme et al., 2003), while if it occurs at the list level, it should vanish in pure lists as previously observed with the production effect (Fawcett, 2013). In addition, it has been suggested that high-arousal words are processed faster than low-arousal words. Therefore, the advantage of high-arousal words should be larger or limited to faster presentation conditions (e.g., 200 ms per word) compared to longer presentation conditions (e.g., 1000 ms per word). To the contrary, other models suggest that arousal affects the rehearsal process requiring a slower presentation rate to influence performance. Consequently, in addition to arousal and list composition, presentation speed was manipulated.

Arousal Hypothesis

According to the arousal hypothesis, derived from the arousal theory, arousing information elicits a biological reaction facilitating memory consolidation (Cahill & McGaugh, 1998). For instance, arousing events would elicit an amygdala reaction that would in turn influence activity of some brain regions like the hippocampus. In support of this hypothesis, fMRI studies showed that the amygdala is activated in response to emotional events, and that such activations are linked to enhanced long-term memory (e.g., Canli et al., 2000). In addition, successful encoding is also linked to amygdala activation in response to emotional arousal (e.g., Labar & Cabeza, 2006). Under this view, in immediate serial recall, high-arousal words should be better recalled than low-arousal words in pure lists and mixed lists, because the effect is assumed to occur at the word level. Furthermore, the proposed mechanisms are not assumed to be time dependent. Therefore, while overall performance could vary with presentation rate, the advantage of high-arousal words should remain.

Priority-Binding Hypothesis

Hadley and MacKay (2006) suggested the priority-binding hypothesis, according to which the advantage of high-arousal words is due to interference with neighbouring low-arousal words. According to this view, item recall relies on an associative link between each list item and its episodic context. The binding process would be sequential and time consuming. Therefore, with a rapid presentation rate, the time required to establish an association between an item node and its context can significantly exceed stimulus duration. This would cause forgetting of the following word because its memory trace would have decayed too much before a link with its episodic context could be formed. It is further assumed that high-arousal words like taboo words are prioritized. In a nutshell, the presentation of a high-arousal word would activate the emotional reaction system that would delay the activation of the currently primed binding nodes of low-arousal items, to enhance binding of the nodes for high-arousal words. Processing of the previous word would only resume once binding of the high-arousal word is completed. Obviously, with pure lists of high-arousal words, such a prioritization process is useless.

Hadley and MacKay (2006) tested the priority-binding hypothesis in immediate recall using pure and mixed lists combined with slow and rapid presentation of taboo and neutral words. As predicted by the priority-binding hypothesis, taboo words were better recalled than neutral words in mixed, but not in pure lists. However, they did not test the effect of presentation rate with mixed lists.

Although supporting the priority-binding hypothesis, taboo words differ from neutral words on more than one emotional dimension. Therefore, it is impossible to know if Hadley and MacKay's (2006) results are due to arousal, to another emotional component or to another factor because they only controlled for familiarity and word length. The necessity to control for more factors is well illustrated by the work of Bireta et al. (2021) on the impact of valence on memory. In the literature, negatively valenced words were usually better remembered than neutral or positively valenced words, although some studies failed to find this advantage and others found the reverse pattern (for a literature review, see Bowen, Kark, & Kensinger, 2018). However, in these studies, the stimuli were equated on a limited number of dimensions. When Bireta et al. controlled for many dimensions, which is now the gold standard in the field, they observed no effect of valence on memory.

Rehearsal Hypothesis and Processing-Time Hypothesis

Hadley and MacKay (2006) also tested the rehearsal and the processing-time hypotheses. According to the rehearsal hypothesis, at presentation, high-arousal words receive more attention than low-arousal words, increasing their probability of being rehearsed. Therefore, at a slower presentation rate, the advantage of high-arousal words should be greater because there are more rehearsing opportunities, and the effect should decrease at a faster presentation speed because less time would be available for rehearsal. Contrary to the rehearsal hypothesis, the processing-time hypothesis suggests that high-arousal information is encoded faster. Therefore, the advantage of high-arousal words should be greater at faster presentation rates.

In the current study, two experiments were conducted to provide a comprehensive examination of the impact of arousal on short-term ordered recall performance. In both experiments, participants performed an immediate serial recall task. We varied presentation rate as a between-experiments factor (200 ms in Experiment a and 1000 ms in Experiment b), and list composition (pure vs. mixed) and arousal (high vs. low) as within-participant factors. In Experiment 1, high- and low-arousal words were equated on 30 dimensions and list composition varied randomly from participant to participant. To further test the robustness of our results, Experiment 2 replicated Experiment 1, but with a new set of stimuli equated on the same number of dimensions.

Experiment 1a

Method

Participants. Forty-four participants (39 female, 5 male, mean age: 22.77) were recruited through the platform Prolific. The necessary sample size was estimated with a power analysis conducted with G*Power (Faul et al., 2009): The a priori analysis suggests that a sample of 44 participants would have a power of .90 to detect a medium effect size of arousal (d = 0.50). The same sample size calculation was used for all subsequent experiments. Participants ought to be between 18 and 30 years old; to be from the United States; to have English as their first language, normal or corrected-to-normal vision, a Prolific approval rate of at least 90%, and not to have reading or writing related disorders, cognitive impairments or dementia. These selection criteria were used for all experiments. Participants were paid 3 £. Participants gave their free and informed consent and this and subsequent experiments were approved by the research ethics committee of the Université de Moncton.

Stimuli. One hundred and twenty high-arousal words and 120 low-arousal words were selected from Warriner at al.'s (2013) word pool, a large word bank containing words rated for emotional valence, arousal and dominance. As shown in Appendix A, our stimuli varied on arousal, while being matched on 30 dimensions. The mean arousal rating was 5.71 (SD = 0.54, range 5.05—7.57) for the high-arousal words and 2.97 (SD = 0.34, range 2.15—3.5) for the low-arousal words. For each participant, lists were created by randomly sampling without replacement from the appropriate word pool.

Design. A 2 x 2 x 6 repeated measures design was used with list type (pure vs. mixed), arousal (high vs. low), and serial position (1 to 6) as factors. The experiment consisted of 40 experimental trials preceded by 4 practice trials. The 40 experimental lists were presented in a different random order to each participant.

Procedure. Participants were tested in a single 20-minute session. The experiment was run online through PsyToolkit (Stoet, 2010, 2017). The stimuli were displayed in capital letters in white 20-point Times New Roman on a black background. Trials were self-initiated by pressing the spacebar. Each trial began with the presentation of a plus sign at the center of the screen for 500ms, after which each word appeared sequentially at a rate of five items per second (200ms on/ 0ms off). After the presentation of the final word of the list, the message "Type the first word" appeared at the top of the screen over a blank text box in which the typed word would appear. At the bottom of the screen, it was indicated that the word “SKIP” should be typed if a word could not be remembered. After typing the first word, participants pressed the enter key and the word disappeared and was replaced by another box in which the second answer could be typed. This procedure continued until the sixth word was recalled, at which point the enter key triggered the beginning of the next trial.

Data Analysis

Before any analyses were conducted, misspellings were corrected if they could be unambiguously identified (e.g., instead of yearbook: yyearbook, yearboo, xearbook). The results were similar with or without spelling corrections, albeit overall performance was slightly better with the corrections. The data with or without spelling corrections are available on the Open Science Framework page. Responses were then scored with a strict serial recall criterion according to which a word must be recalled at its presentation position to be considered correct. Responses were also scored with a free recall scoring according to which a word can be recalled in any position to be considered correct.

We used Bayes Factor analyses via the “BayesFactor” package with the default parameters (Version 0.9.12-4.2; see Morey & Rouder, 2018; Rouder, Morey, Speckman, & Province, 2012). For all analyses, the proportional error was below 5% and we report BF with values superior to 1 representing evidence in favour of an effect and values inferior to 1 representing evidence against an effect. The interpretation of our results was guided by the recommendations of Kass and Raftery (1995): $BF < 3$ indicates weak or anecdotal evidence; $3 \leq BF < 20$ indicates positive evidence; $20 \leq BF < 150$ indicates strong evidence; and $BF > 150$ indicates very strong evidence. For all BF ANOVAs, main effects and interaction models were tested by omitting these effects one at a time from the full model and participants were treated as a random factor. Lastly, for all analyses, we also report the corresponding *F* ratios and partial eta squares via the “ez” package (Version 4.4-0, Lawrence, 2016) as additional descriptive information.

Results

Strict Serial Recall Scoring. As shown in Figure 1, participants’ performance was superior for high arousal words (*M* = 0.39, *SD* = 0.13) than for low arousal words (*M* = 0.35, *SD* = 0.12). Performance was also equivalent between mixed lists (*M* = 0.37, *SD* = 0.13) and pure lists (*M* = 0.37, *SD* = 0.11). The 2 x 2 x 6 ANOVA revealed very strong evidence in favour of a main effect of arousal, $F(1,43) = 46.33$, $\eta_p^2 = .52$, BF > 1000, and a main effect of serial position, $F(5,215) = 161.10$, $\eta_p^2 = .79$, BF > 1000. However, there was no main effect of list composition, $F < 1$, $\eta_p^2 = .02$, BF = 0.032, or interaction between arousal and list composition, $F < 1$, $\eta_p^2 = .00$, BF = 0.033, arousal and position, $F(5,215) = 1.48$, $\eta_p^2 = .03$, BF < 0.000, or position and list composition, $F < 1$, $\eta_p^2 = .02$, BF < 0.000. There was also strong evidence against the three-way interaction between arousal, list composition and serial position, $F < 1$, $\eta_p^2 = .01$, BF < 0.000.

Free Recall Scoring. As shown in Figure 1, overall, participants’ performance was better for high-arousal words (*M* = 0.52, *SD* = 0.11) than for low-arousal words (*M* = 0.45, *SD* = 0.10), but performance was equivalent between mixed lists (*M* = 0.48, *SD* = 0.11) and pure lists (*M* = 0.49, *SD* = 0.10). The 2 x 2 x 6 ANOVA on free recall performance echoes results with strict serial scoring. More specifically, there was a main effect of arousal, $F(1,43) = 76.29$, $\eta_p^2 = .64$, BF > 1000, a main effect of serial position, $F(5,215) = 129.19$, $\eta_p^2 = .75$, BF > 1000, but no main effect of list composition, $F(1, 43) = 2.70$, $\eta_p^2 = .06$, BF = 0.051, and no interaction. More precisely, there was no interaction between arousal and list composition, $F < 1$, $\eta_p^2 = .02$, BF =

0.041, arousal and position, $F(5,215) = 1.26$, $\eta_p^2 = .03$, BF < 0.000, position and list composition, $F(5,215) = 1.03$, $\eta_p^2 = .02$, BF < 0.000, and arousal, list composition and serial position, $F < 1$, $\eta_p^2 = .01$, BF < 0.000.

Experiment 1b

Experiment 1b was identical to Experiment 1a, except that words were presented for a much longer duration. With a slower presentation rate, participants are more likely to rehearse list items and their rehearsal strategy should be more efficient (see, e.g., Tan & Ward, 2008). Therefore, according to the rehearsal hypothesis, the beneficial effect of arousal should be larger than what has been found in Experiment 1a with a fast presentation rate (Hadley & MacKay, 2006). However, according to the processing-time hypothesis, the advantage enjoyed by high-arousal words with a fast presentation rate in Experiment 1a should decrease because there would be enough time to encode low-arousal words.

Method

Participants, Materials and Procedure. Forty-four participants who did not take part in the previous experiment (35 female, 5 male, 4 other, mean age: 21.84) were recruited through the platform Prolific. The method is identical to Experiment 1a except that stimuli were presented at a rate of one word every second (1000 ms on/ 0ms off).

Results

Strict Serial Recall Scoring. As shown in Figure 1, participants recalled more high-arousal words ($M = 0.54$, $SD = 0.20$) than low-arousal words ($M = 0.48$, $SD = 0.21$), but performance was comparable in mixed lists ($M = 0.50$, $SD = 0.20$) and pure lists ($M = 0.52$, $SD = 0.20$). The 2 x 2 x 6 ANOVA showed a main effect of arousal, $F(1,43) = 35.63$, $\eta_p^2 = .45$, BF >

1000, and of serial position, *F*(5,215) = 59.69, η_p^2 = .58, BF > 1000. However, there was positive evidence against a main effect of list composition, *F*(1,43) = 3.30, η_p^2 = .07, BF = 0.309, and the interaction between arousal and list composition, *F*(1,43) = 1.16, η_p^2 = .03, BF = 0.058. There was strong evidence against the interaction between arousal and position, *F*(5,215) = 1.20, η_p^2 = .03, BF < 0.000, position and list composition, *F*(5,215) = 1.22, η_p^2 = .03, BF < 0.000, and arousal, list composition and serial position, *F*(5,215) = 1.81, η_p^2 = .04, BF = 0.003.

Free Recall Scoring. As illustrated in Figure 1, overall, participants' performance was higher for high-arousal words (*M* = 0.64, *SD* = 0.15) than low-arousal words (*M* = 0.57, *SD* = 0.17) and it did not vary between mixed lists (*M* = 0.60, *SD* = 0.16) and pure lists (*M* = 0.61, *SD* = 0.16). The 2 x 2 x 6 ANOVA revealed a main effect of arousal, *F*(1,43) = 44.30, η_p^2 = .51, BF > 1000, and serial position, *F*(5,215) = 38.20, η_p^2 = .47, BF > 1000, but evidence against a main effect of list composition, *F*(1,43) = 1.46, η_p^2 = .03, BF = 0.050. In line with the strict recall scoring analysis, there was no interaction between arousal and list composition, *F*< 1, η_p^2 = .00, BF = 0.031, arousal and position, *F*(5,215) = 1.68, η_p^2 = .04, BF = 0.001, position and list composition, *F*< 1, η_p^2 = .02, BF < 0.000, and arousal, list composition and serial position, *F*< 1, η_p^2 = .01, BF < 0.000.

Discussion

For both presentation rates (200 ms, 1000 ms), the analyses provide evidence against the interaction between arousal and list composition. This pattern of results observed with strict and free serial recall scoring does not lend support to the priority binding hypothesis, but could be accommodated by the arousal hypothesis (Hadley & MacKay, 2006). After the second experiment, a cross-experiments analysis with presentation rate as a factor will be presented.

This analysis provides a critical test of the processing-time hypothesis and the rehearsal hypothesis.

Experiment 2

Results of Experiment 1 are clear and provide compelling evidence for the arousal hypothesis (Hadley & MacKay, 2006). However, given the novelty of the effect, an additional step is required to ensure that it is not due to any specificity of the stimuli. The stimuli specificity issue refers to the outcome of an experiment that would be dependent on some uncontrolled and unknown idiosyncratic features from a given stimulus set (see e.g., Guitard, Saint-Aubin, Tehan, & Tolan, 2018; Neath, Hockley, & Ensor, 2021). Therefore, we replicated Experiment 1 with a novel set of low and high arousal words.

Experiment 2a

Experiment 2a was modeled after Experiment 1a. However, a new set of low and high arousal words was used.

Method

Participants. Forty-four participants who did not take part in the previous experiments (36 female, 7 male, 1 preferred not to report, mean age: 23.18) were recruited through the platform Prolific.

Stimuli. A new set of stimuli was created. More exactly, 108 high-arousal words and 108 low-arousal words were selected from Warriner at al.'s (2013) word pool. The stimuli were matched on the same criteria as the previous stimuli set (see Appendix B). The mean arousal rating was 5.92 (SD = 0.53, range 5.06-7.45) for the high-arousal words and 3.11 (SD = 0.32, range 1.95-3.50) for the low-arousal words. For each participant, lists were created by randomly sampling the words without replacement from the appropriate condition.

Design, Procedure, Data Analysis. The design, the procedure, and the data analysis were identical to those used in Experiment 1a except for the following change. In this experiment, the number of experimental trials was reduced from 40 to 36 because the word pools are slightly smaller than in the previous experiments.

Results

Strict Serial Recall Scoring. Results of Experiment 2 are illustrated in Figure 2. Overall, results with the new word pools are very similar to those observed in Experiment 1. Participants recalled more high arousal words (*M* = 0.38, *SD* = 0.13) than low arousal words (*M* = 0.34, *SD* = 0.13), and their performance was comparable between mixed lists (*M* = 0.36, *SD* = 0.13) and pure lists (*M* = 0.37, *SD* = 0.13). The 2 x 2 x 6 ANOVA revealed very strong evidence in favour of a main effect of arousal, $F(1,43) = 34.48$, $\eta_p^2 = .45$, BF > 1000, and serial position, $F(5,215) = 168.11$, $\eta_p^2 = .80$, BF > 1000. There was no main effect of list composition, $F(1,43) = 1.85$, $\eta_p^2 = .04$, BF = 0.071, and there was strong or very strong evidence against all interactions; there was no interaction between arousal and list composition $F(1,43) = 1.22$, $\eta_p^2 = .03$, $BF_{01} = 0.049$, arousal and position, $F < 1$, $\eta_p^2 = .01$, BF < 0.000, position and list composition, $F(5,215) = 1.57$, $\eta_p^2 = .04$, BF < 0.000, and arousal, list composition and serial position, $F(5,215) = 1.17$, $\eta_p^2 = .03$, BF = 0.001.

Free Recall Scoring. As shown in Figure 2, participants were more likely to recall high-arousal words (*M* = 0.49, *SD* = 0.12) than low-arousal words (*M* = 0.43, *SD* = 0.13), but their performance was equivalent in mixed lists (*M* = 0.45, *SD* = 0.13) and pure lists (*M* = 0.46, *SD* =

0.12). The 2 x 2 x 6 ANOVA revealed a main effect of arousal, $F(1,43) = 55.09$, $\eta_p^2 = .56$, BF > 1000, and of serial position, $F(5,215) = 130.46$, $\eta_p^2 = .75$, BF > 1000, but no main effect of list composition, $F(1,43) = 1.32$, $\eta_p^2 = .03$, BF = 0.042. Again, results showed positive to very strong evidence against all interactions; there was no interaction between arousal and list composition $F(1,43) = 1.65$, $\eta_p^2 = .04$, BF = 0.050, arousal and position, $F(5,215) = 1.23$, $\eta_p^2 = .03$, BF < 0.000, position and list composition, $F(5,215) = 1.32$, $\eta_p^2 = .03$, BF < 0.000, and arousal, list composition and serial position, $F(5,215) = 1.30$, $\eta_p^2 = .03$, BF = 0.001.

Experiment 2b

Experiment 2 follows the same structure as the previous experiments. Therefore, in Experiment 2b, we replicated Experiment 2a with a slower presentation rate of 1000 ms per word.

Method

Participants, Stimuli, Design, Procedure, and Data Analysis. Forty-four participants who did not take part in previous experiments (29 female, 15 male, mean age: 23.66) were recruited through the platform Prolific. The method was identical to the one used in Experiment 2a except that words were presented at a rate of one word every second.

Results

Strict Serial Recall Scoring. Performance was superior for high arousal words (M = 0.52, SD = 0.18) than for low arousal words (M = 0.47, SD = 0.18), and did not differ between mixed lists (M = 0.50, SD = 0.18) and pure lists (M = 0.49, SD = 0.18). The ANOVA showed a main effect of arousal, $F(1,43) = 25.85$, $\eta_p^2 = .38$, BF > 1000, and of serial position, $F(5,215)$ =

77.43, η_p^2 = .64, BF > 1000. Once more, there was strong evidence against a main effect of list composition, *F* < 1, η_p^2 = .01, BF = 0.036, and positive to very strong evidence against all interactions. There was no interaction between arousal and list composition *F* < 1, η_p^2 = .02, BF = 0.057, arousal and position, *F* < 1, η_p^2 = .02, BF < 0.000, position and list composition, *F* < 1, η_p^2 = .01, BF < 0.000, and arousal, list composition and serial position, *F*< 1, η_p^2 = .02, BF < 0.000.

Free Recall Scoring. As shown in Figure 2, participants were better at recalling high-arousal words (*M* = 0.60, *SD* = 0.15) relative to low-arousal words (*M* = 0.54, *SD* = 0.16) and performance was nearly identical between mixed lists (*M* = 0.57, *SD* = 0.16) and pure lists (*M* = 0.57, *SD* = 0.15). The ANOVA showed a main effect of arousal, *F*(1,43) = 42.80, η_p^2 = .50, BF > 1000, and of serial position, *F*(5,215) = 47.87, η_p^2 = .53, BF > 1000, but no main effect of list composition, *F* < 1, η_p^2 = .00, BF = 0.025. Evidence was found against all interactions with the following values: arousal and list composition, *F*(1,43) = 1.10, η_p^2 = .03, BF = 0.052, arousal and position, *F* < 1, η_p^2 = .01, BF < 0.000, position and list composition, *F* < 1, η_p^2 = .01, BF < 0.000, and arousal, list composition and serial position, *F*< 1, η_p^2 = .01, BF < 0.000.

Discussion

Results of Experiment 2, with a new set of stimuli, parallel those observed in the first experiment. This within-study replication contributes to establishing that the effect of arousal on short-term ordered recall is reproducible. This demonstration was essential before memory theories are adapted to account for this effect (Simons, 2014). Before discussing the theoretical implications of current findings, a further test of the hypotheses is necessary. This test was provided by a cross-experiments analysis assessing the impact of presentation rate.

Cross-Experiments Analysis

In this section we reported a cross-experiments analysis for both strict serial scoring and free recall scoring. More specifically, we pooled together data from all experiments and we computed a 2 × 2 × 2 mixed-design ANOVA with list composition (pure vs. mixed) and arousal (low vs. high) as within factors and presentation rate (200 ms vs. 1000 ms) as a between factor.

The cross-experiments analysis was aimed at directly evaluating the validity of three of the competing hypotheses highlighted in the introduction. According to the rehearsal hypothesis, the beneficial effect of arousal should be larger with a slower presentation rate relative to a fast presentation rate. This would occur because a slow presentation rate allows more rehearsal opportunity which would benefit more high-arousal words. Contrary to the rehearsal hypothesis, the processing-time hypothesis suggests that the recall advantage of high-arousal words over low-arousal words should decrease when more time is available because this would afford more time to process low-arousal words. Lastly, according to the arousal hypothesis, time should not interact with arousal, because high-arousal words benefit from biologically induced enhancements at encoding, irrespective of presentation speed.

Strict Serial Recall Scoring

When collapsed across experiments, participants' performance was superior for high-arousal words (M = 0.46, SD = 0.17) relative to low-arousal words (M = 0.41, SD = 0.18). Participants were also better when words were presented at a rate of one word per second (M = 0.50, SD = 0.19) than when words were presented at a rate of five words per second (M = 0.37, SD = 0.12). However, performance was the same in mixed lists (M = 0.44, SD = 0.17) and pure lists (M = 0.43, SD = 0.18).

The ANOVA confirmed these trends with very strong evidence in favour of the presence of a main effect of arousal, $F(1,174) = 136.21$, $\eta_p^2 = .44$, BF > 1000, and presentation rate, $F(1,174) = 31.54$, $\eta_p^2 = .15$, BF > 1000, but strong evidence against a main effect of list composition, $F(1,174) = 2.29$, $\eta_p^2 = .01$, BF = 0.044. Critically, there was strong evidence against an interaction between presentation rate and arousal, $F < 1$, $\eta_p^2 = .00$, BF = 0.022. In addition, there was strong evidence against all other interactions with the following values: arousal and list composition, $F(1,174) = 2.08$, $\eta_p^2 = .01$, BF = 0.034, presentation rate and list composition, $F < 1$, $\eta_p^2 = .00$, BF = 0.016, presentation rate, arousal, and list composition, $F(1,174) = 1.64$, $\eta_p^2 = .01$, BF = 0.031.

Free Recall Scoring

Overall, free recall performance was better for high-arousal words (*M* = 0.56, *SD* = 0.15) than for low-arousal words (*M* = 0.50, *SD* = 0.15), and for the slower presentation rate (*M* = 0.59, *SD* = 0.16) compared to the faster presentation rate (*M* = 0.47, *SD* = 0.11). Again, performance did not differ between mixed lists (*M* = 0.53, *SD* = 0.15) and pure lists (*M* = 0.53, *SD* = 0.15).

The ANOVA showed very strong evidence of a main effect of arousal, $F(1,174) = 208.24$, $\eta_p^2 = .55$, BF > 1000, and presentation rate, $F(1,174) = 30.23$, $\eta_p^2 = .15$, BF > 1000, but strong evidence against a main effect of list composition, $F(1,174) = 2.62$, $\eta_p^2 = .02$, BF = 0.037. Most importantly, there was strong evidence against the critical interaction between presentation rate and arousal, $F < 1$, $\eta_p^2 = .00$, BF = 0.017. There was also strong evidence against the remaining hypotheses: arousal and list composition, $F < 1$, $\eta_p^2 = .01$, BF = 0.017, presentation rate and list composition, $F < 1$, $\eta_p^2 = .00$, BF = 0.019, and presentation rate, arousal, and list composition, $F(1,174) = 2.61$, $\eta_p^2 = .02$, BF = 0.046.

General Discussion

Our results clearly establish that while valence does not impact short-term ordered recall (Bireta et al., 2021), arousal does have a reliable effect. In their discussion, Bireta et al. acknowledged that they cannot dismiss the possibility that their null effect was due to their implementation of many controls in stimulus selection. Therefore, the remaining dimension on which they differ would be insufficient to translate into a sizeable effect of valence. This hypothesis is even more possible in view of the results of Guitard, Gabel, Saint-Aubin, Surprenant, and Neath (2018) who abolished the very robust syllable-based word length effect after controlling 20 dimensions of their stimuli. Here, the presence of a reliable effect of arousal with two sets of stimuli controlled on 30 dimensions further support Bireta et al. 's conclusion that valence—when operationalized by contrasting positive and negative words—does not affect immediate memory. Taken together, our results and those of Bireta et al. could be seen as an additional empirical demonstration that arousal and valence are two independent dimensions of semantics impacting lexical or memory processes (see, e.g., Ishiguro & Saito, 2021; Kuperman, Estes, Brysbaert, & Warrier, 2014).

The presence of a large effect of arousal on item information in pure and mixed lists diverges from results with taboo words which are better recalled in mixed, but not in pure lists (Hadley & MacKay, 2006). How can we account for this discrepancy? It can be argued that because taboo words differ from neutral words on many dimensions, and because of their provocative quality, they are more distinct from neutral words than high-arousal words are from low-arousal words. It is well-known that local distinctiveness has a large and reliable effect on short-term memory performance (e.g., Saint-Aubin, Yearsley, Poirier, Cyr, & Guitard, 2021). In mixed-lists, taboo words contrast very distinctively against a background of neutral words, but the same distinctiveness effect cannot operate with pure lists. With well-controlled stimuli like ours, the contrast is less obvious, thus diminishing distinctiveness.

In our results, the presence of a large effect of arousal with a fast and a slow presentation rate does not support the processing-time hypothesis suggesting that high-arousal words are processed faster than low-arousal words. Similarly, results are contrary to the predictions of the rehearsal hypothesis because with a slower presentation rate, allowing more opportunities for rehearsal (Tan & Ward, 2008), the effect of arousal was not larger than with a rapid presentation rate. The presence of an arousal effect in pure lists is also contrary to the priority-binding hypothesis, but fits well with the arousal hypothesis (Cahill & McGaugh, 1998; Canli et al., 2000). Finally, the presence of sawtooth serial position curves in mixed lists is analogous to previous findings with lexicality (Hulme et al., 2003) and fits well with the arousal hypothesis suggesting an effect at the item level. Overall, the better recall of high- than low-arousal words observed in all conditions provide support for the arousal hypothesis according to which arousing information elicits a biological reaction facilitating its processing.

Current results have implications for models accounting for short-term recall performance. More specifically, it has been suggested that valence, arousal, and dominance are three major dimensions of semantics contributing to short-term memory performance (Ishiguro & Saito, 2021; Majerus & d’Argembeau, 2011). In this context, Majerus and d’Argembeau suggested that valence could affect similarity, because emotional words would share a positivity-negativity dimension. On the other hand, arousal could affect item accessibility with a lower activation threshold for high arousal words compared to low arousal words. Arousal could also increase the level of activation of lexical-semantic representations due to the involvement of the motivation systems involved in emotion processing. Under this view, because they contrasted positively and negatively valence words, Bireta et al. (2021) could not have uncovered an effect of valence; items in both conditions would have been equally similar. However, with a neutral condition, it should be possible to observe an effect of valence. Here, high-arousal words would have benefited from either a higher activation state or a lower activation threshold of their lexical-semantic representations.

Two main classes of hypotheses have been suggested to account for the impact of lexical-semantic factors on short-term memory performance. First, according to the redintegration hypothesis, short-term ordered recall is a two-step process (see, e.g., Hulme, Maughan, & Brown, 1991; Saint-Aubin & Poirier, 2000; Schweickert, 1993). Items would be maintained with phonological codes and at recall, phonological representations would be likely to be degraded. Degraded phonological traces would be reconstructed based on long-term lexical-phonological information. Lexical-semantic factors would influence the availability of long-term representations to support reconstruction. Therefore, with their higher level of activation or their lower threshold of activation, the long-term representations of high-arousal words would be more accessible to complete a degraded short-term phonological representation. This would increase their probability of recall as observed here with a strict and a free recall criterion.

The second class of hypotheses can be referred to as language-based models of short-term memory (e.g., Majerus, 2009; Martin, 2009; Poirier, Saint-Aubin, Mair, Tehan, & Tolan, 2015). Those models are heterogeneous, but they all suggest that semantic and lexical representations are closely related to short-term recall. Broadly speaking, these models suggest that processing linguistic information for recall involves activation of the relevant long-term networks including semantic, lexical and sublexical networks. Consequently, those networks would impact all aspects of performance in short-term memory tasks. In fact, the richer or more easily accessible representations within the language network would contribute to their temporary storage and produce a recall advantage. Majerus and D'Argembeau (2011) nicely illustrated these processes. Therefore, under this view, valence, arousal and dominance should all impact short-term recall performance, as observed here with arousal.

In sum, our results indicate that in short-term ordered recall, emotions are no exception: as a semantic dimension they influence recall performance. With well-controlled stimuli matched on several dimensions, there is a reliable effect of arousal on memory performance which is not influenced by presentation rate or list composition. It remains to be seen if dominance and valence—when a neutral condition is included in the design—have an effect to fully understand the role of emotion on short-term memory.

References

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. http://dx.doi.org/10.3758/BF03193014

Bireta, T. J., Guitard, D., Neath, I., Surprenant, A. M. (2021). Valence does not affect serial recall. *Canadian Journal of Experimental Psychology, 75*(1), 35-47. https://doi.org/10.1037/cep0000239

Bowen, H. J., Kark, S. M., & Kensinger, E. A. (2018). NEVER forget: Negative emotional valence enhances recapitulation. *Psychonomic Bulletin & Review, 25(3),* 870–891. https://doi.org/10.3758/s13423-017-1313-9

Bradley, M. M., Greenwald, M. K., Petry, M. C., Lang, P. J. (1992). Remembering pictures: Pleasure and arousal in memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*(2), 379-390. https://doi.org/10.1037/0278-7393.18.2.379

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*, 1520–1523. http://dx.doi.org/10.3758/s13428-016-0811-4

Brysbaert, M., & New, B. (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. http://dx.doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. http://dx.doi.org/10.3758/s13428-013-0403-5

Cahill, L., McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in Neurosciences*, *21*(1), 294-299. https://doi.org/10.1016/s0166-2236(97)01214-9

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *The Journal of Neuroscience*, *20*(19), RC99. https://doi.org/10.1523/JNEUROSCI.20-19-j0004.2000

Easterbrook, J. A. (1959). The Effect of Emotion on Cue Utilization and the Organization of Behavior. *Psychological Review*, *66*(3), 183-201. https://doi.org/10.1037/h0047707

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(1), 1149-1160. https://doi.org/10.3758/brm.41.4.1149

Fawcett, J. M., (2013). The production effect benefits performance in between-subject designs: A meta-analysis. *Acta Psychologica*, *142*(1), 1-5. http://dx.doi.org/10.1016/j.actpsy.2012.10.001

Guérard, K., & Saint-Aubin, J. (2012). Assessing the effect of lexical variables in backward recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38(2)*, 312–324. https://doi.org/10.1037/a0025481

Guitard, D., Saint-Aubin, J., Tehan, G., & Tolan, A. (2018). Does neighborhood size really cause the word length effect? *Memory & Cognition*, *46(2)*, 244–260. https://doi.org/10.3758/s13421-017-0761-9

Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word length, set size, and lexical factors: Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44(11)*, 1824–1844. https://doi.org/10.1037/xlm0000551

Hadley, C. B., MacKay, D. G. (2006). Does emotion help or hinder immediate memory? Arousal versus priority-binding mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*(1), 79-88. https://doi.org/10.1037/0278-7393.32.1.79

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30(6)*, 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518. https://doi.org/10.1016/S0749-596X(03)00096-2

Ishiguro, S., & Saito, S. (2021). The detrimental effect of semantic similarity in short-term memory tasks: A meta-regression approach. *Psychonomic Bulletin & Review*, *28(2)*, 384–408. https://doi.org/10.3758/s13423-020-01815-7

Kass, R. E., Raftery, A. E. (1995). Bayes factor. *Journal of the American Statistical Association 90*(430), 773-795. https://doi.org/10.1080/01621459.1995.10476572

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General, 143(3),* 1065–1081. https://doi.org/10.1037/a0035669

Labar, K. S., Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience, 7*(1), 54-64. https://doi.org/10.1038/nrn1825

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

Majerus, S., & D'Argembeau, A. (2011). Verbal short-term memory reflects the organization of long-term memory: Further evidence from short-term memory for emotional words. *Journal of Memory and Language, 64(2)*, 181–197. https://doi.org/10.1016/j.jml.2010.10.003

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations: The importance of distinguishing item and order information. In A. S. Thorn & M. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 244–276). Hove, UK: Psychology Press.

Martin, N. (2009). The roles of semantic and phonological processing in short-term memory and learning: Evidence from aphasia. In A. C. Thorn, M. A. Page (Eds.), *Interactions between short-term and long-term memory in the verbal domain* (pp. 220-243). NY: Psychology Press.

Medler, D. A., & Binder, J. R. (2005). *MCWord: An on-line orthographic database of the English language*. Madison, WI: Medical College of Wisconsin, Language Imaging Laboratory. Retrieved from www.neuro.mcw.edu/mcword/

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morey, R. D. & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Neath, I., Hockley, W. E., & Ensor, T. M. (2021). Stimulus-based mirror effects revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* https://doi.org/10.1037/xlm0000901

Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science, 17*(2), 147–152. https://doi.org/10.1111/j.1467-8721.2008.00565.x

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43(3)*, 489–499. https://doi.org/10.3758/s13421-014-0470-6

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(1), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, *7(2)*, 332–340. https://doi.org/10.3758/BF03212990

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language*, *118*(1). https://doi.org/10.1016/j.jml.2021.104219

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21(2)*, 168–175. https://doi.org/10.3758/BF03202729

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods*, *42*, 393–413. https://doi.org/10.3758/BRM.42.2.393

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9(1)*, 76–80. https://doi.org/10.1177/1745691613514755

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42(4)*, 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24-31. https://doi.org/10.1177/0098628316677643

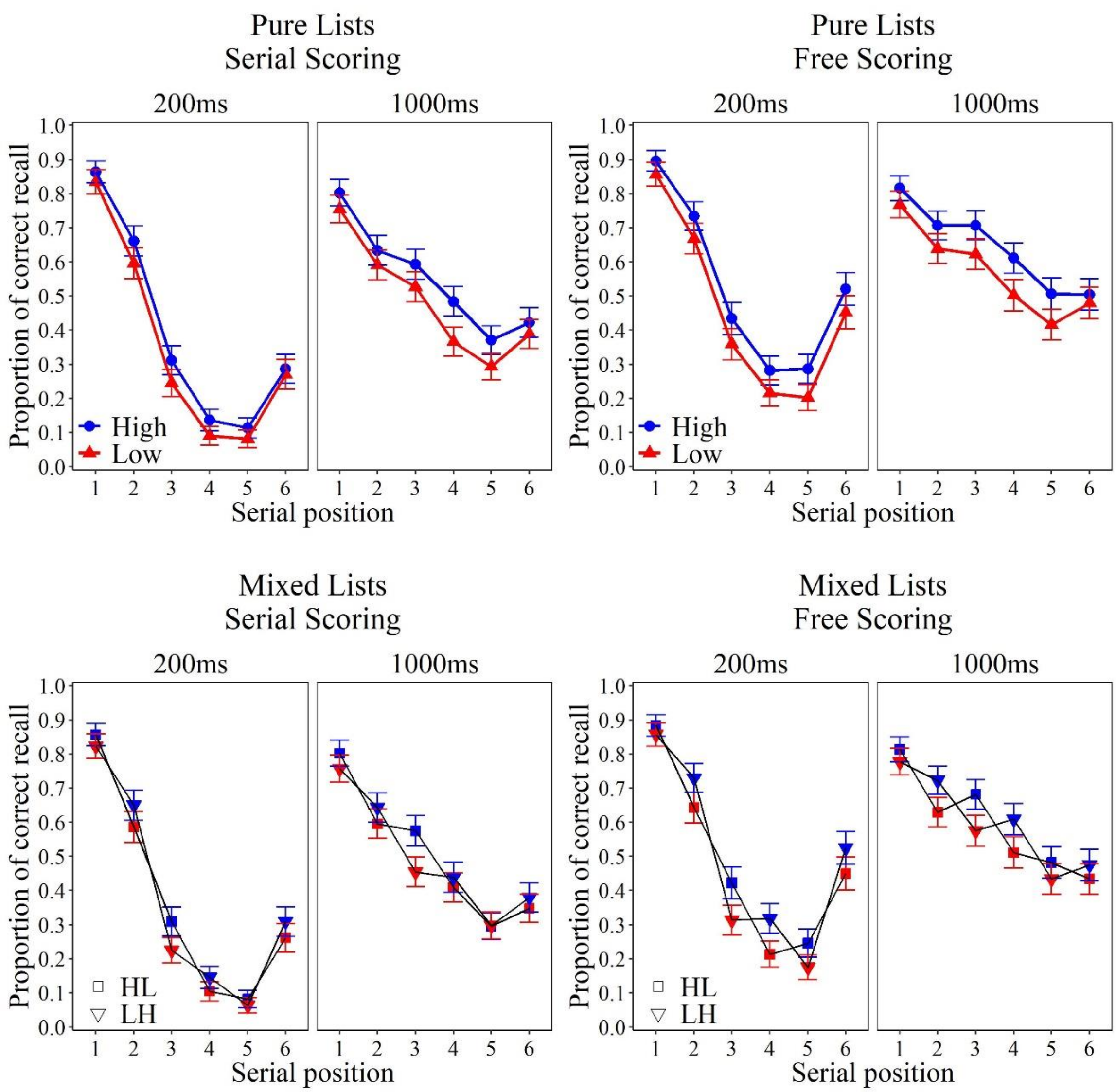
Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15(3)*, 535–542. https://doi.org/10.3758/PBR.15.3.535

Titchener, E.B. (1895). Affective memory, *The Philosophical Review*, *4*(1), 65-76. https://doi.org/10.2307/2175845

Warriner, A. B., Kuperman, V., Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45*(1), 1191 1207. https://doi.org/10.3758/s13428-012-0314-x

Figure 1

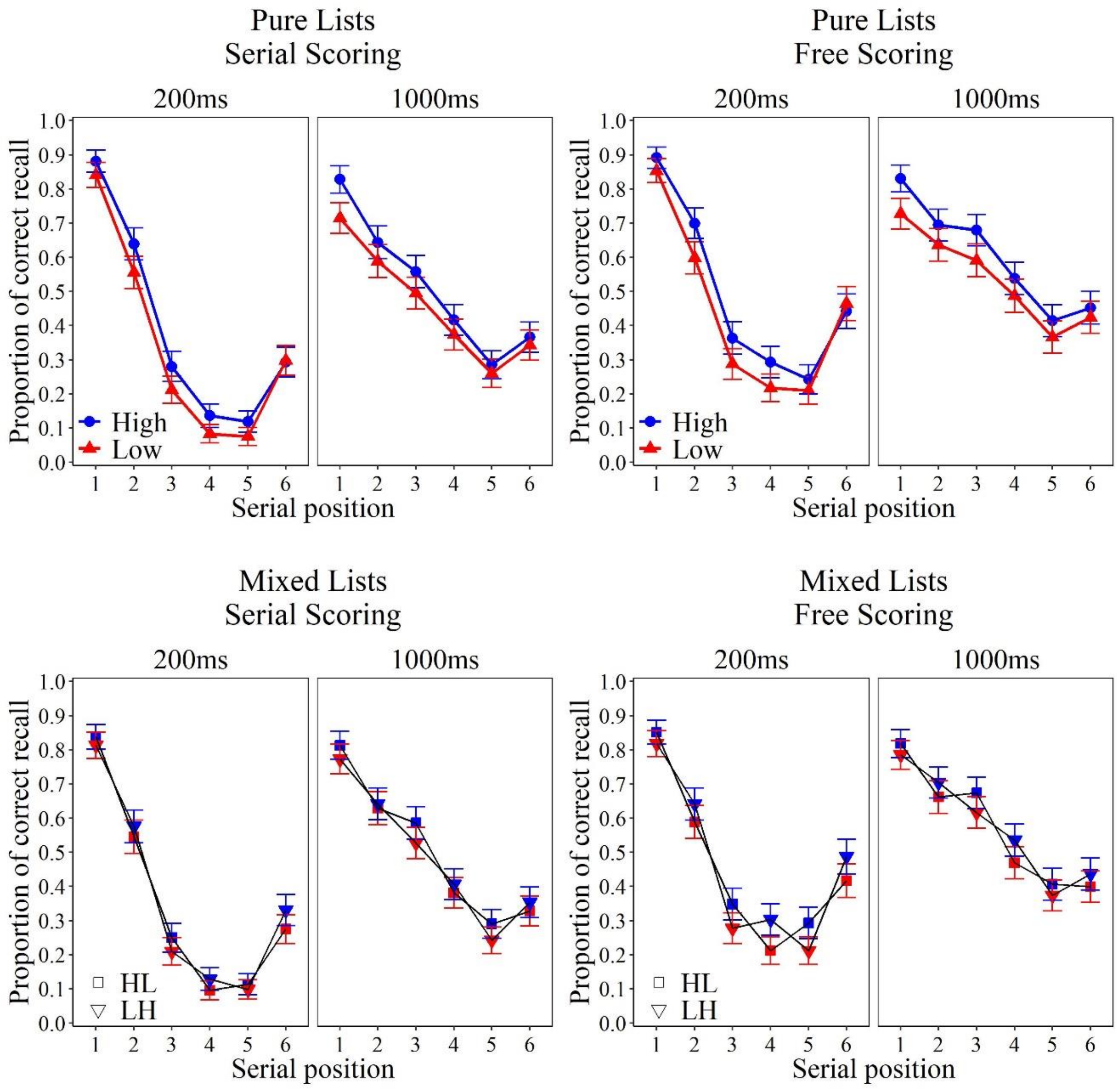
Proportion of words correctly recalled as a function of arousal, list composition, presentation rate, and serial position in Experiment 1 (Experiment 1a, Experiment 1b) with a strict serial recall scoring (left panels) and a free recall scoring (right panels).



Note. In all panels, high arousal words are in blue and low arousal words are in red. The letter combinations "HL" and "LH" indicate the composition of mixed lists; HL lists contain high arousal words at positions 1, 3 and 5, while LH lists contain high arousal words at positions 2, 4 and 6. The error bars represent 95% confidence intervals computed with Morey's (2008) method.

Figure 2

Proportion of words correctly recalled as a function of arousal, list composition, presentation rate, and serial position in Experiment 2 (Experiment 2a, Experiment 2b) with a strict serial recall scoring (left panels) and a free recall scoring (right panels).



Note. In all panels, high arousal words are in blue and low arousal words are in red. The letter combinations "HL" and "LH" indicate the composition of mixed lists; HL lists contain high arousal words at positions 1, 3 and 5, while LH lists contain high arousal words at positions 2, 4 and 6. The error bars represent 95% confidence intervals computed with Morey's (2008) method.

Appendix A

Stimuli used in Experiment 1.

	Low arousal words						High arousal words			
Dimension	*M*	*SD*	Min	Max	*t*	*p*	*M*	*SD*	*Min*	*Max*
LgWF	2.22	0.47	1.42	3.50	0.903	0.368	2.27	0.43	1.36	3.15
LgCD	2.04	0.43	1.30	3.26	0.485	0.628	2.06	0.41	1.28	2.98
NLet	6.04	1.30	3.00	8.00	0.402	0.688	6.11	1.26	3.00	8.00
NPhon	4.90	1.07	3.00	8.00	0.122	0.903	4.92	1.05	2.00	8.00
NSyll	1.80	0.40	1.00	2.00	0.494	0.622	1.83	0.38	1.00	2.00
LgHAL	7.54	1.48	3.69	11.29	0.956	0.340	7.71	1.28	3.85	11.47
OLD	2.23	0.55	1.25	3.80	0.076	0.940	2.23	0.55	1.00	3.40
OLDF	7.25	0.76	3.56	8.88	1.085	0.279	7.14	0.77	5.29	9.09
PLD	2.02	0.58	1.00	3.75	0.366	0.715	2.05	0.56	1.00	3.60
PLDF	7.23	0.92	4.52	9.34	0.345	0.731	7.19	1.01	4.40	11.14
CNC	4.19	0.75	1.82	4.93	1.094	0.275	4.08	0.78	2.30	5.00
SN	1,403.28	2,139.32	0.00	8,122.00	0.342	0.733	1,311.64	1,993.84	0.00	7,964.00
AOA	8.39	2.25	3.00	14.53	1.147	0.252	8.73	2.34	3.75	14.94
VM	5.28	0.69	2.40	6.61	1.153	0.250	5.12	1.31	2.71	8.14
AM	**2.97**	**0.34**	**2.15**	**3.50**	**46.857**	**< .001**	**5.71**	**0.54**	**5.05**	**7.57**
DM	5.07	0.38	3.84	6.05	0.596	0.552	5.02	0.82	3.57	6.96
CELEX	8.38	11.93	0.00	11.93	1.079	0.282	6.84	10.00	0.06	60.68
Orth	1.90	2.43	0.00	2.43	1.016	0.311	2.29	3.43	0.00	20.00
OrthF	15.71	41.56	0.00	41.56	0.837	0.403	11.71	31.52	0.00	201.08
C1_F	6,911.34	3,436.12	1,851.93	3,436.12	0.271	0.786	6,788.23	3,564.92	1,698.52	22,198.04
C1_C	606.29	330.50	17.67	330.50	0.333	0.739	620.42	323.62	51.67	1,855.43
C2_F	851.53	783.96	5.44	783.96	0.852	0.395	934.70	720.11	12.93	2,870.57
C2_C	85.88	86.55	3.00	86.55	0.416	0.678	90.32	77.55	3.67	385.17
C3_F	154.56	284.28	0.63	284.28	1.098	0.273	122.77	137.53	1.01	713.07
C3_C	16.41	30.26	1.00	30.26	1.010	0.313	13.35	13.21	1.00	83.33
U1_F	264,071.21	44,719.26	101,545.86	44,719.26	1.220	0.224	257,172.29	42,485.97	133,846.92	349,349.55
U1_C	34,374.14	5,527.55	16,640.00	5,527.55	0.440	0.660	3,4045.53	5,973.83	18,693.25	48,580.20
U2_F	20,563.91	10,862.66	2,666.91	10,862.66	0.136	0.892	20,375.68	10,559.50	1,968.45	68,562.63
U2_C	3,085.81	1,420.47	584.00	1,420.47	0.166	0.869	3,055.48	1,404.29	373.25	6,309.80
U3_F	1,884.36	2,492.44	2.11	2,492.44	0.230	0.818	1,963.30	2,788.05	5.31	21,613.08
U3_C	332.03	378.97	3.00	378.97	0.146	0.884	325.56	300.73	3.00	1,648.40

Note. LgWF = log frequency; LgCD = log contextual diversity (from Brysbaert & New, 2009); NLet = number of letters; NPhon = number of phonemes; NSyll = number of syllables; LgHAL = log HAL frequency; OLD = mean Levenshtein distance for 20 closest orthographic neighbours ; OLDF = frequency of the 20 closest orthographic neighbours; PLD = mean Levenshtein distance for the 20 closest phonological neighbours; PLDF = frequency of 20 closest phonological neighbours (from Balota et al., 2007); CNC = mean concreteness (from Brysbaert et al., 2014); SN = semantic neighbours (from Shaoul & Westbury, 2010); AOA = Age of acquisition (from Brysbaert & Biemiller, 2017); VM = mean valence; AM = mean arousal; DM = mean dominance (from Warriner et al., 2013); CELEX = Frequency; Orth = number of orthographic neighbours; OrthF = Frequency of orthographic neighbours; C1_F = Constrained unigram frequency; C1_C Constrained unigram count; C2_F = Constrained bigram frequency; C2_C = Constrained bigram count; C3_F = Constrained trigram frequency; C3_C = Constrained trigram count; U1_F = Unconstrained unigram frequency; U1_C = Unconstrained unigram count; U2_F = Unconstrained bigram frequency; U2_C = Unconstrained bigram count; U3_F = Unconstrained trigram frequency; U3_C = Unconstrained trigram count (from Medler & Binder, 2005).

Low arousal words = apron, asphalt, aspirin, axle, barnyard, baseline, bathroom, boiler, borough, breath, cadet, camel, canteen, cargo, carriage, carrier, catcher, cattle, cedar, ceiling, charcoal, choir, chute, clause, clinic, coaster, cobbler, column, cruiser, deuce, dialogue, dirt, doghouse, doormat, dozen, elk, eyebrow, fabric, femur, fixture, flannel, folk, format, freezer, furnace, gavel, giraffe, gospel, grid, guardian, Guinea, gum, hallway, heater, herring, holder, icebox, imprint, item, kilo, lingo, liver, locket, loft, lung, Madam, mantle, medium, midst, mister, moth, norm, novel, nugget, ointment, ordnance, oxide, palate, panel, pastime, pond, postage, prairie, primate, putter, quartet, rerun, reviewer, roller, rowboat, sadness, schooner, scope, sequel, shutdown, sinus, smudge, stairway, stool, suede, sultan, symbol, syrup, tenant, terrain, textbook, thermos, title, township, trailer, trench, truce, tube, unit, valley, vinyl, waitress, waiver, yarn, yearbook.

High arousal words = actress, album, applause, arcade, baboon, bazaar, beast, birth, boa, bomber, bonus, booty, bourbon, brothel, buffoon, buzzer, canon, circus, cleavage, cobra, cocktail, condom, conquest, corset, cougar, cowgirl, crossbow, cult, dagger, dealer, dessert, diver, drinker, duel, dungeon, eclipse, escort, eyeful, fighter, fireball, fitness, foursome, friction, fury, ghetto, giggle, gourmet, grenade, grinder, groin, heroine, hippo, holster, impact, impulse, jet, knockout, lair, leopard, lion, lizard, lover, mace, madman, Mafia, manhood, misfit, mistress, mobster, mustang, nudist, ogre, orgy, outburst, outrage, panther, panties, passion, phantom, pirate, pistol, playboy, porn, pursuit, quest, quickie, racer, rampage, rebel, robber, rocker, runner, sale, scar, shark, shotgun, siege, sinner, speedway, spouse, stadium, striker, stripper, sunshine, surfer, surgeon, swinger, throne, tiger, twister, vaccine, venom, viper, vodka, voltage, warrior, wasp, wealth, zombie, zoo.

Appendix B

Stimuli used in Experiment 2

Note. LgWF = log frequency; LgCD = log contextual diversity (from Brysbaert & New, 2009); NLet = number of letters; NPhon = number of

	Low arousal words						High arousal words			
Dimension	M	SD	Min	Max	t	p	M	SD	Min	Max
LgWF	2.55	0.55	1.63	4.28	0.827	0.409	2.61	0.63	1.36	4.61
LgCD	2.33	0.51	1.40	3.81	0.866	0.388	2.40	0.55	1.00	3.88
NLet	6.40	1.32	3.00	8.00	0.64	0.52	6.51	1.22	4.00	8.00
NPhon	5.32	1.46	2.00	9.00	0.678	0.499	5.45	1.33	2.00	8.00
NSyll	2.16	0.68	1.00	3.00	0.199	0.842	2.18	0.68	1.00	3.00
LgHAL	8.67	1.65	5.33	12.99	0.865	0.388	8.48	1.51	4.80	12.30
OLD	2.28	0.49	1.00	3.80	1.235	0.218	2.37	0.52	1.30	3.85
OLDF	7.21	0.70	5.14	8.94	0.480	0.632	7.17	0.65	5.29	8.82
PLD	2.19	0.60	1.00	3.65	1.186	0.237	2.30	0.68	1.00	3.80
PLDF	7.36	0.91	4.81	10.26	1.376	0.170	7.19	0.83	5.06	10.77
CNC	3.60	0.78	1.41	4.53	0.648	0.518	3.53	0.99	1.45	4.93
SN	3,070.04	3,147.11	0.00	9,110.00	0.23	0.82	2,975.72	2,908.49	0.00	9,478.00
AOA	9.01	2.15	3.80	14.75	0.671	0.503	8.80	2.39	3.81	16.20
VM	5.37	0.80	1.89	7.67	1.134	0.258	5.16	1.70	1.74	8.05
AM	**3.11**	**0.32**	**1.95**	**3.50**	**47.084**	**0.000**	**5.92**	**0.53**	**5.06**	**7.45**
DM	5.31	0.73	2.71	7.00	1.224	0.222	5.15	1.12	2.67	7.42
CELEX	29.15	47.72	0.12	283.19	0.128	0.898	30.34	83.72	0.42	741.22
Orth	1.35	1.75	0.00	8.00	0.343	0.732	1.44	2.18	0.00	12.00
OrthF	18.99	93.93	0.00	946.74	0.658	0.511	12.51	39.14	0.00	263.49
C1_F	6,411.25	3,340.24	2,195.15	30,320.75	0.81	0.42	6,074.01	2,760.30	2,342.13	17,297.52
C1_C	651.62	309.17	34.00	1,402.38	0.47	0.64	670.40	279.40	128.00	1,257.00
C2_F	869.31	599.41	18.26	3,440.26	0.52	0.60	916.45	724.78	80.32	4,403.59
C2_C	84.11	58.27	1.50	258.67	1.36	0.17	97.99	87.73	8.67	494.29
C3_F	172.93	144.32	2.68	899.25	0.77	0.44	194.95	256.75	3.00	1,901.19
C3_C	14.90	12.34	1.00	75.17	1.14	0.25	19.28	37.65	1.00	255.17
U1_F	271,715.69	49,637.33	124,870.56	412,076.00	1.01	0.32	265,104.52	46,401.67	98,257.74	382,045.26
U1_C	35,793.27	6,671.67	15,897.75	49,357.50	0.93	0.35	34,993.51	5,916.53	11,797.00	48,581.60
U2_F	21,789.66	8,326.08	694.76	47,437.92	0.85	0.40	22,846.94	9,876.94	357.25	68,055.67
U2_C	3,406.92	1,353.53	238.00	7,284.40	0.14	0.89	3,382.20	1,296.78	188.67	7,345.14
U3_F	2,372.63	1,734.42	18.09	7,630.98	1.16	0.25	2,095.88	1,768.83	16.42	8,738.11
U3_C	433.71	348.42	14.00	2,146.33	1.31	0.19	371.78	345.44	13.50	1,931.67

phonemes; NSyll = number of syllables; LgHAL = log HAL frequency; OLD = mean Levenshtein distance for 20 closest orthographic neighbours ; OLDF = frequency of the 20 closest orthographic neighbours; PLD = mean Levenshtein distance for the 20 closest phonological neighbours; PLDF = frequency of 20 closest phonological neighbours (from Balota et al., 2007); CNC = mean concreteness (from Brysbaert et al., 2014); SN = semantic neighbours (from Shaoul & Westbury, 2010); AOA = Age of acquisition (from Brysbaert & Biemiller, 2017); VM = mean valence; AM = mean arousal; DM = mean dominance (from Warriner et al., 2013); CELEX = Frequency; Orth = number of orthographic neighbours; OrthF = Frequency of orthographic neighbours; C1_F = Constrained unigram frequency; C1_C Constrained unigram count; C2_F = Constrained bigram frequency;

C2_C = Constrained bigram count; C3_F = Constrained trigram frequency; C3_C = Constrained trigram count; U1_F = Unconstrained unigram frequency; U1_C = Unconstrained unigram count; U2_F = Unconstrained bigram frequency; U2_C = Unconstrained bigram count; U3_F = Unconstrained trigram frequency; U3_C = Unconstrained trigram count (from Medler & Binder, 2005).

Low arousal words = acre, address, advice, agency, agent, aisle, alphabet, analyst, appendix, area, aspect, aunt, avenue, bedside, beginner, birch, booster, brig, brow, caller, canopy, capital, century, chord, citizen, client, closure, coma, company, concept, context, county, cousin, coverage, customer, decency, decision, depot, document, estate, estimate, footage, friar, gallon, gardener, granny, guidance, handicap, haven, history, holiness, horseman, hymn, length, liaison, memo, mileage, monsieur, opener, ore, overcoat, passage, pattern, pension, percent, policy, politics, pouch, prophet, protein, province, putty, ratio, record, recorder, refuge, regiment, region, republic, residue, reverend, rookie, saline, savings, scene, sequence, sermon, session, shrine, sitter, sky, sleeper, slogan, specimen, steward, summary, supplier, syllable, system, teller, temple, transfer, twain, version, window, womb, wrapper, yoga.

High arousal words = action, admirer, airport, burglary, buzzard, caffeine, cannon, carnival, casino, casualty, chaos, cinema, coffee, combat, comedy, concert, crisis, crusade, defiance, democrat, dollar, doomsday, dynamite, ecstasy, elite, energy, erection, explorer, exposure, fantasy, fiesta, football, fortune, galaxy, gambler, gangster, genius, ghost, glory, gunfire, gunner, hero, horror, incident, increase, inferno, invasion, jazz, jury, killer, laser, legend, life, lottery, lust, madness, martini, mastery, medal, mission, money, morphine, murderer, music, musician, nerve, nipple, noise, opponent, orgasm, overtime, penny, phobia, plasma, player, ranch, reptile, revenge, revolver, riches, robbery, romp, science, scorpion, sibling, sterling, strength, stud, suspense, sword, terror, thigh, threat, torment, tornado, tragedy, turbine, uprising, urgency, vendetta, viceroy, victory, warfare, weapon, weekend, winner, wrath, zeal.