# Emotional Arousal Does Not Modulate Stimulus-Response Binding and Retrieval Effects

Carina G. Giesen[1] and Andreas B. Eder[2]

[1]Friedrich Schiller University Jena, Jena, Germany

[2]Julius-Maximilians-University Würzburg, Würzburg, Germany

Cognition & Emotion, in press (2022)

Author Note

Carina C. Giesen, Friedrich-Schiller-Universität Jena, and Andreas B. Eder, Julius-Maximilians-University Würzburg, Germany

Correspondence

Correspondence concerning this article should be addressed to Carina G. Giesen, Friedrich Schiller University Jena, Department of Psychology, General Psychology II, Am Steiger 3, Haus 1, 07743 Jena, Germany. Email: carina.giesen@uni-jena.de.

## Abstract

The adaptation-by-binding account and the arousal-biased competition model suggest that emotional arousal increases binding effects for transient links between stimuli and responses. Two highly-powered, pre-registered experiments tested whether transient stimulus-response bindings are stronger for high versus low arousing stimuli. Emotional words were presented in a sequential prime-probe design in which stimulus relation, response relation, and stimulus arousal were orthogonally manipulated. In Experiment 1 ($N$ = 101), words with high and low arousal levels were presented individually in prime and probe displays. In Experiment 2 ($N$ = 170), a high arousing affective word was presented simultaneously with a neutral word during the prime display; in the subsequent probe display, either the arousing or the neutral word repeated or a different high versus low arousal word was shown. Data from both experiments did not demonstrate a modulation of SRBR effects by stimulus arousal and SRBR effects were of equal magnitude for word stimuli of high and low arousal levels. These null results are not in line with binding accounts that hypothesize a modulatory influence of emotional arousal on perception-action binding.


*Keywords*: stimulus-response binding; event files; stimulus arousal, episodic retrieval, action control

Research findings from the past two decades attest to the ubiquity and pervasiveness of transient stimulus-response binding and retrieval (SRBR) effects (for an overview, see Frings et al., 2020). Carrying out a response in close temporal proximity to a perceived stimulus will transiently link mental representations of both, the stimulus and the executed response, in episodic memory (termed stimulus-response binding or event file, cf. Hommel, 1998; Hommel et al., 2001). After binding, stimulus repetition can retrieve the entire event file from memory, which will facilitate response selection in situations in which the retrieved response is appropriate (resulting in performance benefits when the stimulus and response are repeated from prime to probe), but which will hamper response execution if the retrieved response is not appropriate (reflected in performance costs when the stimulus and/or the response change from prime from to probe). Statistically, retrieval of stimulus-response episodes is reflected in a disordinal interaction between stimulus relation and response relation from trial n-1 to trial n (for recent reviews see Frings et al., 2020; Horner et al., 2014). Processes of stimulus-response binding and retrieval allow for an efficient way of stimulus-based action regulation (Logan, 1988) and may be conceived of as the cognitive basis of more enduring learning phenomena like contingency learning or habit formation (e.g., Giesen & Rothermund, 2015; Giesen et al., 2020; Schmidt et al., 2020; see Moeller & Frings, 2017, for a detailed discussion).

The present study investigated whether SRBR effects will differ for stimuli of arousing versus neutral (i.e., non-arousing) stimuli. This research question was inspired by two influential neuropsychological models on a modulation of binding processes by noradrenergic arousal signals.

According to the adaptation-by-binding model by Verguts and Notebaert (2008, 2009; see also Abrahamse et al., 2016), high arousal states involve a catecholamine boost (noradrenaline or dopamine) across cortex that facilitates ongoing Hebbian learning and thus the binding of currently active perceptual and motor representations. The account was originally proposed to explain conflict adaptation and congruency sequence effects in stimulus-response conflict tasks. Specifically, it was proposed that conflict detection creates an arousal signal in a neuromodulatory locus coeruleus-norepinephrine system (LC-NE; Aston-Jones & Cohen, 2005) that strengthens all connections between representations active at the time the arousal signal arrives, even if they are irrelevant to the task. In trial sequences, the strengthened episodic binding between stimulus and response features in the foregoing trial should facilitate performance in a subsequent trial if all or no features of the episode are repeated relative to a partial (mis)match, which was observed in several experiments (Mayr et al., 2003; Hommel et al., 2004; Verguts & Notebaert 2008). Note that according to Verguts and Notebaert (2009), the arousal signal is not specific to conflict detection and can be also triggered by other stimulations of the neuromodulatory arousal system. Therefore, high arousing, emotional stimuli should analogously boost binding between stimulus and response representations according to this account.

A similar prediction is made by the arousal-biased competition (ABC) model (Mather & Sutherland, 2011; Mather et al., 2016). According to this model, emotional arousal (elicited by external stimuli, internal thoughts, or stress hormones) enhances perception and memory of high-priority information, at the cost of other information, in a winner-takes-more fashion. The model assumes that stimuli must compete for limited mental resources and that both, bottom-up and top-down mechanisms help resolve competition. Numerous studies showed that emotionally relevant stimuli attract more attention and are recalled better than non-arousing, neutral stimuli (e.g., Anderson, 2005; Sheth & Pham, 2008; for an overview, see Mather, 2007). When several stimuli compete for representation, arousal elicited by the same or by another stimulus increases the competitive advantage of the emotional stimuli over lower priority (neutral) stimuli (for evidence see Lee et al., 2014). The arousal-enhanced attention bias for emotionally relevant stimuli improves memorization of these events, which also explains why the gist information of emotional events is typically better recalled relative to mundane events (for a review see Kensinger, 2009). MacKay and colleagues (2004; Hadley & MacKay, 2006) argued in a discussion of so-called 'flashbulb memories' that emotional arousal improves binding of stimuli to contextual features in episodic memory. Other studies, in contrast, argued that high stress (negative emotional arousal) will interfere with episodic binding to the context (Bisby et al., 2018; Payne et al., 2004). The ABC model can reconciliate both perspectives with the assumption that emotional arousal will only increase binding to context features if these features are salient and become a target of focused attention during encoding (Mather, 2007).

Theories on binding in perception, memory, and action control hence agree that (negative) emotional arousal should affect binding and retrieval processes. This assumption is also in line with the *Binding and Retrieval in Action Control* (BRAC) framework (Frings et al.,2020). According to BRAC, processes of stimulus-response binding and retrieval are independently modulated by top-down as well as bottom-up influences. Whereas top-down influences reflect modulations due to expectations, goals, or instructions that are mediated via attentional weighting (Memelink & Hommel, 2013), bottom-up influences in this framework reflect experience-based modulations of binding and retrieval processes due to differences in stimulus perception. For instance, Schmalbrock and colleagues (2021) reported stronger SRBR effects for salient than nonsalient target stimuli. Giesen and Rothemund (2011) observed stronger SRBR effects for affectively matching (vs. non-matching) stimuli. In similar fashion, the expected modulatory influence of (negative) emotional arousal of SRBR effects would reflect a bottom-up modulation of SRBR effects that is in line with BRAC and the ABC model.

To investigate whether emotional arousal modulates SRBR effects, we conducted two studies: We used a sequential prime-probe design for each study and presented individual words in each display (Experiment 1) or created stimulus competition by presenting an arousing and neutral word simultaneously above each other during prime displays (Experiment 2). Figure 1 shows the basic design of the experiments (for details see the respective *Method* sections). Experiment 1 was designed to examine an arousal influence on stimulus-response binding and retrieval processes in line with prediction of the adaptation-by-binding model. In turn, the design of Experiment 2 with presentations of multiple stimuli was optimized for a test of the ABC model. In both experiments, affective words were presented as distractor items that either were high in arousal or affectively neutral. Only negative affective words were selected because previous research suggested that the arousal influence on binding could be specific for negative emotion (Kensinger, 2007, 2009). High-arousing words were selected based on extensive normative ratings (in Experiment 1) and, additionally, based on individual ratings (in Experiment 2). Previous studies demonstrated that the processing of high arousing negative words elicits autonomic reactions in line with activation of a neuromodulatory arousal system (e.g., Herbert et al., 2006; McGinnies, 1949; Kissler et al., 2007; Lewis et al., 2007; Weis & Herbert, 2017). In one example study, participants reacted with increased skin conductance levels (indexing an autonomic arousal reaction) after subliminal presentations of aversive words rated high in arousal (Silvert et al., 2004; for a review see van der Ploeg et al., 2017). Therefore, we assumed that the minimal conditions for the generation of arousal signals were met in the present studies.

## Experiment 1

Experiment 1 presented emotional words as distractor items that could repeat or change from prime to probe displays in a color classification task. The words were the carriers of the task-relevant color information and thus in the focus of the participant's attention. Note though that word meaning was irrelevant for the color classification task. Importantly, emotional words were either high in arousal or affectively neutral. Based on the adaption-by-binding model, we hypothesized larger SRBR effects following presentations of high arousing words relative to neutral words. Statistically, this is expressed in a three-way interaction in reaction times to probe displays between word relation (repetition *vs* change), response relation (repetition *vs* change), and stimulus arousal level (high *vs* low). Reaction times to probe displays were statistically analyzed using a linear mixed model approach (LMM) that included word arousal values as metric predictor variable in the model.

**Method**

**Ethics vote, pre-registration, and open access.** Study procedures were approved by the local ethics committee of the Department of Psychology, JMU Würzburg (ref. no. GZ 2018-22). Prior to data collection, the exact method, design, hypotheses, data preparation, and planned analyses were pre-registered online at the Open Science Framework (OSF; pre-registrations: Exp 1: https://osf.io/khv29; Exp. 2: https://osf.io/vy2kq). All materials, data, and analyses are available online at the Open Science Foundation (https://osf.io/4bgxc/).

**Sample size and a-priori power calculations.** According to a pre-registered power analysis calculated with G*Power3 (Version 3.1.9.2, Faul et al., 2007), a minimum sample size of $N = 97$ is needed to detect a medium-sized effect of $dz = 0.4$ with a statistical power of (1-ß) = 0.95 in two-tailed dependent-samples *t*-tests with the alpha-level adjusted to 0.025 according to Bonferroni. Note that the dependent-samples *t*-tests are equivalent to the planned contrasts of the two-way and three-way ANOVA interaction tests specified in the preregistration of Experiment 1 with $t^2=F$ and df=1. In the absence of an informed effect size estimate, Brysbaert (2019) suggested *d* = 0.4 as a good first estimate for most psychological studies. Furthermore, SRBR effects in previous studies were typically larger than *dz* = 0.4 (e.g., Giesen & Rothermund, 2011, 2015, 2016). Therefore, we planned our studies with *dz* = 0.4 as the smallest effect size of interest.

**Participants.** Participants were recruited online via the internet platform Prolific Academic. A total of 103 participants completed the experiment. Two participants produced excessive error rates (> 25% errors) and were excluded in line with our pre-registered criteria for data exclusions. The final sample was *n*=101 participants (41 females, 60 male). Participants were native speakers of German language, between 18 and 35 years old (*M*=25.5 years), and they had a Prolific approval rate of at least 65-100% in prior studies.

The experiment was run on the participant's personal desktop or notebook computer. Experiment duration was about 15 minutes. Participants received £1.88 (approximately €2.10) as monetary compensation for participation. All participants explicitly provided informed consent to taking part in the study.

**Design.** The experiment had a 2×2×2 sequential prime-probe design with the within-subject factors *Word Relation* (50% identical repetition [ID] vs. 50% word change [half of these word change sequences presented different words of the same arousal category [SC] and the other half presented words of different arousal categories [DC]]) × *Response Relation* (response repetition [RR], 50% vs. response change [RC], 50%) × *Stimulus Arousal Level* (high, 50% vs. low, 50%). The dependent variable of main interest was the probe reaction time (RT). For exploratory purposes, we also analyzed the error rates (for a report see the supplementary information file).

**Materials**. Ten German words of high arousal (all negative) and ten neutral words were selected from The Berlin Affective World List Reloaded database (BAWL-R; Võ et al., 2009) on the basis of normative ratings of emotional arousal (see Table 1). The high and low-arousing word sets were matched for word length and frequency of appearance per million words (see the word list in the supplement, Table 1). Mean arousal ratings (rated on a 5-point scale ranging from 1= *low* to 5= *high*) differed significantly between high ($M$=4.6, $SD$=0.08) and low ($M$=1.32, $SD$=0.09) arousal words, $t(9)=48.41$, $p<.001$, $d_z=15.3$. High arousing words ($M$=-2.35, $SD$=0.51) were also rated more negatively than low arousing words ($M$=1.26, $SD$=0.65), $t(9)=15.53$, $p<.001$, $d_z=4.9$. For task practice, an additional set of ten neutral German adjectives (e.g., quiet, small, edgy) was selected.

**Procedure.** The experiment was programmed with PsychoPy (version 3). Participants first answered demographic questions about their gender, age, native language, and dominant hand. Instructions for the color classification task stated that a white word will appear in each trial that will change color to yellow or blue. Task instruction was to react to a yellow word with a press of the left key (D) and to a blue word with a press the right key (L) as quickly and as accurately as possible. After a brief instruction check with guided questions, a practice block followed with 32 prime-probe sequences. The practice block was repeated if the participant committed many errors (>20%) and/or slowly in many trials (>20% RTs slower than 1000 ms). The practice block was repeated two times when necessary; if the performance criteria were not met after three practice blocks, participation was terminated.

Following the practice block, the main block started which comprised 160 prime-probe sequences. Prime-probe sequences were constructed with respect to the 2x2x2 factorial design. Half of all prime words were presented in yellow, the other half were presented in blue. Note that probe color varied as a result of the Prime Color × Response Relation factorial combinations. Each of the 20 words was presented 2 times in prime-probe sequences with identical repetition and 2 times in sequences with word change (1 presentation for same and different arousal category each) for every level of the response relation factor, resulting in 8 prime and 8 probe presentations overall.

The events in a prime-probe sequence were the following: Displays had black background and white font as default. The prime trial started with a fixation cross ('+') for 250 ms, followed by the prime word in white font on a black background for a variable duration (150-350 ms). Then, the prime word changed its color to yellow or blue. The colored word was displayed until registration of a keypress or for a maximum of 1500 ms (response omission). Erroneous prime responses produced an error feedback "Error! Continue with correct key press" (presented in red font until correct keypress); slow (RT> 1000ms) oromitted responses generated the feedback message "Too slow! Please respond faster! Continue with correct key press" (presented in gray font until correct keypress). Immediately after registration of the prime reaction, a fixation cross appeared again for 250 ms, initiating, the probe trial. The procedure of the probe trial was identical with that of the prime trial. The prime-probe trial sequence ended with a blank black screen (500 ms). Then, the next prime-probe sequence started.

## Results

**Data preparation.** In line with our preregistered criteria, prime-probe sequences with an incorrect prime and/or probe response were removed for RT analyses (8.1% of the data). In addition, probe RTs faster than 200 ms or slower than 1.5 interquartile ranges above the third quartile of the participant's individual RT distribution ("outliers" according to Tukey, 1977) were excluded (3.5% of the data).

RT data were analyzed with a linear mixed-effects model (LMM), using trials as units of analysis[1]. Analyses were performed with R (packages ‘lme4’ and ‘lmerTest’). Participants were treated as level-2 predictor (i.e., trials were nested within participants) to account for statistical dependence between trial-based performance. In detail, we computed a random intercept model and included trial RT as dependent variable. We added fixed effects for *Word Relation* (contrast coded with identical repetition =1, word change/same arousal category=-1; the word change/different arousal category level was not included in the analyses), *Response Relation* (response repetition=1, response change=-1) and *Arousal Rating* (mean score ± SD) for each probe stimulus (metric predictor, grand mean centered) and their interactions. Last, we included participants as a random effect. The model specification was as follows: ProbeRT~1+wordRelation * ResponseRelation * AdjustedArousalRating + (1 | Subject).

**Probe RT performance.** The analysis yielded significant effects of word relation, $b$=-3.04, $t$(14214.2007)=5.92, $p$<.001, response relation, $b$=-18.67, $t$(14214.2007)=36.44, $p$<.001, and for their interaction, $b$=-6.40, $t$(14214.2007)=12.49, $p$<.001, reflecting the typical pattern of SRBR effects[2]. Importantly, however, this interaction was not qualified by arousal ratings, as the three-

---

[1] The LMM analysis was only preregistered for Experiment 2. However, we decided to use the same data-analytic approach for both studies given the methodological strengths of LMM and for a better comparability of results across experiments. Preregistered analyses of Experiment 1 using an ANOVA approach are reported in the Supplement.

[2]A 2 (arousal level) × 2 (word relation: identical repetition vs. word change/same arousal category) × 2 (response relation) ANOVA yielded a significant main effect of Response Relation, $F(1, 100) = 251.59$ $p < .001$, $\eta_p^2 = .72$, and a significant Word Relation × Response Relation interaction effect, $F(1, 100) =77.94$, $p < .001$, $\eta_p^{2}$= .44. Follow-up analyses on the Word Relation x Response Relation interaction revealed a disordinal interaction effect pattern typical for the retrieval of stimulus-response bindings: Word repetition (compared with word changes) produced a significant performance benefit for response repetition probes of about $\Delta$=18 ms, t(100)=9.56, p<.001, $d_z$=0.96. However, for response change probes, word repetition (compared with word change) produced a significant performance cost of about $\Delta$= -7 ms, t(100)=4.04, p<.001, $d_z$=0.40 (see also the ANOVA results of Experiment 1 in the supplement).

way interaction was not significant, $b$=-0.51, $t$(14214.2007)=1.06, $p$=.286. The pattern of results shows that SRBR effects were of equal magnitude for probe words of high and low arousal level (Fig. 2a). No other effect was significant.

To follow up on the nonsignificant three-way interaction, we computed Bayes Factors for a modulation of SRBR effects by stimulus arousal using the JASP software package (2022, Version 0.16.2). In a first computational step, we computed interaction effects for the three-way interaction (IA= $SRBR_{high\ arousal} - SRBR_{low\ arousal}$ = $[(SCRR\text{-}IDRR)\text{-}(SCRC\text{-}IDRC)]_{high\ arousal}$ - $[(SCRR\text{-}IDRR)\text{-}(SCRC\text{-}IDRC)]_{low\ arousal}$; see Design for explanation of abbreviations). Then, we submitted interaction effect scores to a Bayesian one sample $t$-test against zero, which yielded a $BF_{01}$ =8.749, indicating that the null hypothesis (nonexistence of the three-way interaction) was nearly nine times more likely than the alternative hypothesis. This provides moderate evidence for the validity of the null hypothesis (van Doorn et al., 2021).

## Discussion

Experiment 1 produced strong SRBR effects. Importantly however, SRBR effects did not differ in magnitude between arousing and non-arousing stimuli. These findings argue against the hypothesis derived from the adaptation-by-binding model that stimulus-induced arousal signals increase stimulus-response binding effects. The results are also at odds with the ABC model, which assumes that emotional stimuli increase binding to high-priority (emotional) stimuli. For a fair test of this model, however, stimuli must compete for representation. This was not the case in Experiment 1, in which we used presentations of a single stimulus item in prime and probe. For Experiment 2, we therefore included multiple stimuli in the prime display, creating competition between arousing and non-arousing stimuli.

## Experiment 2

In Experiment 2, two different words were shown simultaneously in the prime display on the screen (Figure 1b). Both words had the same color, which was the relevant feature for the colour categorization task. One of the two words was emotionally arousing, whereas the other word was non-arousing (neutral). According to the ABC model, the emotionally arousing word should become prioritized in competition for mental representation at the expense of the non-arousing word. This arousal-induced processing advantage should generate stronger binding between arousing distractor words and categorization responses. For a comparison of SRBR effects, either the arousing word or the non-arousing word was repeated from prime to probe (identical repetition) or a different arousing or non-arousing word was presented in the probe (word change). In the probe display, the selected probe word was presented twice to make the display similar to the prime (Figure 1b). In line with the ABC model, we expected to find stronger SRBR effects with arousing words, which is statistically expressed in a three-way interaction between word relation, response relation, and arousal levels. In addition to these changes, each participant provided emotional valence and arousal ratings of the words presented in Experiment 2. These subjective ratings were entered as metric predictor variables in the LMM analysis.

### Method

**Sample Size Planning.** As described in the preregistration document (see https://osf.io/vy2kq), sample size was planned with a more conservative effect size estimate of d*z*=0.3. According to a-priori power calculated with G*Power3, a total of n = 171 was needed to have good statistical power (1-ß = 0.95) for the detection of dz = 0.3 in two-tailed dependent-samples t-tests with the alpha-level adjusted to 0.025 according to Bonferroni.

**Participants.** Participants were recruited online via Prolific. In total, *N*=171 participants completed the experiment. One person participated twice and the second data set was therefore excluded, leaving *n*=170 participants (79 female, 90 males, 1 gender not reported, $M_{age}$=30.12 years, *SD* = 10.49). All participants were pre-screened as native German speakers with an approval rate of at least 65-100% in prior studies (we meant to pre-screen for age range, too, but that filter was not saved prior to data collection). Experiment duration and payment was equal to Experiment 1. Participants explicitly provided informed consent to participation.

**Design.** Experiment 2 comprised a 2×2×2 sequential prime-probe design with the within-subject factors *Word Relation* (identical repetition, 50% vs. word change, 50%) × *Response Relation* (response repetition, 50% vs. response change, 50%) × *Arousal Level* (high, 50% vs. low, 50%).

**Materials and Procedure.** Materials and procedure were identical to Experiment 1 except for the following changes: Eight high arousing word and eight neutral words selected from the BAWL-R data base were presented as distractor items (for a list, see Table 2). Before the main experimental task, participants were asked to rate the feelings elicited by each word ("*How do you feel when reading this word?*") on 9-point self-assessment manikin scales (Bradley & Lang, 1994) for arousal (1= very calm, 9= very aroused) and pleasantness (1= very unpleasant; 9= very pleasant). The main experimental task comprised 160 prime-probe sequences that were constructed with respect to the factorial design. Two words were simultaneously presented in the prime trial – a high arousal word and a low arousal word – that were presented centrally, one word slightly above and the other word slightly below the screen center. The position of the high arousal word in the prime display was balanced (50% top, 50% bottom). In the probe display, the same word was shown on both positions (see Figure 1b). Probe word identity depended on the factorial combinations of the Arousal Level × Word Relation factors: In trial sequences with word repetition from prime to probe, the high arousal prime word was repeated in 50% of the trials; in the other half, the low arousal prime word was displayed in the probe. In trial sequences with a word change from prime to probe, a high arousal probe word appeared in the probe that was not shown in the preceding prime display; for the other half, a different low arousal probe word was shown in the probe (see Fig. 1b). Words were first presented in white font and then the color of both words changed to either blue or yellow following the task procedures as described for Experiment 1. Half of the prime-probe sequences required response repetition from prime to probe (e.g., yellow-yellow; blue-blue) whereas the remaining prime-probe sequences required a response change from prime to probe (e.g., yellow-blue; blue-yellow). Prime color was balanced (50% yellow, 50% blue).

**Results**

**Word ratings.** Mean arousal and pleasantness ratings of each word were compared in a dependent *t*-test. Arousal ratings were higher for (BAWL-R pre-selected) high ($M$=7.07) than low ($M$=2.64) arousal words, $t(7)=19.50$, $p<.001$, $d_z=6.89$. Furthermore, feelings were rated more unpleasant during presentations of high arousing ($M$=2.20) relative to low arousing words ($M$=7.04), $t(7)=13.96$, $p<.001$, $d_z=-4.93$.

**Data preparation.** As preregistered, erroneous prime and/or probe responses (7.9%) and RT outlier values (3.7%) were discarded from RT analyses. Data analyses followed the same procedure as for Experiment 1 with the only difference that subjective arousal ratings of each participant (centered within person) were entered as metric predictor in the model. The model specification was as follows: ProbeRT~1+WordRelation * ResponseRelation * ArousalRating + (1|Subject).

**Probe RT performance.** The analysis yielded significant effects of response relation, $b$=-20.76, $t$(23,940)=13.66, $p$<.001, and for the Word Relation by Response Relation interaction, $b$=-3.27, $t$(23,940)=8.25, $p$<.001, reflecting the typical pattern of SRBR effects[3]. The three-way interaction was not significant, $b$=0.09, $t$(23,940)=0.64, $p$=.525, indicating that SRBR effects were of equal magnitude for probe words with high and low arousal levels (see Figure 2b). No other effect was significant.

Using the calculation method described in Experiment 1, we computed Bayes Factors for the modulation of SRBR effects by stimulus arousal level. This analysis yielded a $BF_{01}$ = 10.320, indicating that the null hypothesis (nonexistence of the three-way interaction) was about ten times more likely than the alternative hypothesis. This Bayes factor can be regarded as strong evidence for the validity of the null hypothesis (van Doorn et al., 2021).

## Discussion

In Experiment 2, high arousing and low arousing stimuli were presented simultaneously in the prime display. Based on the ABC model (Mather, 2007), we hypothesized a relative processing advantage for the high (compared with low) arousing words, which should promote binding processes for high arousing stimuli. To test this hypothesis, probe RTs were analyzed as a function of subjective ratings of the stimulus arousal level collected from participants in a linear mixed model with word relation and response relation as fixed effects. This data-analytic approach

[3] A 2 (arousal level) × 2 (word relation) × 2 (response relation) ANOVA yielded a significant main effect of Response Relation, $F(1, 169) = 325.04$, $p < .001$, $\eta_p^2 = .66$, and a significant Word Relation × Response Relation interaction effect, $F(1, 169) = 52.64$, $p < .001$, $\eta_p^2 = .24$. Follow-up tests indicated that word repetition (compared with word changes) produced a significant performance benefit for response repetition probes of about $\Delta$=7 ms, t(169) = 5.85, p < .001, $d_z$ = 0.45. However, for response change probes, word repetition (compared with word change) produced a significant performance cost of about $\Delta$= -6 ms, t(169) = 5.07, p < .001, $d_z$ = 0.39.

revealed a large SRBR effect, demonstrating robust binding processes with a complex stimulus display. Importantly, the magnitude of SRBR effects was not affected by the arousal value of the stimuli, which is in line with the result of Experiment 1 and at odds with predictions from the ABC model.

## General Discussion

In two highly-powered, preregistered experiments, we tested whether transient stimulus-response bindings are stronger for high versus low arousing word stimuli. This research question was inspired by two influential neuropsychological models—the adaptation-by-binding account (Verguts & Notebaert, 2008, 2009) and the ABC model (Mather & Sutherland, 2011) that proposed a modulation of binding processes by noradrenergic arousal signals. Both experiments investigated a modulation of stimulus-response binding processes with a sequential prime-probe design, in which stimulus relation and response relation from prime to probe trial were independently manipulated for high and low arousing word stimuli. Results of Experiment 1, with single word presentations in prime and probe trials, yielded robust SRBR effects. Crucially, and in contradiction to the hypotheses, these effects did not differ as a function of stimulus arousal level, meaning that SRBR effects were of equal magnitude for high and low arousal stimuli. As such, these findings oppose the adaptation-by-binding model. For an additional test of the ABC account, Experiment 2 created competition between stimuli by simultaneous presentations of high and low arousal words during prime trials. Results of Experiment 2 demonstrated strong and robust SRBR effects that did not differ between high and low arousing stimuli. These findings are in line with the results of Experiment 1 but at odds with the ABC model. In sum, findings from both studies argue against a modulatory influence of stimulus arousal level on transient stimulus-response binding and retrieval processes.

Note that both studies yielded robust evidence for SRBR effects. Follow-up tests on these effects (see Footnotes 2 and 3 for details) demonstrate disordinal interaction effects for the Word Relation by Response Relation interaction. These results are consistent with the typical pattern of advantages for full repetitions/full alternations and costs for partial repetitions that typically occur for bindings between task-relevant stimuli and responses (e.g., Hommel, 1998, 2007). For distractor-based binding effects, one often obtains an additional main effect of response relation that slightly tilts the disordinal interaction (resulting in a pattern where full stimulus- and response repetitions go along with faster performance than stimulus change and response changes). This tilted interaction is a common finding in studies on distractor-response bindings (e.g., Frings & Rothermund, 2011, 2017; Frings et al., 2007; Giesen & Rothermund, 2011, 2015, 2016; Giesen et al., 2012; Janczyk et al., 2022; Moeller & Frings, 2011, 2014; Moeller et al., 2016). Overall, the pattern of stimulus-based binding and retrieval effects is therefore consistent with previous findings from studies that investigate bindings for irrelevant word stimuli (i.e., distractors) and responses.

**Limitations**

It should be highlighted that this conclusion is based on the interpretation of null findings, that is, the absence of a modulatory effect of stimulus arousal level on SRBR effects. Given that both experiments were highly powered, we view it unlikely that the absence of a modulation was merely a consequence of insufficient statistical power (see the *a-priori* power calculations in the Method sections). This reasoning is further supported by our Bayes Factor Analyses, which yielded moderate (Exp. 1) to strong (Exp. 2) evidence for the validity of the null hypothesis (i.e., no difference in SRBR effects for high vs. low arousing stimuli).

In the present research, high arousal was confounded with emotional valence, because the high arousing words were also more negative than the low arousing words. It is possible that arousal signals associated with positive affect have a modulatory effect on SRBR effects (for suggestive evidence see e.g., Colzato et al., 2007; Eder et al., 2020). Studies on memory binding, however, suggested larger binding effects with (threatening) high arousing, negative stimuli (Kensinger, 2009). Furthermore, higher arousal values for negative stimuli are also typical for large sets of emotional stimuli (e.g, the IAPS; Lang et al., 2005), suggesting that negative emotional arousal is also characteristic for most emotional situations outside of the laboratory. More research is needed to disentangle effects of emotional arousal and emotional valence on binding in memory and action control.

**Theoretical Implications**

The null findings in the present studies are at odds with theoretical claims of a ubiquitous neuromodulatory influence of arousal on binding processes (Mather & Sutherland, 2011; Verguts & Notebaert, 2009). Possible revisions could address the nature of arousal signals as well as the nature of the binding processes. The original adaptation-by-binding account hypothesized arousal signals generated by the monitoring of response conflicts. Subsequent studies demonstrated that arousal signals motivating conflict adaptations are not specific for response conflict and are also evoked by conflicting stimuli and emotional states (Dreisbach & Fischer, 2012a; Dignath & Eder, 2015; Inzlicht et al., 2015, for a review see Dignath et al., 2020). However, one could argue that the phenomenal experience of arousal states generated by response conflict is qualitatively different from arousal produced by negatively arousing stimuli (for suggestive evidence see

Questienne et al., 2018; Morsella et al., 2009). If binding is modulated by the meta-cognitive experience of arousal states (see e.g., Questienne et al., 2016), rather than by noradrenergic arousal signals per se, then differences in phenomenal experiences could explain why binding effects are modulated by response conflict but not by arousing emotional stimuli. Relatedly, studies on the role of affect for conflict adaptation suggest that positive events have a different impact depending on whether they are task- or performance-specific (indicating rewards) or not (positive affect that is unrelated to task performance). For instance, performance-continent rewards strengthened conflict adaptation effects, whereas random gains or losses did not (Stürmer et al., 2011; for an overview, see Dreisbach & Fischer, 2012b). It is possible that the same pattern holds for arousal. This would imply that arousal-based modulations of stimulus-response binding effects might be restricted to situations where arousal is task- or performance-specific. In turn, arousal-based modulations might be absent when arousal is task-irrelevant (like in the present experiments). Future studies could examine these post-hoc explanations with a comparison of different arousal induction procedures and by controlling for differences in the phenomenal experience of arousal on a trial-to-trial basis.

In respect to the nature of binding processes, it is meaningful to distinguish feature binding within objects and modalities (Treisman, 1996) from feature binding across stimulus and response modalities (Hommel et al., 2001). This is especially important since previous studies showed a selective impact of specific neurochemicals on stimulus-stimulus (i.e., within-object), and not on stimulus-response bindings (Colzato et al., 2004; 2005). The present study only addressed binding across stimulus and response modalities, and it is possible that (neuro)modulatory effects of arousal are more pronounced for within-object bindings (as these bindings were also selectively affected by caffeine and moderate levels of alcohol consumption). Furthermore, the ABC model assumes that arousal signals potentiate binding to high-priority stimuli, whereas binding to low-priority, distracting items is diminished. In the present studies, the task-relevant information was the color—and not the identity—of the word. Arousal signals could hence have strengthened connections to the task-relevant colour information at the expense of binding to the distracting word identity information. In this case, however, binding of responses to distractor features should have been smaller, or even absent, with arousing items, which was clearly not observed. To sum up, the present studies do not support theoretical claims that emotional arousal enhances perceptual-action binding. Actions become temporarily associated with perceptual events, forming transient stimulus-response links, but this binding process appears to operate independently of arousal states evoked by emotionally relevant stimuli.

## References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142*(7), 693–728. https://doi.org/10.1037/bul0000047

Anderson, A. K. (2005). Affective Influences on the Attentional Dynamics Supporting Awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. https://doi.org/10.1037/0096-3445.134.2.258

Aston-Jones, G., & Cohen, J. D. (2005). An Integrative Theory of Locus Coeruleus-Norepinephrine Function: Adaptive Gain and Optimal Performance. *Annual Review of Neuroscience*, *28*, 403–450. https://doi.org/10.1146/annurev.neuro.28.061604.135709

Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. *Journal of Experimental Psychology: General*, *147*(2), 243–256. https://doi.org/10.1037/xge0000356

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition.*, *2*(1), 1–38. https://doi.org/10.5334/joc.72

Colzato, L. S., Erasmus, V., & Hommel, B. (2004). Moderate alcohol consumption in humans impairs feature binding in visual perception but not across perception and action. *NEUROSCIENCE LETTERS*, *360*(1-2), 103–105. https://doi.org/10.1016/j.neulet.2004.01.054

Colzato, L. S., Fagioli, S., Erasmus, V., & Hommel, B. (2005). Caffeine, but not nicotine, enhances visual feature binding. *European Journal of Neuroscience*, *21*(2), 591–595. https://doi.org/10.1111/j.1460-9568.2005.03868.x

Dignath, D., & Eder, A. B. (2015). Stimulus conflict triggers behavioral avoidance. *Cognitive, Affective & Behavioral Neuroscience*, *15*(4), 822–836. https://doi.org/10.3758/s13415-015-0355-6

Dignath, D., Eder, A. B., Steinhauser, M., & Kiesel, A. (2020). Conflict monitoring and the affective-signaling hypothesis—an integrative review. *Psychonomic Bulletin & Review*, *27*(2), 193–216. https://doi.org/10.3758/s13423-019-01668-9

Dreisbach, G., & Fischer, R. (2012a). Conflicts as aversive signals. *Brain and Cognition*, *78*(2), 94–98. https://doi.org/10.1016/j.bandc.2011.12.003

Dreisbach, G., & Fischer, R. (2012b). The role of affect and reward in the conflict-triggered adjustment of cognitive control. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00342

Eder, A. B., Erle, T. M., & Kunde, W. (2020). Reward strengthens action–effect binding. *Motivation Science*, *6*(3), 297–302. https://doi.org/10.1037/mot0000153

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., Kiesel, A., Kunde, W., Mayr, S., Moeller, B., Möller, M., Pfister, R., & Philipp, A. (2020). Binding and retrieval in action control (BRAC). *Trends in Cognitive Sciences*, *24*(5), 375–387. https://doi.org/10.1016/j.tics.2020.02.004

Frings, C., & Rothermund, K. (2011). To be or not to be…included in an event file: Integration and retrieval of distractors in stimulus–response episodes is influenced by perceptual grouping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1209–1227. https://doi.org/10.1037/a0023915

Frings, C., & Rothermund, K. (2017). How Perception Guides Action: Figure-Ground Segmentation Modulates Integration of Context Features into S-R Episodes. *Journal of Experimental*

*Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0000403

Frings, C., Rothermund, K., & Wentura, D. (2007). Distractor repetitions retrieve previous responses to targets. *The Quarterly Journal of Experimental Psychology*, *60*(10), 1367–1377. https://doi.org/10.1080/17470210600955645

Giesen, C., Frings, C., & Rothermund, K. (2012). Differences in the strength of distractor inhibition do not affect distractor-response bindings. *Memory & Cognition*, *40*(3), 373–387. https://doi.org/10.3758/s13421-011-0157-1

Giesen, C., & Rothermund, K. (2011). Affective matching moderates S-R binding. *Cognition & Emotion*, *25*(2), 342–350. https://doi.org/10.1080/02699931.2010.482765

Giesen, C., & Rothermund, K. (2015). Adapting to stimulus–response contingencies without noticing them. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(6), 1475–1481. https://doi.org/10.1037/xhp0000122

Giesen, C., & Rothermund, K. (2016). Multi-level response coding in stimulus-response bindings: Irrelevant distractors retrieve both semantic and motor response codes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1643–1656. https://doi.org/10.1037/xlm0000264

Giesen, C. G., Schmidt, J. R., & Rothermund, K. (2020). The law of recency: An episodic stimulus-response retrieval account of habit acquisition. *Frontiers in Psychology*(10), Article 2927, 2927. https://doi.org/10.3389/fpsyg.2019.02927

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

Herbert, C., Kissler, J., Junghöfer, M., Peyk, P., & Rockstroh, B. (2006). Processing of emotional adjectives: Evidence from startle EMG and ERPs. *Psychophysiology*, *43*(2), 197–206. https://doi.org/10.1111/j.1469-8986.2006.00385.x

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*(1), 183–216. https://doi.org/10.1080/713756773

Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research/Psychologische Forschung*, *71*(1), 42–63. https://doi.org/10.1007/s00426-005-0035-1

Hommel, B. , Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–937. https://doi.org/10.1017/S0140525X01000103

Hommel, B., Proctor, R. W., & Vu, K.-P. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*(1), 1–17. https://doi.org/10.1007/s00426-003-0132-y

Inzlicht, M., Bartholow, B. D., & Hirsh, J. B. (2015). Emotional foundations of cognitive control. *Trends in Cognitive Sciences*, *19*(3), 126–132. https://doi.org/10.1016/j.tics.2015.01.004

Janczyk, M., Giesen, C. G., Moeller, B., Dignath, D., & Pfister, R. (2022). Perception and action as viewed from the theory of event coding: A multi-lab replication and effect size estimation of common experimental designs. *Psychological Research.* Advance online publication. https://doi.org/10.1007/s00426-022-01705-8

Kensinger, E. A. (2007). Negative emotion enhances memory accuracy: Behavioral and neuroimaging evidence. *Current Directions in Psychological Science*, *16*(4), 213–218. https://doi.org/10.1111/j.1467-8721.2007.00506.x

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*(2), 99–113. https://doi.org/10.1177/1754073908100432

Kissler, J., Herbert, C., Peyk, P., & Junghofer, M. (2007). Buzzwords: Early cortical responses to emotional words during reading. *Psychological Science*, *18*(6), 475–480. https://doi.org/10.1111/j.1467-9280.2007.01924.x

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. *Technical Report*, A-6.

Lee, T.-H., Sakaki, M., Cheng, R., Velasco, R., & Mather, M. (2014). Emotional arousal amplifies the effects of biased competition in the brain. *Social Cognitive and Affective Neuroscience*, *9*(12), 2067–2077. https://doi.org/10.1093/scan/nsu015

Lewis, P. A., Critchley, H. D., Rotshtein, P., & Dolan, R. J. (2007). Neural correlates of processing valence and arousal in affective words. *Cerebral Cortex*, *17*(3), 742–748. https://doi.org/10.1093/cercor/bhk024

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, *32*(3), 474–488. https://doi.org/10.3758/BF03195840

Mather, M. (2007). Emotional arousal and memory binding: An object-based framework. *Perspectives on Psychological Science*, *2*(1), 33–52. https://doi.org/10.1111/j.1745-6916.2007.00028.x

Mather, M., Clewett, D., Sakaki, M., & & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, *39*. https://doi.org/10.1017/S0140525X15000667

Mather, M., Mitchell, K. J., Raye, C. L., Novak, D. L., Greene, E. J., & Johnson, M. K. (2006). Emotional arousal can impair feature binding in working memory. *Journal of Cognitive Neuroscience*, *18*(4), 614–625. https://doi.org/10.1162/jocn.2006.18.4.614

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*(2), 114–133. https://doi.org/10.1177/1745691611400234

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*(5), 450–452. https://doi.org/10.1038/nn1051

McGinnies, E. (1949). Emotionality and perceptual defense. *Psychological Review*, *56*(5), 244–251. https://doi.org/10.1037/h0056508

Memelink, J., & Hommel, B [Bernhard] (2013). Intentional weighting: A basic principle in cognitive control. *Psychological Research*, *77*(3), 249–259. https://doi.org/10.1007/s00426-012-0435-y

Moeller, B., & Frings, C. (2014). Attention meets binding: Only attended distractors are used for the retrieval of event files. *Attention, Perception, & Psychophysics*, *76*(4), 959–978. https://doi.org/10.3758/s13414-014-0648-9

Moeller, B., & Frings, C. (2017). Dissociation of binding and learning processes. *Attention, Perception, & Psychophysics*, *79*(8), 2590–2605. https://doi.org/10.3758/s13414-017-1393-7

Moeller, B., Frings, C., & Pfister, R. (2016). The structure of distractor-response bindings: Conditions for configural and elemental integration. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(4), 464–479. https://doi.org/10.1037/xhp0000158

Morsella, E., Wilson, L. E., Berger, C. C., Honhongva, M., Gazzaley, A., & Bargh, J. A. (2009). Subjective aspects of cognitive control at different stages of processing. *Attention, Perception, & Psychophysics*, *71*(8), 1807–1824. https://doi.org/10.3758/APP.71.8.1807

Payne, J. D., Nadel, L., Britton, W. B., & Jacobs, W. J. (2004). The Biopsychology of Trauma and Memory. In D. Reisberg & P. Hertel (Eds.), *Series in affective science. Memory and emotion* (pp. 76–128). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195158564.003.0003

Questienne, L., van Dijck, J.-P., & Gevers, W. (2018). Introspection of subjective feelings is sensitive and specific. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 215–225. https://doi.org/10.1037/xhp0000437

Questienne, L., van Opstal, F., van Dijck, J.-P., & Gevers, W. (2016). Metacognition and cognitive control: Behavioural adaptation requires conflict experience. *The Quarterly Journal of Experimental Psychology*, *71*(2), 411–423. https://doi.org/10.1080/17470218.2016.1251473

Schmalbrock, P., Laub, R., & Frings, C. (2021). Integrating salience and action— Increased integration strength through salience. *Visual Cognition*, *29*(2), 91–104. https://doi.org/10.1080/13506285.2020.1871455

Schmidt, J. R., Giesen, C. G [C. G.], & Rothermund, K. (2020). Contingency learning as binding? Testing an exemplar view of the colour-word contingency learning effect. *Quarterly Journal of Experimental Psychology*(73(5)), 739–761.

Sheth, B. R., & Pham, T. (2008). How emotional arousal and valence influence access to awareness. *Vision Research*, *48*(23-24), 2415–2424. https://doi.org/10.1016/j.visres.2008.07.013

Silvert, L., Delplanque, S., Bouwalerh, H., Verpoort, C., & Sequeira, H. (2004). Autonomic responding to aversive words without conscious valence discrimination. *International Journal of Psychophysiology*, *53*(2), 135–145. https://doi.org/10.1016/j.ijpsycho.2004.03.005

Stürmer, B., Nigbur, R., Schacht, A., & Sommer, W. (2011). Reward and punishment effects on error processing and conflict control. *Frontiers in Psychology*, *2*. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2017-40069-001&site=ehost-live

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178. https://doi.org/10.1016/s0959-4388(96)80070-5

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

van der Ploeg, M. M., Brosschot, J. F., Versluis, A., & Verkuil, B. (2017). Peripheral physiological responses to subliminally presented negative affective stimuli: A systematic review. *Biological Psychology*, *129*, 131–153. https://doi.org/10.1016/j.biopsycho.2017.08.051

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, Š., Ly, A., Marsman, M., Matzke, D., Gupta, Akash R. Komarlu Narendra, Sarafoglou, A., Stefan, A., Voelkel, J. G., & Wagenmakers, E.-J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826. https://doi.org/10.3758/s13423-020-01798-5

Verguts, T., & Notebaert, W. (2008). Hebbian learning of cognitive control: Dealing with specific and nonspecific adaptation. *Psychological Review*, *115*(2), 518–525. https://doi.org/10.1037/0033-295X.115.2.518

Verguts, T., & Notebaert, W. (2009). Adaptation by binding: A learning account of cognitive control. *Trends in Cognitive Sciences*, *13*(6), 252–257. https://doi.org/10.1016/j.tics.2009.02.007

Võ, M. L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL–R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Weis, P. P., & Herbert, C. (2017). Bodily reactions to emotional words referring to own versus other people's emotions. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01277

**Acknowledgements**

We thank Georg Reinhardt for help with programming the experiments. This research was supported by grants from the German Research Foundation (GI 1295/2-1) to Carina G. Giesen.

**Table 1**

*Emotional Words (with English translations) Selected from BAWL-R for Experiment 1*

| Arousal level | Word | Arousal | | Valence | |
|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* |
| High | Attentat (assassination) | 4.71 | 0.59 | -2.40 | 0.70 |
| | Psychose (psychosis) | 4.69 | 0.48 | -2.10 | 0.99 |
| | Folter (torture) | 4.68 | 0.58 | -2.80 | 0.52 |
| | Nazi (nazi) | 4.67 | 0.69 | -2.90 | 0.32 |
| | Massaker (massacre) | 4.61 | 0.70 | -2.80 | 0.42 |
| | Panik (panic) | 4.58 | 0.69 | -1.80 | 1.28 |
| | Krieg (war) | 4.57 | 0.60 | -2.90 | 0.32 |
| | Unheil (harm) | 4.53 | 0.62 | -2.15 | 0.67 |
| | Terror (terror) | 4.52 | 0.68 | -2.25 | 1.07 |
| | Alarm (stress) | 4.44 | 0.62 | -1.40 | 0.60 |
| Low | Murmel (marble) | 1.47 | 0.51 | 1.25 | 0.79 |
| | Klee (clover) | 1.39 | 0.70 | 1.80 | 0.92 |
| | Pause (break) | 1.38 | 0.50 | 1.15 | 1.14 |
| | Aquarium (aquarium) | 1.36 | 0.66 | 0.59 | 1.33 |
| | Allee (alley) | 1.33 | 0.49 | 1.20 | 0.63 |
| | Lesesaal (reading room) | 1.33 | 0.59 | 0.00 | 0.47 |
| | Balsam (ageratum) | 1.32 | 0.67 | 1.30 | 0.73 |
| | Erholung (pickup) | 1.29 | 0.59 | 2.30 | 0.67 |
| | Wiege (bassinet) | 1.28 | 0.46 | 1.10 | 0.91 |
| | Schlaf (dormancy) | 1.11 | 0.32 | 1.90 | 0.99 |

*Note*. Emotional arousal was rated on a 5-point scale ranging from 1 (*low arousal*) to 5 (*high arousal*) and emotional valence was rated on a 7-point scale ranging from -3 (*very negative*) to +3 (*very positive*). For LMM analyses, arousal ratings were adjusted by using (M+SD) scores for high arousal stimuli and (M-SD) scores for low arousal stimuli to also account for systematic variance differences between both arousal level categories.

**Table 2**

*Emotional Words (with English translations) with Subjective Ratings of Emotional Arousal and Valence in Experiment 2*

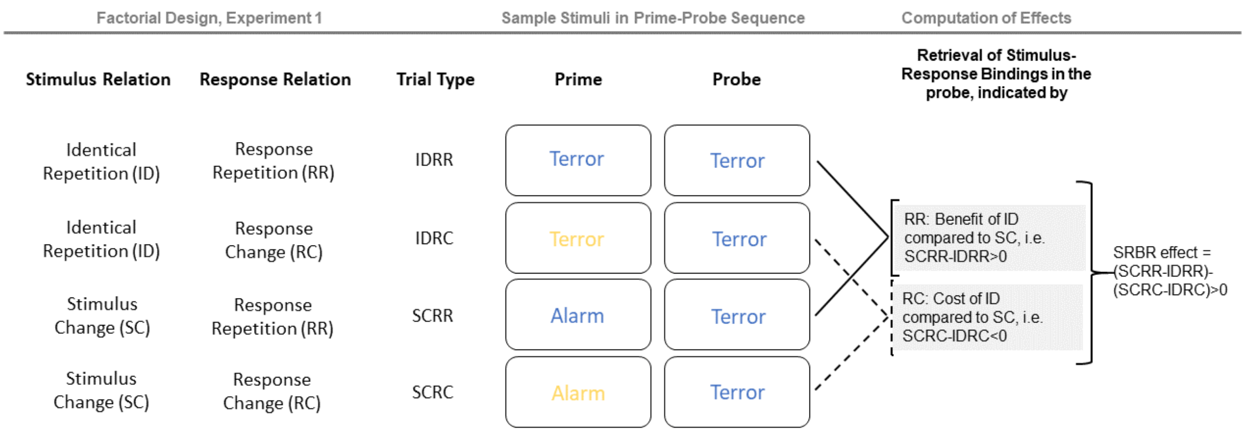
| Arousal level | Word | Arousal | | Valence | |
|---|---|---|---|---|---|
| | | *M* | *SD* | *M* | *SD* |
| High | Alarm (stress) | 7.25 | 1.49 | -2.33 | 1.24 |
| | Angst (anxiety) | 6.93 | 1.59 | -2.77 | 1.29 |
| | Furcht (fear) | 6.45 | 1.80 | -2.55 | 1.27 |
| | Gewalt (violence) | 7.32 | 1.71 | -3.13 | 1.12 |
| | Panik (panic) | 7.71 | 1.55 | -3.11 | 1.11 |
| | Schrei (scream) | 6.78 | 1.69 | -2.4 | 1.17 |
| | Terror (terror) | 7.73 | 1.66 | -3.44 | 1.01 |
| | Unheil (harm) | 6.47 | 1.71 | -2.64 | 1.26 |
| Low | Balsam (ageratum) | 2.13 | 1.38 | 2.57 | 1.39 |
| | Beutel (bag) | 3.76 | 1.56 | 0.33 | 1.02 |
| | Friede (peace) | 1.91 | 1.36 | 3.15 | 1.14 |
| | Harfe (harp) | 2.70 | 1.54 | 1.98 | 1.49 |
| | Schale (bowl) | 3.64 | 1.48 | 0.41 | 1.01 |
| | Schlaf (dormancy) | 1.76 | 1.54 | 3.19 | 1.15 |
| | Seide (silk) | 2.84 | 1.58 | 2.06 | 1.46 |
| | Wiese (meadow) | 2.36 | 1.58 | 2.61 | 1.31 |

*Note.* Emotional arousal and valence were rated on 9-point self-assessment manikin scales ranging from 1 (*low arousal*) to 9 (*high arousal*) and -4 (*very negative*) to +4 (*very positive*), respectively.

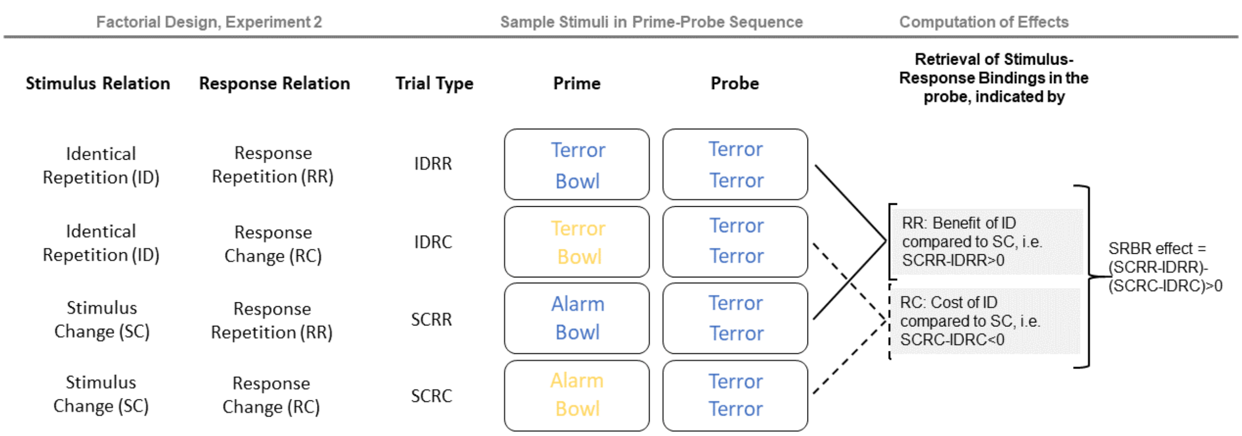
**Figure 1**

*Sample trial displays for (a) Experiment 1 and (b) Experiment 2*

**(a)**




**(b)**

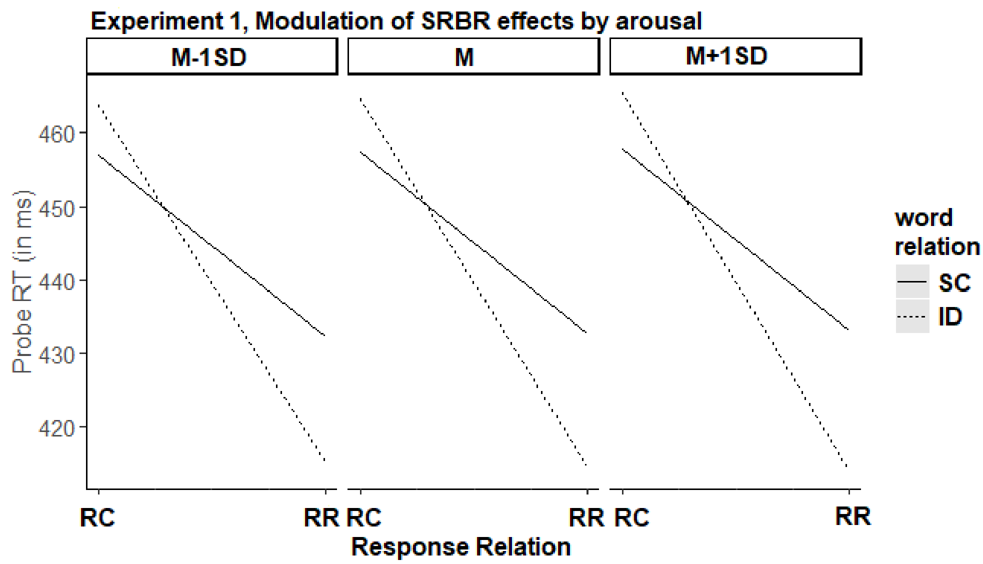



*Note.* Stimuli are not drawn to scale. For illustrative purposes, background colors are inverted (in the studies, stimuli were presented on black background). Only probe displays with high arousing stimuli are presented; however, both studies also used probe displays with neutral stimuli that were constructed in analogue fashion. For Experiment 1, only word changes for words of the same stimulus arousal category are displayed as these reflect the appropriate baseline for stimulus repetition effects.

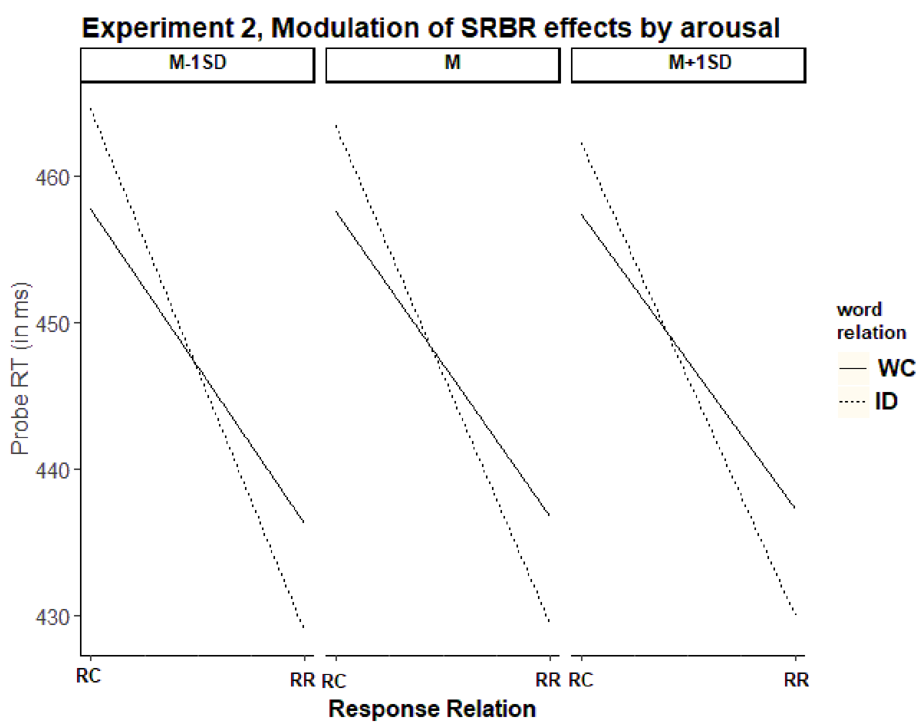
**Figure 2**

*Regression Plots for SRBR Effects as a Function of Different Levels of Arousal Ratings in Experiment 1 (a) and Experiment 2 (b)*

**(a)**




**(b)**



*Note*. Regression plots for SRBR effects (probe RTs as function of the Word Relation by Response Relation interaction) for different levels of arousal ratings (M-1SD, M, M+1SD). The upper panel (a) shows results for Experiment 1 and the lower panel (b) shows results for Experiment 2. ID=identical word repetition; SC= word change/same arousal category. WC=word change. RR=response repetition; RC=response change.

## Supplementary Material

### S1: Preregistered analysis for Experiment 1

**Data preparation.** In line with our preregistered criteria, prime-probe sequences with an incorrect prime and/or probe response were removed for RT analyses (8.1% of the data). In addition, probe reaction times (RT) faster than 200 ms or slower than 1.5 interquartile ranges above the third quartile of the participant's individual RT distribution ("outliers" according to Tukey, 1977) were excluded from all analyses (3.5% of the data). For error analyses, only probe errors that followed correct prime responses were analyzed. We computed mean RT (cf. Table S1) and error rates (cf. Table S2) for the factorial design and subjected these to a 3 (*Word relation*: Identical repetition vs. word change/same arousal level vs. word change/different arousal level) × 2 (*Response relation*: repetition vs. change) × 2 (*Word arousal*: high vs. low) repeated measures analysis of variance (ANOVA). The factor *Word Relation* was split in two orthogonal contrasts (Contrast 1: identical repetition vs. word change/same arousal category; Contrast 2: word change/same arousal category vs. word change/different arousal category). Regarding the *Word Relation by Response Relation* interaction (which evidences standard SRBR effects) as well as the *Word Relation by Response Relation by Word arousal* three-way interaction (which indicates a modulation of SRBR effects by stimulus arousal level), it was hypothesized that response retrieval effects are driven by identical word repetition, meaning that effects should be due to the first contrast (which tests identical word repetition against word change within the same arousal level). To control for inflated risks of type I errors due to multiple comparisons, α-level was Bonferroni-adjusted to .025. Data were analyzed with R (package 'afex').

**Probe RT performance.** The 3×2×2 ANOVA yielded main effects of *Word Relation*, $F(2,200)=14.20$, $p<.001$, $\eta_p^2=.12$, and *Response Relation*, $F(1,100)=209.80$, $p<.001$, $\eta_p^2=.68$, which were further qualified by a significant *Word Relation* × *Response Relation* interaction, $F(2,200)=58.70$, $p<.001$, $\eta_p^2=.37$. A-priori contrasts analyses showed that the interaction was due to the first contrast, $t(100)=8.83$, $p<.001$ (i.e., driven by the repetition/change of words from prime to probe) and not due to the repetition/change of the prime arousal level in the probe, as the second contrast was not significant, $|t|<1$, $p=.879$ (see Fig. S1a). We conducted follow-up *t*-tests on the interaction for the first contrast. For prime-probe sequences with response repetition, identical word repetition facilitated performance compared to word change/same arousal category and led to a performance benefit of $\Delta=18$ ms, $t(100)=9.55$, $p<.001$, $d_z=0.96$; with response change, identical word repetition impeded performance compared to word change/same arousal category and produced performance costs of $\Delta=-7$ ms, $t(100)=4.04$, $p<.001$, $d_z=0.40$. Overall, this pattern is typical for SRBR effects. In contradiction to the arousal-modulation hypotheses, the three-way interaction of stimulus arousal, word relation, and response relation was not significant, $F(2,200)=0.36$, $p=.681$, $\eta_p^2<.01$, which implies that SRBR effects did not differ for words of different arousal levels (Figure 1a). No other effect was significant (all *F*s<1.85, *p*s>.177).

**Probe error rates.** The same ANOVA on probe error rates yielded a significant main effect of Response Relation, $F(1,100)=12.44$, $p=.001$, $\eta_p^2=.11$, that was qualified by an significant Word Relation × Response Relation interaction, $F(2,200)=18.96$, $p<.001$, $\eta_p^2=.16$. A-priori contrast analyses showed that the interaction was due to the first contrast, $t(100)=5.29$, $p<.001$, whereas the second contrast was not significant, $|t|<1$, $p=.462$. Similar to RT analyses, follow-up *t*-tests on the interaction for the first contrast showed that for prime-probe sequences with response repetition, identical word repetition facilitated performance compared to word change/same arousal category and led to fewer errors of $\Delta=1.7\%$, $t(100)=4.14$, $p<.001$, $d_z=0.41$. In turn, for prime-probe sequences with response change, identical word repetition impaired performance compared to word change/same arousal category and increased errors $\Delta=-2.4\%$, $t(100)=4.21$, $p<.001$, $d_z=0.42$, which again represents a pattern that is indicative of SRBR effects. The three-way interaction of stimulus arousal, word relation, and response relation failed the significance threshold, $F(2,200)=2.81$, $p=.067$, $\eta_p^2=.03$. Planned contrast analyses revealed a descriptive trend according to which the three-way interaction was due to the first contrast, reflecting stronger retrieval effects for high compared with low arousal words, $t(100)=1.93$, $p=.056$, but also due to the second contrast, $t(100)=-2.01$, $p=.046$, reflecting smaller retrieval effects for presentations of probe words with same versus different arousal category (Figure S1b). Given that we used a Bonferroni-adjusted α-level of .025, none of these effects can be considered significant. No other effect was significant.

Table S1

*Means and standard deviations for Probe reaction times (in ms) as a function of the factorial design, Experiment 1*

| | | Response Relation | | | |
|---|---|---|---|---|---|
| | | RR | | RC | |
| Stimulus Arousal | Word Relation | *M* | *SD* | *M* | *SD* |
| High | ID-Identical Repetition | 413.96 | 49.91 | 465.79 | 54.61 |
| | SC-Word Change, Same Arousal category | 432.98 | 54.36 | 458.54 | 62.33 |
| | DC-Word Change, Different Arousal category | 432.63 | 52.89 | 459.83 | 61.57 |
| | *Contrast comparisons* | *Δ[SE]* | | *Δ[SE]* | |
| | Contrast 1 (ΔSC-ID) | 19 [1.9] | | -7 [1.7] | |
| | Contrast 2 (ΔDC-SC) | 0 [2.0] | | 1 [2.1] | |
| | SRBR Contrast 1 (=(ΔSC-ID)$_{RR}$-(ΔSC-ID)$_{RC}$) | 26 [3.6] | | | |
| | SRBR Contrast 2 (=(ΔDC-SC)$_{RR}$-(ΔDC-SC)$_{RC}$) | -1 [3.8] | | | |
| Low | ID | 414.95 | 47.86 | 464.50 | 53.52 |
| | SC | 432.77 | 49.15 | 457.45 | 55.57 |
| | DC | 435.60 | 53.97 | 457.85 | 55.91 |
| | *Contrast comparisons* | *Δ[SE]* | | *Δ[SE]*0 | |
| | Contrast 1 (ΔSC-ID) | 18 | | -7 | |
| | Contrast 2 (ΔDC-SC) | 3 | | 1 | |
| | SRBR Contrast 1 (=(ΔSC-ID)$_{RR}$-(ΔSC-ID)$_{RC}$) | 25 [4.1] | | | |
| | SRBR Contrast 2 (=(ΔDC-SC)$_{RR}$-(ΔDC-SC)$_{RC}$) | 2 [3.6] | | | |

*Note.* RR= response repetition. RC=response change. *M* and *SD* represent mean and standard deviation, respectively. SRBR= Stimulus-response binding and retrieval effect. For contrast comparisons, standard errors of the mean are presented in brackets.

Table S2

*Means and standard deviations for Probe error rates as a function of the factorial design, Experiment 1*

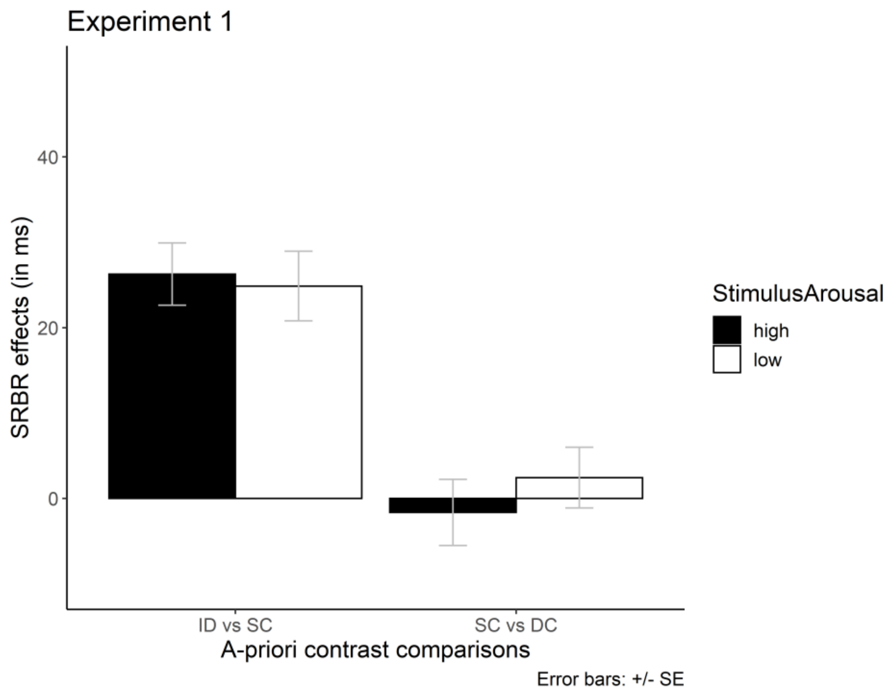
| | | Response Relation | | | |
|---|---|---|---|---|---|
| | | RR | | RC | |
| Stimulus Arousal | Word Relation | *M* | *SD* | *M* | *SD* |
| High | ID-Identical Repetition | 0.01 | 0.02 | 0.05 | 0.07 |
| | SC-Word Change, Same Arousal category | 0.03 | 0.06 | 0.02 | 0.05 |
| | DC-Word Change, Different Arousal category | 0.03 | 0.06 | 0.04 | 0.08 |
| | *Contrast comparisons* | *Δ[SE]* | | *Δ[SE]* | |
| | Contrast 1 (ΔSC-ID) | 0.02 [0.01] | | -0.03 [0.01] | |
| | Contrast 2 (ΔDC-SC) | -0.00 [0.01] | | 0.02 [0.01] | |
| | SRBR Contrast 1 (=(ΔSC-ID)$_{RR}$-(ΔSC-ID)$_{RC}$) | 0.05 [0.01] | | | |
| | SRBR Contrast 2 (=(ΔDC-SC)$_{RR}$-(ΔDC-SC)$_{RC}$) | -0.02 [0.01] | | | |
| Low | ID | 0.01 | 0.03 | 0.05 | 0.06 |
| | SC | 0.02 | 0.05 | 0.03 | 0.06 |
| | DC | 0.03 | 0.06 | 0.03 | 0.05 |
| | *Contrast comparisons* | *Δ[SE]* | | *Δ[SE]* | |
| | Contrast 1 (ΔSC-ID) | 0.01 [0.01] | | -0.02 [0.01] | |
| | Contrast 2 (ΔDC-SC) | 0.01 [0.01] | | 0.00 [0.01] | |
| | SRBR Contrast 1 (=(ΔSC-ID)$_{RR}$-(ΔSC-ID)$_{RC}$) | 0.03 [0.01] | | | |
| | SRBR Contrast 2 (=(ΔDC-SC)$_{RR}$-(ΔDC-SC)$_{RC}$) | 0.01 [0.01] | | | |

*Note.* RR= response repetition. RC=response change. *M* and *SD* represent mean and standard deviation, respectively. SRBR= Stimulus-response binding and retrieval effect. For contrast comparisons, standard errors of the mean are presented in brackets.

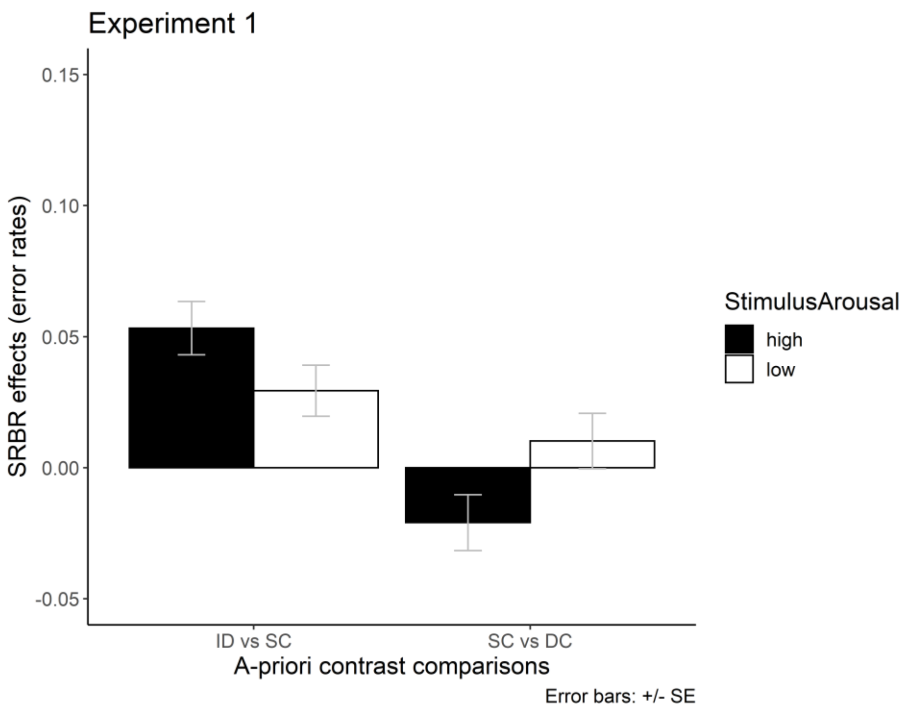
**Figure S1**

*SRBR effects for (a) probe RT and (b) error rates for preregistered analyses of Experiment 1.*

**(a)**




**(b)**



*Note*. SRBR effects as a function of Prime Arousal and a-priori contrast comparison for levels of the *Word Relation* factor (ID=identical repetition; SC= word change/same arousal category; DC=word change/different arousal category; see Table S1 for computation of SRBR effects for each contrast, respectively).

**S2: Analyses of error rates in Experiment 1 and 2**

In both studies, probe reaction times (RT) served as dependent variable of interest (see main text and preregistrations). For exploratory purposes, we ran similar analyses on probe error rates as well.

**Data preparation.** Only probe errors that followed correct prime responses were analyzed with a generalized linear mixed-effects model (GLMM), using trials as unit of analysis. Participants were treated as level-2 predictor (with trials nested within participants). Similar to RT analyses, we computed a random intercept model. Probe error rates served as dependent measure. We added fixed effects for word relation (contrast coded with identical repetition =1, word change =-1 [for Experiment 1, word change referred to the word change/same arousal category level, whereas the word change/different arousal category level was discarded for analyses]), response relation (response repetition=1, response change=-1) and a arousal ratings for each probe stimulus (metric predictor, grand mean centered) and their interactions. Last, we included participants as random effect. The model specification was as follows: ProbeERR~1+WordRelation*ResponseRelation*ArousalRating + (1|Subject).

**Experiment 1**

The analysis yielded significant effects of word relation, $b$=-0.14, z=-2.46, $p$=.013, response relation, $b$=-0.41, z=-7.36, $p$<.001, and for their interaction, $b$=-0.42, $z$=-7.32, $p$<.001. In contrast to RT, the three-way interaction was also significant, $b$=-0.14, z=-2.58, $p$=.009, indicating that SRBR effects differed as a function of stimulus arousal level (Figure S2a) which increasing SRBR effects for stimuli with higher arousal level. No other effect was significant.

**Experiment 2**

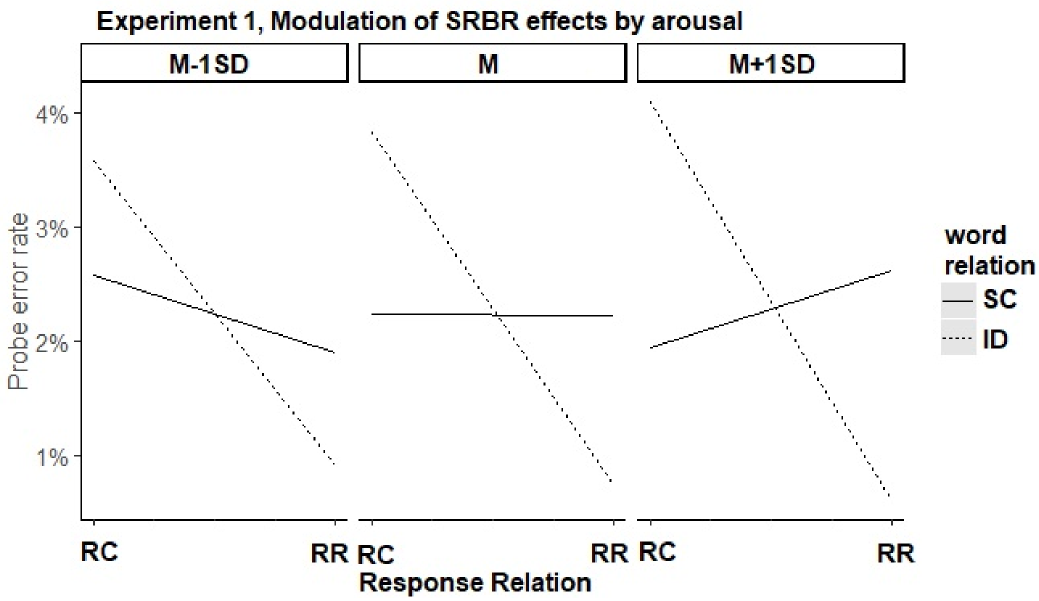
For error rates in Experiment 2, the same analysis yielded significant effects of response relation, $b$=-0.29, z=-2.73, $p$=.006, and for the Word Relation by Response Relation interaction,

$b$=-0.51, $z$=-3.27, $p$=.001. Similar to RT, the three-way interaction was not significant, $b$=0.07, z=1.31, $p$=.189, indicating that SRBR effects were of equal magnitude for probe words of high and low arousal level (Figure S2b). No other effect was significant.

**Figure S2**

Regression plots for SRBR effects for different levels of arousal ratings for error rates of (a) Experiment 1 and (b) Experiment 2.

**(a)**




**(b)**