Generalization Versus Contextualization in Automatic Evaluation Revisited: A Meta-Analysis of Successful and Failed Replications

Bertram Gawronski
University of Texas at Austin, USA
Xiaoqing Hu
University of Texas at Austin, USA
Robert J. Rydell
Indiana University, USA
Bram Vervliet
University of Leuven, Belgium
Jan De Houwer
Ghent University, Belgium

To account for disparate findings in the literature on automatic evaluation, Gawronski, Rydell, Vervliet, and De Houwer (2010) proposed a representational theory that specifies the contextual conditions under which automatic evaluations reflect initially acquired attitudinal information or subsequently acquired counterattitudinal information. The theory predicts that automatic evaluations should reflect the valence of expectancy-violating counterattitudinal information only in the context in which this information had been learned. In contrast, automatic evaluations should reflect the valence of initial attitudinal information in any other context, be it the context in which the initial attitudinal information had been acquired (*ABA renewal*) or a novel context in which the target object had not been encountered before (*ABC renewal*). The current article presents a meta-analysis of all published and unpublished studies from the authors' research groups regardless of whether they produced the predicted pattern of results. Results revealed average effect sizes of $d = 0.249$ for ABA renewal (30 studies, $N = 3,142$) and $d = 0.174$ for ABC renewal (27 studies, $N = 2,930$), both of which were significantly different from zero. Effect sizes were moderated by attention to context during learning, order of positive and negative information, context-valence contingencies during learning, and sample country. Although some of the obtained moderator effects are consistent with the representational theory, others require theoretical refinements and future research to gain deeper insights into the mechanisms underlying contextual renewal.

Keywords: automatic evaluation; context effects; evaluative learning; meta-analysis; occasion setting; renewal effects

In 2010, we published an article in the *Journal of Experimental Psychology: General* that outlined a representational theory of context effects on automatic evaluation (Gawronski, Rydell, Vervliet, & De Houwer, 2010). The theory aimed to reconcile disparate findings in the literature by specifying the contextual conditions under which automatic evaluations reflect initially acquired attitudinal information or subsequently acquired counterattitudinal information. In addition to presenting the core assumptions of the theory, the article reported the results of four experiments that tested several novel predictions derived from this theory.

Since the publication of this article, we conducted several follow-up studies to further examine the empirical implications of our theory. Although many of these studies replicated our initial findings (e.g., Gawronski, Ye, Rydell, & De Houwer, 2014), some of them failed to reproduce the basic effects that served as the foundation for our theory (cf. Rydell & Gawronski, 2009). Such replication failures can be problematic if selective reports of successful studies lead to inadequate conclusions about the average size of these effects (Murayama, Pekrun, & Fiedler, 2014). In the worst case, our initial findings may turn out as false positives if a comprehensive analysis of all studies fails to obtain statistically significant results. Of course, a certain proportion of replication failures have to be expected, because failed replications can also be due to unknown moderators and naturally occurring variations in effect sizes as a result of measurement and sampling error (Cummings, 2012; LeBel & Paunonen, 2011; Stanley & Spence, 2014). Nevertheless, the mere existence of several failed replications requires a thorough reassessment of the available evidence to reevaluate the validity of our theory.

The current article addresses this issue by means of a meta-analysis that includes all relevant studies from our research groups—published and unpublished—regardless of whether they replicated the basic effects that provided the foundation for our theory. By conducting such a meta-analysis, we aimed to achieve three goals. First, by aggregating all datasets from our labs, we aimed to provide a more accurate estimate of the average effect sizes on the basis of a relatively large sample, and thus a more rigorous test of whether the effects reported in our earlier article are indeed genuine. Second, we aimed to determine whether our conflicting findings are due to sampling error or systematic procedural factors that varied across studies (Hunter & Schmidt, 2004). Finally, to the extent that our meta-analysis reveals a systematic contribution of procedural factors, we aimed to test the role of specific moderators that may influence the size of the obtained effects.

The Representational Theory

Our theory was inspired by conflicting findings in the literature on automatic evaluation, suggesting that automatic evaluations can be highly robust and difficult to change, highly malleable and easy to change, and highly

context-dependent (for a review, see Gawronski & Sritharan, 2010). To account for these findings, we proposed a representational theory that specifies the contextual conditions under which automatic evaluations reflect initially acquired attitudinal information or subsequently acquired counterattitudinal information.

Like many other theories of evaluative learning, our theory assumes that the encoding of evaluative information about a target object creates a memory trace that links the object to that information. To the extent that this memory trace is sufficiently strong, it will be automatically activated upon future encounters with the object, thereby eliciting an automatic evaluative response that is in line with the valence of the stored information (Fazio, 2007; Gawronski & Bodenhausen, 2006).

Although context is typically irrelevant when stored information about an object is evaluatively homogenous, contextual cues are assumed to constrain the activation of evaluatively inconsistent information, thereby moderating automatic evaluations across contexts. According to our representational theory, such context effects depend on the integration of contextual cues into the representation of evaluative information. Specifically, our theory assumes that whether contextual cues are integrated into the mental representation of evaluative information depends on perceivers' attention to the context during encoding. If perceivers pay attention to the context during learning, contextual cues will be integrated into the representation of the newly acquired information. If, however, perceivers do not pay attention to the context during learning, contextual cues will not be integrated into the representation. Whereas the former case leads to the formation of a *contextualized representation*, the latter case leads to the formation of a *context-free representation*.

Another central assumption of our theory is that attention to context is typically low during the encoding of initial attitudinal information, but enhanced by exposure to expectancy-violating counterattitudinal information (Roese & Sherman, 2007). In addition, the theory assumes that newly acquired counterattitudinal information is simply added to the existing memory structures instead of erasing initially acquired attitudinal information from memory (cf. Petty, Tormala, Briñol, & Jarvis, 2006; Wilson, Lindsey, & Schooler, 2000). Hence, if a person is confronted with counterattitudinal information about an object that conflicts with initially acquired attitudinal information, the mental representation of the object is assumed to acquire a "dual" nature that involves (1) a context-free representation of the initial attitudinal information, and (2) a contextualized representation of the expectancy-violating counterattitudinal information. Combined with the notion of pattern matching in memory retrieval (Smith, 1996), these assumptions imply that automatic evaluations should reflect the valence of the counterattitudinal information only in the context in which this information had been learned. In contrast, automatic evaluations should reflect the valence of the initial

attitudinal information in any other context, be it the context in which this information had been acquired or a novel context in which the target object had not been encountered before (see Figure 1).

Adopting terminology to describe similar patterns in animal learning, we used the term *renewal effect* to describe the context-dependent recurrence of an initially acquired attitudinal response (see Bouton, 2004) and the term *occasion setting* to describe the modulating function of the context of counterattitudinal information (see Schmajuk & Holland, 1998). *ABA renewal* refers to cases in which initial attitudinal information is acquired in Context A, counterattitudinal information is subsequently acquired in a different Context B, and the initial attitudinal information determines evaluative responses in the initial Context A (cf. Bouton & Bolles, 1979; Bouton & Peck, 1989). Correspondingly, *ABC renewal* refers to cases in which initial attitudinal information is acquired in Context A, counterattitudinal information is subsequently acquired in a different Context B, and the initial attitudinal information determines evaluative responses in a novel Context C (cf. Bouton & Bolles, 1979; Bouton & Brooks, 1993). These patterns differ from ABB scenarios in which initial attitudinal information is acquired in Context A, counterattitudinal information is subsequently acquired in a different Context B, and the counterattitudinal information determines evaluative responses in Context B. Together with the dominance of initial attitudinal information in Contexts A and C, the modulating function of Context B can be described as an instance of occasion setting, in that the presence versus absence of Context B modulates the evaluative response that is elicited by the attitude object (for a review, see Gawronski & Cesario, 2013).

To investigate contextual renewal and occasion setting in automatic evaluation, Gawronski et al. (2010) adopted an evaluative learning paradigm by Rydell and Gawronski (2009). Participants were first presented with positive or negative information about a target person against a meaningless, colored background. In a second block, participants were presented with new information about the target that was evaluatively opposite to the information provided in the first block, and this information was presented against a different colored background. After the impression formation task, automatic evaluations of the target were assessed with an affective priming task in which the target was presented against the background of the first block (Context A), the background of the second block (Context B), or a novel background that was not part of the impression formation task (Context C). Confirming the effectiveness of the counterattitudinal information, results showed that automatic evaluations in Context B reflected the valence of the information that was presented in the second block of the impression formation task. However, consistent with the notion of contextual renewal, automatic evaluations reflected the valence of the initial information when the target was presented against the

initial Context A (ABA renewal) or the novel Context C (ABC renewal).

The Current Meta-Analysis

The current meta-analysis aimed to provide a more accurate estimate of the average effect sizes of ABA and ABC renewal in Gawronski et al.'s (2010) evaluative learning paradigm. By including all published and unpublished studies from our research groups regardless of whether they did or did not replicate our initial findings, this meta-analysis offers a more rigorous test of contextual renewal in automatic evaluation. To the extent that the meta-analysis confirms the reliability of our initial findings, the obtained effect sizes will also provide more accurate information for decisions about appropriate sample sizes in future research using Gawronski et al.'s (2010) paradigm (Cohen, 1988).

Moderator Analyses

In addition to providing a more accurate estimate of the average effect sizes of contextual renewal in automatic evaluation, we aimed to determine whether our conflicting findings are due to sampling error or systematic procedural factors (Hunter & Schmidt, 2004). To the extent that the meta-analysis reveals a systematic contribution of procedural factors, we aimed to test the contribution of several moderators that may influence the size of the obtained effects.

Attention to context during learning. The concepts of ABA and ABC renewal are essential for our representational theory, in that they describe the default pattern of how automatic evaluations should vary across contexts when the representation of an attitude object includes evaluatively inconsistent information. Yet, the theory also makes specific predictions about the conditions under which the two kinds of renewal effects should *not* occur. For example, if attention to the context is enhanced during the encoding of initial attitudinal information, both attitudinal and counterattitudinal information are assumed to be stored in contextualized representations (Gawronski et al., 2010). Thus, to the extent that a novel context is equally (dis)similar to either of these contexts, encountering the target in such a context should activate the two representations to the same extent, thereby producing an averaging effect of the two kinds of information rather than a renewal effect. Yet, automatic evaluations in either of the two known contexts should be driven by the contextualized representations that include the respective information that had been learned in these contexts. Together, these assumptions imply that enhanced attention to contextual cues during the encoding of initial attitudinal information should reduce ABC renewal, but not ABA renewal. These predictions have been confirmed in two experiments by Gawronski et al. (2010) that manipulated participants' attention to contextual cues during the encoding of initial attitudinal information.

Another prediction of our theory is that contextual cues should not be integrated into the representation of counterattitudinal information when attention to context during the encoding of subsequent counterattitudinal information is low. In this case, initial attitudinal and subsequent counterattitudinal information should be stored in a single context-free representation. As a result, context effects should be eliminated altogether, such that automatic evaluations reflect a mixture of the available information regardless of the context. These predictions have been confirmed in an experiment by Gawronski et al. (2010) that manipulated participants' attention to contextual cues during encoding of subsequent attitudinal information.

To provide a more rigorous test of our representational theory, we determined for all experimental conditions of the included studies whether they involved (1) default conditions without a direct manipulation of attention, (2) an experimental manipulation designed to enhance attention to the context during initial attitudinal learning, or (3) an experimental manipulation designed to reduce attention to the context during subsequent counterattitudinal learning. Our primary question was whether effect sizes of ABA and ABC renewal vary in line with the predictions of our theory regarding the impact of attention to context during learning.

Valence order. Another question of theoretical importance is whether renewal effects depend on the order in which positive and negative information is acquired. Research on negativity bias has shown that negative information has a stronger impact than positive information (for reviews, see Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001; Cacioppo & Berntson, 1994; Skowronski & Carlston, 1989). On the basis of these findings, one could argue that subsequently acquired negative information may be more effective in influencing automatic evaluations across contexts than positive information, thereby leading to weaker renewal effects when initial positive information is challenged by negative information than when initial negative information is challenged by positive information. To address the possibility of such valence asymmetries, the current meta-analysis tested whether the relative size of renewal effects differs as a function of valence order.

Context-valence contingencies. A central assumption of our representational theory is that contextual cues constrain the activation of evaluative information about the target object, such that their presence versus absence moderates the evaluative response that is elicited by the object (*occasion setting*). This hypothesis is based on earlier theorizing by Bouton (1994) who argued that Context B constrains the spread of activation from the target object to the available evaluative information by virtue of inhibitory links. If Context B is absent, activation of the target object is assumed to spread to the initial attitudinal information, which in turn inhibits the activation of the counterattitudinal information. In contrast, if Context B is present, activation of the context node is assumed to inhibit the link between the target object and the initial attitudinal information, thereby gating the spread of activation from the target object to the counterattitudinal information, which further inhibits the activation of the initial attitudinal information. Yet, an alternative

possibility is that the contexts themselves become associated with the valence of counterattitudinal experience during encoding (*evaluative binding*), and thus directly elicit a corresponding evaluative response (see Vervliet, Baeyens, Van den Bergh, & Hermans, 2013). From this perspective, context effects on evaluative responses may reflect additive effects of independent excitatory links between (1) the target object and the available information about that object and (2) Context B and the counterattitudinal experience. As a result, activation of the counterattitudinal experience should be stronger when the target is encountered in Context B than when it is encountered in any other context.

In the experimental design of Gawronski et al.'s (2010) studies, either mechanism can lead to the observed pattern of results, because the learning paradigm involved a perfect contingency between the valence of evaluative information and the context in which this information had been presented (cf. De Houwer, Thomas, & Baeyens, 2001). To provide a more stringent test of the occasion setting hypothesis, Gawronski et al. (2014) tested the emergence of renewal effects in a modified variant of Gawronski et al.'s (2010) paradigm involving two targets instead of one. To avoid contingencies between context and valence, one of the two targets was described as positive in the first learning block while the other one was described as negative in the first learning block. In the second learning block, the initially positive target was presented with negative information and the initially negative target was presented with positive information. Consistent with the notion of occasion setting, evaluative responses toward the two targets were moderated by the presence versus absence of the context color of the second learning block despite the absence of any contingency between context and valence.

Although Gawronski et al.'s (2014) findings are consistent with the notion of occasion setting, the mere demonstration of contextual renewal in the absence of context-valence contingencies does not rule out the possibility that direct associations between context and valence contribute to renewal effects over and above occasion setting when there is a contingency between context and valence. To test this possibility, we coded whether renewal effects were investigated with (1) a between-subjects manipulation of valence order that involved a contingency between context and valence (see Gawronski et al., 2010) or (2) a within-subjects manipulation of valence order that did not involve a contingency between context and valence (see Gawronski et al., 2014). A contribution of occasion setting would be indicated by significant renewal effects in studies that used a within-subjects design. A contribution of evaluative binding would be indicated if the effect sizes of contextual renewal are significantly larger in between-subjects designs than within-subject designs.

Type of measure. Gawronski et al.'s (2010) research was primarily concerned with renewal effects on automatic responses. Toward this end, they relied on a variant of affective priming to measure automatic evaluations: the affect misattribution procedure (AMP; Payne, Cheng, Govorun, & Stewart, 2005). On each trial of the AMP, participants were briefly presented with the target object against one of the three background colors that represented the initial Context A, the subsequent Context B, and the novel Context C. The target object was followed by a brief presentation of a Chinese ideograph, which was replaced by a black-and-white pattern mask. Participants' task was to indicate whether they considered the Chinese ideograph as visually more pleasant or visually less pleasant than the average Chinese ideograph. Priming effects in the AMP are indicated by larger proportions of "pleasant" responses when participants were primed with a positive stimulus than when they were primed with a negative stimulus. AMP effects are typically considered "automatic" in the sense that the primes influence responses to the targets in an unintentional manner (see Gawronski & Ye, 2015; Payne, Brown-Iannuzzi, Burkley, Arbuckle, & Cooley, 2013). Yet, an important question is whether renewal effects generalize to other measures beyond the AMP. In the current meta-analysis, we addressed this question by testing whether the size of renewal effects differs as a function of the measurement procedure. On the basis of the available data, we were able to investigate the generality of renewal effects by comparing effect sizes in studies using the AMP (see Payne et al., 2005), studies using Fazio, Jackson, Dunton, and Williams's (1995) evaluative priming task (EPT), studies that used a speeded evaluation task (SET) involving intentional evaluations under time pressure (see Ranganath, Smith, & Nosek, 2008), and studies that used a modified version of the AMP involving a longer stimulus onset asynchrony (SOA) between the presentation of the primes and the targets (see Hofmann, Friese, & Roefs, 2009).

Repeated measurement. An important procedural aspect of the included studies is whether evaluative responses were measured only after the second block of the evaluative learning task, or additionally after the first block. With the exception of Gawronski et al.'s (2010) Experiment 3, all of their studies measured automatic evaluations only after the second learning block. Yet, there is considerable heterogeneity in the dataset of the current meta-analysis, such that some studies measured evaluative responses only after the second learning block, whereas others measured evaluative responses after both learning blocks. Although our representational theory does not include specific hypotheses about the effects of repeated measurement, a delay between the learning of initial attitudinal and subsequent counterattitudinal information might influence renewal effects through processes of memory consolidation. Because the measurement of evaluative responses after the first learning block involves a longer delay between initial attitudinal and subsequent counterattitudinal information, the size of renewal effects may differ as a function of prior measurement. In addition, it is possible that the mere act of completing an evaluation

measure influences renewal effects by signaling a new period of evaluative information (e.g., Lipp & Purkis, 2006), which may increase attention to the context of counterattitudinal information. Thus, an important question is whether the size of renewal effects differs as a function of prior measurement of evaluative responses.

Sample characteristics. A final question addressed in the current meta-analysis concerns sample characteristics, particularly the cultural background of the participants. Drawing on evidence that individuals from East Asian cultures tend to pay more attention to contextual information than individuals from Western cultures (e.g., Chua, Boland, & Nisbett, 2005; Masuda & Nisbett, 2001), Gawronski et al. (2010) speculated that East Asians may show enhanced attention to the context of both initial attitudinal and subsequent counterattitudinal information. In contrast, Westerners may be more likely to show the default pattern hypothesized by our representational theory, such that they pay attention to the context only during the encoding of expectancy-violating counterattitudinal information, but not during the encoding of initial attitudinal information. Together with Gawronski et al.'s (2010) finding that enhanced attention to the context during initial attitudinal information reduces ABC renewal, but not ABA renewal, these considerations suggest that East Asians may be less likely to show ABC renewal than Westerners. Yet, there should be no cultural differences with regard to ABA renewal.

More recently, Ye and Gawronski (in press) discussed an alternative prediction derived from research on cultural differences in the tolerance for inconsistency (Choi & Nisbett, 2000; Peng & Nisbett, 1999). Specifically, they argued that higher tolerance for inconsistency among East Asians may lead them to pay less attention to the context during the encoding of counterattitudinal information than Westerners. As a result, East Asians may be more likely to store initial attitudinal and subsequent counterattitudinal information in a single context-free representation, which should eliminate both ABA and ABC renewal. The counterintuitive, yet very interesting, prediction derived from these assumptions is that automatic evaluations should be less susceptible to context effects in East Asians than Westerners. Whereas automatic evaluations in Westerners should reflect the valence of counterattitudinal information in the context in which this information had been learned and the valence of initial attitudinal information in any other context, automatic evaluations in East Asians may reflect a mixture of the available information regardless of the context. In the current meta-analysis, we had the opportunity to investigate cultural differences in contextual renewal by comparing effect sizes across samples with different cultural backgrounds.

Procedures

Data Inclusion

All datasets were from studies conducted by the research teams of the authors of Gawronski et al.'s (2010) article. The dataset included 30 individual experiments (10 published, 20 unpublished) with a total sample size of 3,142 participants, 54 independent samples, and 94 experimental conditions. All of the included studies utilized Gawronski et al.'s (2010) evaluative learning paradigm (cf. Rydell & Gawronski, 2009). Although most of the studies examined both ABA and ABC renewal, there was a small number of studies that tested only ABA renewal, but not ABC renewal. The dataset for ABC renewal included 27 individual experiments (7 published, 20 unpublished) with a total sample size of 2,930 participants, 49 independent samples, and 78 experimental conditions.

Coding

Studies were coded in a hierarchical manner. At the study level, we assigned an identification number to each study (study-ID) and recorded the reference, publication status, the year in which the data were collected, and the sample size of the study. At the procedure level, we recorded for each experimental condition: (1) the sample size of each cell of the experimental design; (2) whether it involved an experimental manipulation of attention to the context during learning (default without manipulation, enhanced during first learning block, reduced during second learning block); (3) the order in which evaluative information was learned (positive-negative vs. negative-positive); (4) whether valence order was manipulated between-subjects or within-subjects; (5) the measure that was used to assess automatic evaluations (AMP, EPT, SET, AMP with long SOA); (6) whether evaluations were measured only after the second learning block or additionally after the first learning block; and (7) the country in which the study was conducted (United States, Canada, Singapore, Italy, Belgium).

Effect Size Calculation, Correction, and Weights

For effect size estimates, we used Cohen's d for individual effect size. Because renewal effects depend on the availability of evaluatively inconsistent information, all calculations were based on participants' evaluative responses after the second learning block. First, we calculated the means and standard deviations of participants' evaluative responses to a given target in the three contexts: the initial learning Context A, the second learning Context B, and the novel Context C. ABA renewal was defined as the difference in evaluative responses between Context A and Context B; ABC renewal was defined as the difference in evaluative responses between Context C and Context B (see Figure 1). Because evaluative responses in the three contexts were measured on a within-subjects basis for each individual participant, we calculated Cohen's ds and their standard errors on the basis of the mean values, standard deviations, and the corresponding correlations between the two measures (see Lipsey & Wilson, 2001). Each individual effect size was then corrected for differences in precision using the formula proposed by Hedges (1981, see also Lipsey & Wilson, 2001). The inverse of the squared standard error of each effect size was used as weight (i.e., inverse variance

weight) for the meta-analysis. Thus, studies with high precision (i.e., studies with large sample sizes) weighted more than studies with low precision (i.e., studies with small sample sizes).

Combining Multiple Effect Sizes within Experiments

Including multiple effect sizes from the same sample violates the independent effect size assumption (see Lipsey & Wilson, 2001). In the current meta-analysis, there were several samples that yielded more than one effect size. For example, in some experiments, the same participants completed multiple measures of evaluation (e.g., AMP and EPT) with each measure generating one effect size. In such cases, we first identified the individual effect sizes, and then averaged the identified scores in a single weighted effect size. This procedure was repeated for each moderator to ensure that each individual effect size was independent for the moderator analyses (see below).

Meta-Analytic Computations

For the present meta-analysis, we chose a random-effects model over a fixed-effects model. The key difference between the two models is that, whereas a fixed-effects model assumes homogeneity across studies and effect sizes (i.e., all observed errors are sampling errors), a random-effects model assumes systematic variations across studies (Hunter & Schmidt, 2004). For the current analysis, we chose a random-effects model for three reasons. First, one of the motivations for the current meta-analysis is the observed heterogeneity of findings across studies. Because the included studies differ in terms of multiple procedural parameters, it seems unlikely that the observed heterogeneity can be completely attributed to sampling error. Second, our representational theory implies specific predictions about whether and when ABA and ABC renewal should occur (e.g., enhanced attention to context during initial attitudinal learning should reduce ABC renewal, but not ABA renewal). Thus, the parameters identified by our theory should lead to systematic differences in effect sizes over and above sampling error. Third, a fixed-effects model is known to produce a high rate of false-positive Type I errors, because the model assumes a fixed effect size across studies, presupposing that between-study variance is relatively small. In contrast, a random-effects model imposes less constraints on variations in effect sizes, and thereby allows one to estimate the level of heterogeneity across studies. Thus, a random-effects model can avoid potential false positive findings and, if anything, provides more conservative estimates than a fixed-effects model.

For the actual analyses, we used the SPSS macro developed by David Wilson (Lipsey & Wilson, 2001) to perform the main analysis and the tests of categorical moderators. As mentioned above, each individual effect size was corrected and weighted by their inverse variance. Cochran's Q statistic was used to estimate the degree of heterogeneity across effect sizes. The categorical moderator analysis was analogous to an ANOVA, in which total variances across individual effect sizes were partitioned into between-group variances (Q_B, i.e., deviation for each group's mean around the grand mean) and within-group variances (Q_W, i.e., deviation for each individual effect size within the group around the group's mean). Q_B is then tested in a Chi-square distribution for significance with df $= j$ -1 where j is the number of groups.

Results

Overall Effects

The effect sizes of ABA renewal ranged from -0.177 to 1.175, with a meta-analytic effect size of 0.249, $SE =$ 0.035. The 95% confidence interval excluded zero, CI [0.180, 0.317], and the effect size was significantly different from zero, $Z = 7.13$, $p < .001$. Cochran's Q statistic for homogeneity analysis revealed a significant effect, $Q(53) = 72.72$, $p = .04$, suggesting that systematic differences between studies contributed to variations in effect sizes.

The effect sizes of ABC renewal ranged from -0.191 to 1.178, with a meta-analytic effect size of 0.174, $SE =$ 0.033. The 95% confidence interval excluded zero, CI [0.109, 0.238], and the effect was significantly different from zero, $Z = 5.26$, $p < .001$. Cochran's Q statistic for homogeneity analysis revealed a significant effect, $Q(48) =$ 72.25, $p = .01$, again suggesting that systematic differences between studies contributed to variations in effect sizes.

Although effect sizes for ABA renewal were numerically larger than effect sizes for ABC renewal (ds $=$ 0.249 vs. 0.174), the two effects did not significantly differ in terms of their relative size, $Q(1) = 1.85$, $p = .17$.[1]

Moderator Analyses

The mean effect sizes of ABA renewal as a function of the reviewed moderators are depicted in Figure 2; the mean effect sizes of ABC renewal are depicted in Figure 3. Statistical information on standard errors and confidence intervals is provided in Table 1.

Attention to context. On the basis of our representational theory, we determined for each experimental condition if it involved an experimental manipulation of attention to the context during learning. The three coding categories were (1) default condition without experimental manipulation of attention, (2) experimental manipulation designed to enhance attention to the first learning context, or (3) experimental manipulation designed to reduce attention to the second learning context.

---

[1] Although our representational theory does not predict differences in the relative size of ABA and ABC renewal under default conditions, numerically smaller effects of ABC renewal are consistent with the assumptions that ABC renewal should be reduced when either (1) attention to the first learning context is enhanced or (2) attention to the second learning context is reduced. In contrast, ABA renewal should be reduced only when attention to the second learning context is reduced, but not when attention to the first learning context is enhanced. Because the current meta-analysis includes studies with both kinds of attention manipulations, ABC renewal can be expected to be numerically smaller than ABA renewal when the data are aggregated across studies and experimental conditions.

With regard to ABA renewal, our representational theory predicts that effect sizes should be reduced when attention to the second learning context is reduced. However, effect sizes should be unaffected by enhanced attention to the first learning context. Overall, effect sizes of ABA renewal significantly differed as a function of attention, Q_B (2) = 7.59, p = .02. Consistent with the predictions of our theory, effect sizes of ABA renewal tended to be smaller when attention to the second learning context was reduced than when attention was not manipulated, $Q(1) = 2.97$, $p = .08$. Interestingly, effect sizes were significantly larger when attention to the first learning context was enhanced than when attention was not manipulated, $Q(1) = 5.39$, $p = .02$, and attention to the second learning context was reduced, $Q(1) = 8.87$, $p = .003$. The meta-analytic effect size of ABA renewal was significantly larger than zero when attention to the context was not manipulated, $Z = 7.50$, $p < .001$, and when attention to the first learning context was enhanced, $Z = 4.01$, $p < .001$, but it did not differ from zero when attention to the second learning context was reduced, $Z = 0.46$, $p = .65$.

With regard to ABC renewal, our representational theory predicts that effect sizes should be reduced when (1) attention to the first learning context is enhanced or (2) attention to the second learning context is reduced. Counter to these predictions, effect sizes for ABC renewal did not significantly differ as a function of attention, Q_B (2) = 2.49, $p = .29$. When attention to context was not manipulated, the meta-analytic effect size was significantly larger than zero, $Z = 5.64$, $p < .001$. However, ABC renewal was still significantly larger than zero when attention to the first learning context was enhanced, $Z = 2.32$, $p = .02$. If anything, ABC renewal was enhanced rather than reduced compared to default conditions (see Figure 3). When attention to the second learning context was reduced, ABC renewal was not significantly different from zero $Z = 0.70$, $p = .48$. Nevertheless, there were no significant differences between the three attention conditions when they were directly compared to each other, all $Qs(1) < 2$, all $ps > .20$.

Valence order. An important question is whether renewal effects depend on the order in which positive and negative information had been acquired. To address this question, we tested whether effect sizes of ABA and ABC renewal differ as a function of valence order. Because the current dataset included several studies that used within-subjects designs to manipulate valence order (20 out of 54 samples for ABA renewal, 19 out of 49 samples for ABC renewal), we had to ensure that the independent effect size assumption was not violated (see Lipsey & Wilson, 2001). Toward this end, we randomly selected the effect sizes of the positive-negative order from half of the samples using a within-subjects design, and the effect sizes of the negative-positive order from the remaining samples. To reduce the influence of random sampling error, we used a bootstrapping procedure. For ABA renewal, we randomly sampled 10 of the 20 positive-negative effect sizes (with replacements) using 100 iterations and then calculated the mean effect size resulting from this procedure. This procedure was repeated 10 times, which resulted in 10 individual effect sizes corresponding to the number of samples. The same procedure was applied to the effect sizes of the negative-positive order. For the 19 samples that used a within-subjects design to investigate ABC renewal, we utilized the same bootstrapping procedure, using 10 samples to calculate effect sizes for the positive-negative order and the remaining 9 samples to calculate effect sizes for the negative-positive order. Results showed that valence order did not moderate ABA renewal, $Q_B(1) = 0.07$, $p = .79$. However, there was a marginally significant effect for ABC renewal, $Q_B(1) = 3.20$, $p = .07$, indicating the ABC renewal tended to be more pronounced when initial positive information was challenged by subsequent negative information than when initial negative information was challenged by subsequent positive information. Nevertheless, effect sizes for ABC renewal were statistically significant in both valence order conditions, $Z = 5.89$, $p < .001$ for positive-negative and $Z = 3.72$, $p = .002$ for negative-positive.[2]

Context-valence contingencies. According to our representational theory, contextual cues constrain the activation of evaluative information about the target object, such that their presence versus absence moderates the evaluative response that is elicited by the object (*occasion setting*). However, to the extent that there is a contingency between context and valence, it is possible that the contexts themselves become associated with the valence of the evaluative experience during encoding (*evaluative binding*), and thus directly elicit a corresponding evaluative response (see Vervliet et al., 2013). In the current meta-analysis, we tested the contribution of occasion setting and evaluative binding by comparing effect sizes in studies that used a between-subjects manipulation of valence order to the effect sizes in studies that used a within-subjects manipulation of valence order. Whereas between-subjects manipulations of valence order involve a contingency between context and valence, there is no contingency between context and valence in studies using a within-subject manipulation. Thus, whereas occasion setting may contribute to renewal effects in either type of design, a potential contribution of evaluative binding is limited to between-subjects designs.

The moderator analysis indicated that ABA renewal was significantly larger in studies using a between-subjects design than studies using a within-subjects design, Q_B (1)

---

[2] Further analysis revealed that this effect was primarily driven by differences in ABC renewal in studies using within-subjects manipulations of valence order. Although the results of this analysis should be treated with caution because it violates the independent effect size assumption (see Lipsey & Wilson, 2001), a paired-sample t-test

revealed a statistically significant difference between the two valence order conditions for studies using a within-subjects manipulation, $t(23) = 2.45$, $p = .02$; the effect of valence order was not significant among between-subjects studies Q_B (1) = 0.10, $p = .75$.

= 6.03, $p = .01$. However, both between-subjects designs, $Z = 7.36$, $p < .001$, as well as within-subjects designs, $Z = 3.93$, $p < .001$, yielded effects that were significantly larger than zero. A similar pattern occurred for ABC renewal, which tended to show larger effect sizes in between-subjects designs than within-subjects designs, $Q_B(1) = 3.10$, $p = .08$. Yet, again, both between-subjects designs, $Z = 5.35$, $p < .001$, as well as within-subjects designs, $Z = 3.11$, $p = .002$, yielded effects that were significantly larger than zero. Together, these results indicate that contextual cues constrain the activation of evaluative information about the target object, such that their presence versus absence moderates the evaluative response that is elicited by the object (*occasion setting*). However, if there is a contingency between context and valence, the contexts themselves can become directly associated with the valence of the counterattitudinal experience (*evaluative binding*), thereby eliciting a corresponding evaluative response.

Type of measure. The datasets of the current meta-analysis included four different measures of evaluative responses: the AMP (Payne at el. 2005), the EPT (Fazio et al., 1995), the SET (Ranganath et al., 2008), and a modified variant of the AMP with an SOA of 1000 ms instead of 200 ms. Seven samples completed two measures: three samples completed the AMP and the SET; two samples completed the AMP and a modified AMP variant with a longer SOA; and two samples completed the AMP and the EPT. For the samples that included two measures, we selected one effect size for the moderator analysis to avoid a violation of the independent effect size assumption (see Lipsey & Wilson, 2001). In such cases, we chose the effect sizes for the tasks other than the AMP, because the majority of the samples completed the AMP (44 out of 54 studies for ABA renewal, 40 out of 49 studies for ABC renewal).[3]

For ABA renewal, there was no significant effect of type of measure, $Q_B(3) = 5.96$, $p = .11$. However, a closer inspection of the data indicated that only the AMP, $Z = 6.99$, $p < .001$, and the SET, $Z = 4.27$, $p < .001$, produced a significant effect size that was different from zero. The effects for the EPT, $Z = 1.59$, $p = .11$, and the AMP variant with a longer SOA, $Z = -0.11$, $p = .90$, failed to reach statistical significance. Nevertheless, effect sizes did not significantly differ for the four types of measures when they were directly compared to each other, all $Qs(1) < 2.2$, all $ps > .14$.

For ABC renewal, there was also no significant effect of type of measure, $Q_B(3) = 5.23$, $p = .16$. Yet, again, only the AMP, $Z = 5.67$, $p < .001$, and the SET, $Z = 2.83$, $p < .01$, produced a significant effect size that was different from zero. There was no significant effect for the EPT, $Z = 0.48$, $p = .64$, and the AMP variant with a longer SOA, $Z = 0.01$, $p = .99$. Nevertheless, effect sizes did not significantly

differ for the four types of measures when they were directly compared to each other, all $Qs(1) < 3$, all $ps > .10$.

Repeated measurement. In most studies, evaluative responses were measured only after the second learning block. Yet, in some studies, evaluative responses were additionally measured after the first learning block (14 out of 54 samples for ABA renewal; 12 out of 49 samples for ABC renewal). Our meta-analytic results suggest that repeated measurement did not influence the size of ABA renewal, $Q_B(1) = 0.11$, $p = .74$, or ABC renewal, $Q_B(1) = 0.36$, $p = .55$.

Sample characteristics. The available dataset included studies from five countries: United States, Canada, Singapore, Italy, and Belgium. Because the studies conducted in Italy and Belgium had relatively small samples and we did not expect any significant differences among European countries, Italy and Belgium were combined in a single category *Europe* for the current analysis.[4] The meta-analytical results yielded a significant sample effect for ABA renewal, $Q_B(3) = 18.42$, $p < .001$. Overall, effect sizes of ABA renewal were significantly larger for participants from the United States compared to participants from the other three regions, all $Qs(1) > 6$, all $ps < .01$. There were no significant differences in effect sizes for participants from Canada, Singapore, and Europe, all $Qs(1) < 1$, all $ps > .30$. Effect sizes were significantly above zero for studies conducted in the United States, $Z = 7.96$, $p < .001$, and Canada, $Z = 3.92$, $p < .001$, and marginally significant for studies conducted in Europe, $Z = 1.69$, $p = .09$. Effect sizes were not statistically significant for studies conducted in Singapore, $Z = 0.62$, $p = .54$.

A similar pattern emerged for ABC renewal, which also showed a significant sample effect, $Q_B(3) = 14.93$, $p = .002$. Replicating the pattern for ABA renewal, effect sizes of ABC renewal were significantly larger for participants from the United States compared to participants from the other three regions, all $Qs(1) > 6$, all $ps < .02$. Moreover, there were no significant differences in effect sizes for participants from Canada, Singapore, and Europe, all $Qs(1) < 1$, all $ps > .50$. Effect sizes of ABC renewal were significantly above zero for studies conducted in the United States, $Z = 6.38$, $p < .001$, and Canada, $Z = 2.92$, $p = .004$, but not for studies conducted in Europe, $Z = 0.67$, $p = .50$, and Singapore, $Z = 0.71$, $p = .47$.

Discussion

The main goal of the current meta-analysis was to provide a more accurate estimate of the average effect sizes of ABA and ABC renewal in Gawronski et al.'s (2010) evaluative learning paradigm. Considering all published and unpublished studies from our research groups regardless of whether they did or did not replicate our initial findings, the meta-analysis revealed a mean effect size of $d = 0.249$ for ABA renewal, which counts as a small effect

---

[3] We also conducted additional analyses that chose the AMP instead of the other three measures. This selection strategy decreased the number of effect sizes for the SET, the EPT, and the AMP variant with a longer SOA, thereby increasing their associated confidence intervals. Nevertheless, all

results remained the same when the effects sizes of the AMP were selected.

[4] All of the reported results replicate when Belgium and Italy are coded as two different countries.

according to Cohen (1988). The mean effect size for ABC renewal fell slightly below Cohen's criterion for a small effect with a value of $d = 0.174$. Nevertheless, both effects were significantly different from zero, indicating the reliability of the two effects. Together, these findings support our representational theory, which predicts that automatic evaluations tend to reflect the valence of counterattitudinal information only in the context in which this information had been learned. In contrast, automatic evaluations should reflect the valence of initial attitudinal information in any other context, be it the context in which the initial attitudinal information had been acquired or a novel context in which the target object had not been encountered before.

Moderators

In addition to providing a more accurate estimate of the average effect sizes of ABA and ABC renewal, our meta-analysis revealed systematic differences in effect sizes across studies. Although effect sizes were not significantly moderated by the type of evaluation measure and repeated measurement of evaluative responses, there were systematic differences as a function of attention to the context during learning, the order in which positive and negative information had been acquired, the presence of context-valence contingencies during learning, and sample characteristics.

Attention to context during learning. According to our representational theory, enhanced attention to contextual cues during the encoding of initial attitudinal information should reduce ABC renewal, but not ABA renewal. Moreover, both types of renewal effects should be eliminated when attention to contextual cues during the encoding of subsequent counterattitudinal information is reduced. Consistent with these predictions, we found that ABA renewal was attenuated when attention to the second learning context was experimentally reduced, but not when attention to the first learning context was experimentally enhanced. However, the meta-analytic results failed to support our predictions for ABC renewal, which did not show any effect of attention. Although effect sizes did not differ significantly from zero when attention to the second learning context was reduced, enhanced attention to the first learning context failed to reduce ABC renewal. Together, these results provide only partial support for our representational theory. Although the current findings confirmed our predictions for the emergence of ABA renewal, the insensitivity of ABC renewal to attentional manipulations require further theoretical refinements to fully account for the obtained pattern of results.

A potential explanation for the unexpected pattern of results is that the manipulations designed to enhance attention to the first learning context were simply ineffective. Although this argument accounts for the ineffectiveness of these manipulations in *reducing* the two

kinds of renewal effects, it is inconsistent with the finding that effects sizes for ABA renewal were significantly *larger* compared to default conditions when attention to the first learning context was enhanced. This finding clearly indicates that our attention manipulations were effective. An alternative interpretation is that our manipulations enhanced attention to the context of both initial attitudinal and subsequent counterattitudinal information. Together with the attentional effect of expectancy violation (Roese & Sherman, 2007), a joint impact on both attitudinal and counterattitudinal learning implies that Context A should be less strongly represented than Context B, although both contexts may be included in contextualized representations. As a result, manipulations designed to increase attention to the context of initial attitudinal information may be less powerful in reducing ABC renewal than manipulations designed to reduce attention to the context of counterattitudinal information. Although these assumptions are clearly speculative, future research may help to further clarify the contribution of attentional processes to the emergence of ABC renewal.

Valence order. Drawing on the large body of evidence for negativity bias (for reviews, see Baumeister et al., 2001; Cacioppo & Berntson, 1994; Skowronski & Carlston, 1989), one could argue that counterattitudinal negative information may be more effective in influencing automatic evaluations across contexts than counterattitudinal positive information. These considerations suggest that renewal effects may be more pronounced when initial negative information is challenged by positive information than when initial positive information is challenged by negative information. This hypothesis was not supported in the current meta-analysis, which showed comparable effect sizes for ABA renewal regardless of valence order. In fact, ABC renewal showed a marginally significant tendency for the opposite pattern, such that ABC renewal tended to be more pronounced when initial positive information was challenged by subsequent negative information than when initial negative information was challenged by subsequent positive information. Drawing on the assumptions of our representational theory, a potential explanation for this asymmetry is that negative information captures more attention than positive information (e.g., Pratto & John, 1991), which may also enhance attention to its momentary context.[5] As a result, contextual cues may be integrated into the representation of both attitudinal and counterattitudinal information when initial negative information is challenged by subsequent positive information. According to our theory, such differences in attention should lead to weaker ABC renewal, but not ABA renewal, when initial negative information is challenged by subsequent positive information compared to conditions when initial positive information is challenged by

---

[5] This attentional asymmetry may be limited to conditions in which negative information about one target object occurs in the context of positive information about another target object, which would explain

why the effect of valence order was limited to studies using a within-subjects manipulation of valence order (see Footnote 1).

subsequent negative information. Although this interpretation is clearly speculative at this point, future research may help to clarify the processes underlying valence asymmetries in ABC renewal.

Context-valence contingency. Counter to the strong emphasis on occasion setting in the initial presentation of our theory (Gawronski et al., 2010), our meta-analysis suggests that renewal effects can be the result of two distinct mechanisms. Specifically, our results indicate that both ABA and ABC renewal are more pronounced when there is a contingency between context and valence during encoding than when there is no such contingency. In the current meta-analysis, this difference was reflected in the finding that both kinds of renewal effects were larger when valence order was manipulated on a between-subjects basis than when valence order was manipulated on a within-subjects basis. Whereas between-subjects designs involve a perfect contingency between context and valence, there is no such contingency in within-subjects designs. Although we did find significant renewal effects for both experimental designs, renewal effects were more pronounced in studies that used between-subjects designs compared with studies that used within-subjects designs. The finding that contextual renewal occurs in the absence of a contingency between context and valence supports the notion of occasion setting (see Gawronski et al., 2014); the finding that contextual renewal is more pronounced when there is a contingency between context and valence supports the notion of evaluative binding (see Vervliet et al., 2013). Occasion setting implies that contextual cues constrain the activation of evaluative information about the target object, such that their presence versus absence moderates the evaluative response that is elicited by the object. Evaluative binding implies that the contexts themselves become directly associated with the valence of counterattitudinal experience during encoding, and thus directly elicit a corresponding evaluative response. Thus, deviating from the emphasis on occasion setting in our representational theory (Gawronski et al., 2010), our meta-analytic findings indicate that both occasion setting and evaluative binding can jointly contribute to contextual renewal effects.

Type of measure. Expanding on the exclusive use of the AMP in our initial studies (Gawronski et al., 2010), the datasets of the current meta-analysis allowed us to compare renewal effects across four different measures of evaluation: the AMP (Payne et al., 2005), the EPT (Fazio et al., 1995), the SET (Ranganath et al., 2008), and an AMP variant with a longer SOA (Hofmann et al., 2009). Although measurement type did not produce a statistically significant moderation overall, effect sizes for ABA and ABC renewal reached statistical significance only for the AMP and the SET, but not for the EPT and the AMP variant with a longer SOA.

The converging effects sizes on the AMP and the SET suggest that renewal effects might be independent of intention as a central feature of automaticity. Whereas the AMP measures unintentional effects of prime stimuli on evaluative judgments, the SET involves intentional evaluations under time pressure. Moreover, the absence of renewal effects on the AMP variant with a longer SOA is consistent with earlier speculations by Gawronski et al. (2010), who argued that contextual cues may function like retrieval cues by determining which information comes to mind most rapidly upon encountering the target object (see also Gawronski & Cesario, 2013). With increasing delays, other information may be retrieved from memory, including information that has been learned in other contexts (cf. Cunningham, Zelazo, Packer, & Van Bavel, 2007; Wojnowicz, Ferguson, Dale, & Spivey, 2009). As a result, renewal effects may decrease as a function of time, which is consistent with the obtained absence of renewal effects on the AMP variant with a longer SOA.

Although these findings are consistent with the assumptions of our representational theory, the situation is more ambiguous for the EPT, which failed to produce significant renewal effects in the current meta-analysis. On the one hand, one could argue that the reduced effect sizes on the EPT are due to its lower reliability compared to the other three measures (see Gawronski & De Houwer, 2014), which reduces the likelihood of replicating actually existing effects with the EPT (LeBel & Paunonen, 2011). On the other hand, the absence of contextual renewal on the EPT may indicate that renewal effects involve a more substantial contribution of non-automatic processes than postulated in our initial article (Gawronski et al., 2010). In that case, the significant effects on the AMP might reflect intentional rather than unintentional evaluations (i.e., intentional use of the primes to evaluate the targets), which would be consistent with recent concerns about the role of intentional processes in the AMP (Bar-Anan & Nosek, 2012). Although the validity of these concerns has been questioned in several recent studies (e.g., Gawronski & Ye, 2015; Payne et al., 2013), future research may help to clarify the role of intentional and unintentional processes in contextual renewal by using measures that are more reliable than the EPT and less susceptible to intentional influences than the AMP.

Repeated measurement. Although most of the studies included in the current meta-analysis measured evaluative responses only after the second learning block, there were some studies that measured evaluative responses after both learning blocks. Because delays between initial attitudinal and subsequent counterattitudinal learning may influence renewal effects through processes of memory consolidation and the partitioning of evaluative information, we were interested in whether renewal effects are moderated by repeated measurement. Our results suggest that neither ABA renewal nor ABC renewal are influenced by prior measurement of evaluative responses, and thus the delays implied by prior assessments.

Sample characteristics. A final question concerned potential effects of sample characteristics, particularly the role of culture. Based on Gawronski et al.'s (2010) findings

regarding the role of attentional processes, Ye and Gawronski (in press) discussed two potential patterns of cultural differences in contextual renewal. First, drawing on evidence that individuals from East Asian cultures tend to pay more attention to contextual information than individuals from Western cultures (e.g., Chua et al., 2005; Masuda & Nisbett, 2001), one could argue that East Asians are less likely than Westerners to show ABC renewal while showing similar levels of ABA renewal. Second, drawing on evidence that individuals from East Asian cultures tend to have a higher tolerance for inconsistency than individuals from Western cultures (e.g., Choi & Nisbett, 2000; Peng & Nisbett, 1999), one could argue that East Asians are less likely than Westerners to show either type of renewal effect. Consistent with the latter hypothesis, we found significant effects of ABA and ABC renewal in samples from the United States and Canada, but not in samples from Singapore. However, counter to an interpretation of this pattern in terms of cultural differences in the tolerance for inconsistency, there were no significant differences between samples from Canada, Europe, and Singapore. Because thinking styles in Europe and Canada tend to be more similar to those in the United States than in Singapore, a simple East-West dichotomy fails to account for the obtained pattern of results. Although the available samples from Singapore and Europe were relatively small compared to the ones from the United States and Canada, future research may help to clarify the role of cultural differences in the emergence of renewal effects.

Limitations

The main goal of the current meta-analysis was to reassess the reliability of Gawronski et al.'s (2010) findings in order to reevaluate the validity of our representational theory. Toward this end, the current meta-analysis was based on published and unpublished studies from our own research groups regardless of whether they did or did not obtain significant renewal effects. Thus, the data base of our meta-analysis is limited, in that it did not include studies from other labs using the same paradigm.[6] It also did not include studies that investigated contextual renewal in other paradigms, such as renewal effects after extinction of conditioned fear responses (for a review, see Vervliet et al., 2013). Although a broader meta-analysis would be helpful to provide deeper insights into the size and the conditions of contextual renewal, our primary concern was to reassess the reliability of our earlier findings in light of several unsuccessful replications in our own labs; it was not meant to assess the reliability of contextual renewal for other kinds of learning. Our meta-analytic findings indicate that the results reported in our earlier article are indeed reliable. However, the average size of ABA and ABC renewal seems to be much smaller compared to what might be inferred from the data reported by Gawronski et al.

(2010). As such, the current meta-analysis provides valuable information for power analyses in future research using Gawronski et al.'s evaluative learning paradigm (Cohen, 1988). In addition, our findings help to identify gaps in the theoretical explanation of renewal effects. Although our representational theory fared reasonably well with regard to the prediction of moderator effects, our meta-analysis identified several other effects that were not expected on the basis of our theory. One example is the moderating role of context-valence contingencies, which requires theoretical refinements in the assumptions about the underlying mental structures (cf. Bouton, 2010). Another example is the unqualified occurrence of ABC renewal under conditions where our theory would predict a reduction. Although a broader meta-analysis would be helpful to provide deeper insights into the boundary conditions of contextual renewal, the consistent use of the same learning paradigm allows us to draw stronger conclusions about the impact of the identified parameters, because they are less likely to be confounded with other procedural parameters.

Nevertheless, it is important to note that the identification of moderator effects in meta-analyses is based on correlational data rather than random assignment to experimental conditions. Hence, it is possible that at least some of the identified effects are driven by third variables that happen to be confounded in the current data. For example, to the extent that the levels of one moderator tend to overlap with the levels of another one, it is possible that a causal effect of one moderator leads to a spurious effect of the other. Although meta-analyses offer valuable information about the average size of a given effect and its moderators, they do not provide irrevocable truths that are immune to future experimental evidence. In line with this concern, we interpret our meta-analysis as an important correction of potentially inadequate conclusions about the average effect sizes of ABA and ABC renewal as well as their moderators. Thus, future experimental research is needed to provide further evidence for the reliability of the obtained moderator effects and to gain deeper insights into their underlying mechanisms.

Conclusion

The main goal of the current meta-analysis was to provide a more accurate estimate of the average effect sizes of ABA and ABC renewal in Gawronski et al.'s (2010) evaluative learning paradigm. Although our meta-analytic findings support the reliability of contextual renewal in automatic evaluation, the obtained effect sizes turned out to be much smaller than what might be expected on the basis of our initial findings. A comprehensive analysis of all published and unpublished studies from our labs revealed an average effect size of $d = 0.249$ for ABA renewal and an average effect size of $d = 0.174$ for ABC renewal, both of

---

[6] We are aware of only one other study that investigated contextual renewal in Gawronski et al.'s (2010) evaluative learning paradigm. Comparing renewal effects on speeded and non-speeded self-reports, Klein and Ratliff (2014) found significant effects of ABA and ABC

renewal on both measures when initial positive information was challenged by subsequent negative information. The study did not include conditions in which initial negative information was challenged by subsequent positive information.

which were significantly different from zero. Moreover, renewal effects were moderated by attention to the context during learning, the order in which positive and negative information had been acquired, the presence of context-valence contingencies during learning, and the country in which the study was conducted. Although some of the obtained effects are consistent with the assumptions of our representational theory, others require theoretical refinements and future research to provide deeper insights into the processes and representations underlying contextual renewal.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Bar-Anan, Y., & Nosek, B. A. (2012). Reporting intentional rating of the primes predicts priming effects in the Affective Misattribution Procedure. *Personality and Social Psychology Bulletin, 38,* 1194-1208.

Baumeister, R. F., Bratslavsky, E., Finkenauer, E., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5,* 323-370.

Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science, 3,* 49-53.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning and Memory, 11,* 485-494.

Bouton, M. E. (2010). The multiple forms of "context" in associative learning theory. In B. Mesquita, L. Feldman-Barrett, & E. R. Smith (Eds.), *The mind in context* (pp. 233-258). New York, NY: Guilford Press.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10*, 445-466.

Bouton, M. E., & Brooks, D. C. (1993). Time and context effects on performance in a Pavlovian discrimination reversal. *Journal of Experimental Psychology: Animal Behavior Processes, 19,* 165-179.

Bouton, M. E., & Peck, C. A. (1989). Context effects on conditioning, extinction, and reinstatement in an appetitive conditioning preparation. *Animal Learning and Behavior, 17,* 188-198.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115,* 401-423.

*Carraro, L., Castelli, L., & Piccione, E., & Gawronski, B. (2011). *Political ideology and contextual renewal.* Unpublished raw data.

*Carraro, L., Castelli, L., Tosatto, F., & Gawronski, B. (2013). *Construal level and contextual renewal.* Unpublished raw data.

Choi, I., & Nisbett, R. E. (2000). Cultural psychology of surprise: Holistic theories and recognition of contradiction. *Journal of Personality and Social Psychology, 79,* 890-905.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences, USA, 102,* 12629-12633.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.

Cummings, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis.* New York: Taylor & Francis.

Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition, 25,* 736-760.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127,* 853–869.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, 25, 603-637.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013-1027.

*Gawronski, B. (2013). *Individual differences in contextual renewal (Study 1).* Unpublished raw data.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132,* 692-731.

*Gawronski, B., & Brannon, S. M. (2014). *Individual differences in contextual renewal (Study 2).* Unpublished raw data.

Gawronski, B., & Cesario, J. (2013). Of mice and men: What animal research can tell us about context effects on automatic responses in humans. *Personality and Social Psychology Review, 17,* 187-215.

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2nd edition, pp. 283-310). New York, NY: Cambridge University Press.

*Gawronski, B., & Rydell, R. J. (2010). *Automatic and controlled processes in contextual renewal: Intentionality (Study 1).* Unpublished raw data.

*Gawronski, B., & Rydell, R. J. (2010). *Automatic and controlled processes in contextual renewal: Time (Study 1).* Unpublished raw data.

*Gawronski, B., & Rydell, R. J. (2010). *Automatic and controlled processes in contextual renewal: Intentionality (Study 2).* Unpublished raw data.

*Gawronski, B., & Rydell, R. J. (2011). *Automatic and controlled processes in contextual renewal: Intentionality (Study 3).* Unpublished raw data.

*Gawronski, B., & Rydell, R. J. (2012). *Automatic and controlled processes in contextual renewal: Time (Study 2)*. Unpublished raw data.

*Gawronski, B., & Rydell, R. J., & De Houwer, J. (2011). *Occasion setting in contextual renewal (Study 1)*. Unpublished raw data.

*Gawronski, B., & Rydell, R. J., & De Houwer, J. (2012). *Occasion setting in contextual renewal (Study 2)*. Unpublished raw data.

*Gawronski, B., & Rydell, R. J., & De Houwer, J. (2012). *Occasion setting in contextual renewal (Study 3)*. Unpublished raw data.

*Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2009). *Contextual renewal after counterattitudinal learning in single versus multiple contexts*. Unpublished raw data.

*Gawronski, B., Rydell, R. J., Vervliet, B., & De Houwer, J. (2010). Generalization versus contextualization in automatic evaluation. *Journal of Experimental Psychology: General, 139,* 683-701.

Gawronski, B., & Sritharan, R. (2010). Formation, change, and contextualization of mental associations: Determinants and principles of variations in implicit measures. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 216-240). New York, NY: Guilford Press.

Gawronski, B., & Ye, Y. (2015). Prevention of intention invention in the affect misattribution procedure. *Social Psychological and Personality Science, 6,* 101-108.

*Gawronski, B., Ye, Y., Rydell, R. J., & De Houwer, J. (2011). *Perceptual versus conceptual processing in contextual renewal*. Unpublished raw data.

*Gawronski, B., Ye, Y., Rydell, R. J., & De Houwer, J. (2014). Formation, representation, and activation of contextualized attitudes. *Journal of Experimental Social Psychology, 54,* 188-203.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107-128.

Hofmann, W., Friese, M., & Roefs, A. (2009). Three ways to resist temptation: The independent contributions of executive attention, inhibitory control, and affect regulation to the impulse control of eating behavior. *Journal of Experimental Social Psychology, 45,* 431-435.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd Edition). Thousand Oaks, CAS: Sage.

Klein, R. A., & Ratliff, K. A (2014, February). *Context effects on implicit and explicit evaluation*. Poster presented at the 15th Annual Meeting of the Society for Personality and Social Psychology, Austin, TX.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: Impact of unreliability on the replicability of experimental findings involving implicit measures. *Personality and Social Psychology Bulletin, 37,* 570-583.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 32,* 431-440.

Lipsey, M. W., & Wilson, D. S. (2001). *Practical meta-analysis. Volume 49*. Thousand Oaks, CA: Sage.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology* 81, 992-934.

Murayama, K., Pekrun, R., & Fiedler, K. (2014). Research practices that can prevent an inflation of false-positive rates. *Personality and Social Psychology Review, 18,* 107-118.

Payne, B. K., Brown-Iannuzzi, J., Burkley, M., Arbuckle, N. L., Cooley, E., Cameron, C. D., & Lundberg, K. B. (2013). Intention invention and the Affect Misattribution Procedure: Reply to Bar-Anan and Nosek (2012). *Personality and Social Psychology Bulletin, 39,* 375-386.

Payne, B. K., Cheng, S. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277-293.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist, 54,* 741-754.

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST Model. *Journal of Personality and Social Psychology, 90,* 21-41.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-391.

Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement methods. *Journal of Experimental Social Psychology, 44,* 386-396.

Roese, N. J., & Sherman, J. W. (2007). Expectancies. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (2nd Ed., pp. 91-115). New York, NY: Guilford Press.

*Rydell, R. J., & Gawronski, B. (2006). *Formation of context-dependent automatic attitudes*. Unpublished raw data.

*Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition and Emotion, 23,* 1118-1152.

Schmajuk, N. A., & Holland, P. C. (1998). *Occasion setting: Associative learning and cognition in animals*.

Washington, DC: American Psychological Association.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105,* 131-142.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70,* 893-912.

Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic? *Perspectives on Psychological Science, 9,* 305-318.

*Struyf, D., & Vervliet, B. (2012). *Memory consolidation and contextual renewal*. Unpublished raw data.

Vervliet, B., Baeyens, F., Van den Bergh, O., & Hermans, D. (2013). Extinction, generalization, and return of fear: A critical review of renewal research in humans. *Biological Psychology, 92,* 51-58.

Wilson, T. D., Lindsey, S. & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101-126.

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science, 20,* 1428-1435.

Ye, Y., & Gawronski, B. (in press). Contextualization of mental representations and evaluative responses: A theory-based analysis of cultural differences. In J. Spencer-Rodgers, & K. Peng (Eds.), *The psychological and cultural foundations of East Asian cognition: Contradiction, change, and holism*. New York, NY: Oxford University Press.

*Ye, Y., Tong, Y.-Y., Chiu, C.-Y., & Gawronski, B. (2012). *Culture and contextual renewal (Study 1)*. Unpublished raw data.

*Ye, Y., Tong, Y.-Y., Chiu, C.-Y., & Gawronski, B. (2012). *Culture and contextual renewal (Study 2)*. Unpublished raw data.

*Ye, Y., Tong, Y.-Y., Chiu, C.-Y., & Gawronski, B. (2014). *Culture and contextual renewal (Study 3)*. Unpublished raw data.

Author Note

Bertram Gawronski , Department of Psychology, University of Texas at Austin, USA; Xiaoqing Hu, Department of Psychology, University of Texas at Austin, USA; Robert J. Rydell, Department of Psychological and Brain Sciences, Indiana University, USA; Bram Vervliet, Department of Psychology, University of Leuven, Belgium; Jan De Houwer, Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium.

Correspondence concerning this article should be sent to: Bertram Gawronski, Department of Psychology, University of Texas at Austin, 108 E Dean Keeton A8000, Austin, TX 78712-1043, USA, Email: gawronski@utexas.edu

Table 1. Mean weighted effect sizes (Cohen's *d*), standard errors, and 95% confidence intervals for ABA and ABC renewal as a function of procedural moderators.

	ABA Renewal			ABC Renewal		
	ES *d*	*SE*	95% CI	ES *d*	*SE*	95% CI
Attention during Learning						
default without manipulation	0.234	0.031	0.173 – 0.295	0.159	0.028	0.104 – 0.215
enhanced first context	0.574	0.143	0.293 – 0.856	0.296	0.127	0.046 – 0.546
reduced second context	0.048	0.103	-0.155 – 0.250	0.060	0.086	-0.108 – 0.228
Valence Order						
positive-negative	0.242	0.042	0.161 – 0.324	0.234	0.040	0.156 – 0.312
negative-positive	0.227	0.043	0.143 – 0.311	0.137	0.037	0.065 – 0.210
Context-Valence Contingency						
present	0.305	0.042	0.224 – 0.387	0.205	0.038	0.130 – 0.280
absent	0.162	0.041	0.081 – 0.243	0.112	0.037	0.039 – 0.186
Type of Measure						
AMP	0.261	0.037	0.188 – 0.334	0.199	0.035	0.130 – 0.268
EPT	0.137	0.086	-0.032 – 0.306	0.038	0.081	-0.121 – 0.198
SET	0.266	0.062	0.144 – 0.389	0.160	0.057	0.049 – 0.271
AMP with long SOA	-0.014	0.125	-0.259 – 0.232	0.001	0.111	-0.216 – 0.218
Repeated Measurement						
yes	0.252	0.042	0.132 – 0.371	0.185	0.056	0.077 – 0.294
no	0.229	0.041	0.163 – 0.295	0.148	0.030	0.090 – 0.206
Sample						
United States	0.412	0.052	0.311 – 0.513	0.313	0.049	0.217 – 0.410
Canada	0.165	0.042	0.082 – 0.247	0.108	0.037	0.035 – 0.180
Singapore	0.062	0.100	-0.135 – 0.256	0.061	0.086	-0.107 – 0.229
Europe	0.148	0.088	-0.023 – 0.320	0.053	0.079	-0.102 – 0.207

Note. ES *d* = Effect Size Cohen's *d*; *SE* = Standard Error; CI = Confidence Interval; AMP = Affect Misattribution Procedure; EPT = Evaluative Priming Task; SET = Speeded Evaluation Task; SOA = Stimulus Onset Asynchrony.

Figure 1. Predicted pattern of evaluations as a function of valence order (positive-negative vs. negative-positive) and context of evaluation. Context A refers to the context of initial attitudinal learning; Context B refers to the context of subsequent counterattitudinal learning; Context C refers to a novel context in which the target object has not been encountered before. Values above zero indicate positive evaluations; values below zero indicate negative evaluations.
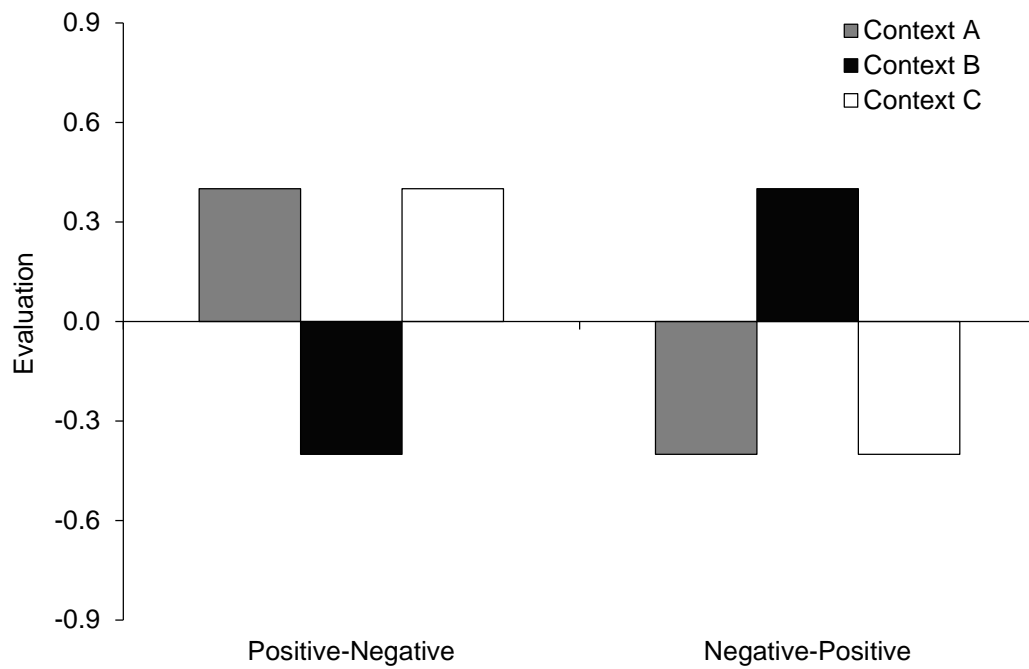
Figure 2. Mean weighted effect sizes (Cohen's *d*) of ABA renewal as a function of procedural moderators. Error bars depict 95% confidence intervals.
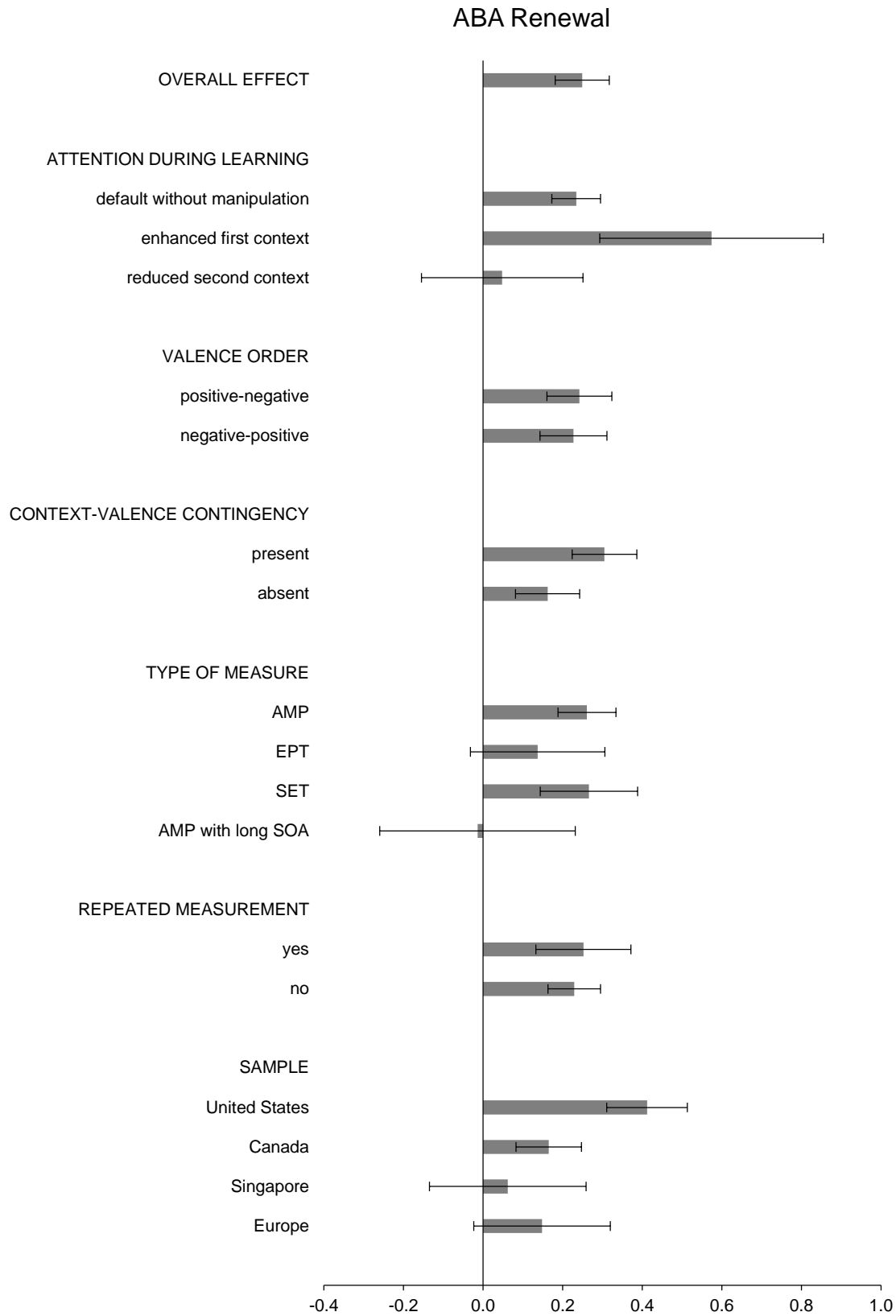
Figure 3. Mean weighted effect sizes (Cohen's *d*) of ABC renewal as a function of procedural moderators. Error bars depict 95% confidence intervals.

ABC Renewal