Contextualized Attitude Change

Bertram Gawronski

University of Texas at Austin

Robert J. Rydell

Indiana University

Jan De Houwer

Ghent University

Skylar M. Brannon

University of Texas at Austin

Yang Ye

Ghent University

Bram Vervliet

University of Leuven

Xiaoqing Hu

University of Hong Kong

Abstract

The current chapter reviews the findings of an ongoing research program suggesting that changes in attitudes can be limited to the context in which counterattitudinal information was learned. The reviewed findings indicate that, although counterattitudinal information may effectively influence evaluations in the context in which this information was learned, previously formed attitudes may continue to influence evaluations in any other context. According to the representational theory of contextualized attitude change, such patterns of contextual renewal occur because exposure to expectancy-violating information enhances attention to context, which leads to an integration of the context into the representation of expectancy-violating counterattitudinal information. The chapter reviews research that (a) tested novel predictions derived from the representational theory of contextualized attitude change, (b) explored the nature of contextualized representations, and (c) investigated the boundary conditions of contextualized attitude change. Theoretical challenges, future directions, and implications for basic and applied research are discussed.

Keywords: attitude change, attitude representation, context effects, expectancy violation, renewal effects

People have a natural predisposition to appraise objects as "good" or "bad". This propensity is the hallmark of attitudes, which can be defined as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken, 1993, p. 1). Attitudes have been shown to influence behavior by shaping the construal of the current situation, eliciting spontaneous approach-avoidance tendencies, and guiding the formation of deliberate action plans (Fazio, 1990; Strack & Deutsch, 2004). Expanding on the ubiquitous effects of attitudes on behavior, a great deal of research has sought to identify the determinants of attitude change. The basic idea underlying this research is that, when attitudes change, the behaviors that are known to be influenced by attitudes will change accordingly (Chaiken, Liberman, & Eagly, 1989; Petty & Cacioppo, 1986). This idea has become a central tenet in various applied areas, including research on the effects of commercial advertisements, health communication, and interventions to reduce intergroup prejudice (Perloff, 2003; Stiff & Mongeau, 2003).

An important question in research on attitude change concerns the temporal stability of observed changes in evaluations. Whereas some factors tend to produce changes that remain temporally stable, others lead to ephemeral shifts in evaluations that dissipate over time (e.g., Chaiken, 1980; Petty, 1977). Although temporal stability has been a central issue in social psychological theories of attitude change (e.g., Petty & Cacioppo, 1986), hardly any attention has been devoted to the question of whether observed changes in evaluations generalize across contexts. The current chapter reviews the findings of an ongoing research program, suggesting that changes in evaluations can be limited to the context in which counterattitudinal information was learned. That is, although counterattitudinal information may influence evaluations of an object within the context in which this information was learned, previously formed attitudes

toward the object can continue to influence evaluations in any other context. For example, if Sally forms an initial positive impression of her new colleague Harry at work and later sees him being rude to another person at the gym, the new negative information about Harry may influence Sally's evaluative response to Harry only when she encounters him within the context of the gym. In contrast, Sally's evaluative response to Harry may continue to reflect the initial positive impression when she encounters him at work (original learning context) or at a bookstore where she has not encountered him before (completely novel context). Expanding on a review of evidence for this idea, we describe the core assumptions of a representational theory that specifies the mental processes and representations underlying contextualized attitude change. Using this theory as a framework, we review research that (a) tested novel predictions derived from the theory, (b) explored the nature of contextualized representations, and (c) investigated potential boundary conditions of contextualized attitude change. In the final sections, we discuss theoretical challenges, future directions, and implications for basic and applied research.

The Representational Theory of Contextualized Attitude Change

The two most prominent paradigms to study the formation and change of attitudes are the persuasion paradigm (see Chaiken et al., 1989; Petty & Cacioppo, 1986) and the evaluative conditioning (EC) paradigm (see De Houwer, Thomas, & Baeyens, 2001). The persuasion paradigm focuses on attitudinal effects of verbal arguments and characteristics of the message source (e.g., expertise, likeability). The EC paradigm investigates attitudinal effects of observed co-occurrences of a conditioned stimulus (CS) with a positive or negative unconditioned stimulus (US). In a broad sense, both paradigms involve the learning of new evaluative

information. To the extent that this information conflicts with a previously formed attitude, this information can be described as counterattitudinal.

The basic notion of contextualized attitude change is that mental representations of counterattitudinal information about an object often become bound to the context in which this information was learned. As a result, activation of the counterattitudinal information is limited to that context, such that previously formed attitudes toward the object continue to determine evaluations of the object in any other context. According to the representational theory of contextualized attitude change (Gawronski, Rydell, Vervliet, & De Houwer, 2010), a central determinant of such effects is the violation of attitude-related expectancies during the learning of counterattitudinal information. Although exposure to counterattitudinal information does not always produce an expectancy-violation (e.g., when a political opponent argues for a position that conflicts with one's political views), it can lead to a violation of expectancies when a previously formed attitude gives rise to the expectation of positive or negative information about an object and this expectation is violated by information of the opposite valence (e.g., when a positively evaluated person engages in negative behavior, or vice versa).

Drawing on previous research on the effects of expectancy-violation (for a review, see Roese & Sherman, 2007), the representational theory of contextualized attitude change further assumes that exposure to expectancy-violating information enhances attention to the momentary context (see Rosas & Callejas-Aguilera, 2007). As a result, incidental context cues become integrated into the representation of the newly acquired counterattitudinal information. In line with earlier theories (e.g., Petty, Tormala, Briñol, & Jarvis, 2006; Wilson, Lindsey, & Schooler, 2000), the theory further assumes that counterattitudinal information does not erase previously formed attitudes from memory. Instead, the newly formed contextualized representation is

simply added to the existing memory structures. Thus, to the extent that attention to context is typically low during the learning of initial attitudinal information (Gilbert & Malone, 1995; Jones, 1991), the mental representation of the object can be said to acquire a "dual" nature, in that it comprises (a) a *context-free representation* of initially acquired attitudinal information about the object, and (b) a *contextualized representation* of the subsequently acquired counterattitudinal information about the object.

Contextual Renewal

Drawing on the notion of pattern matching in memory retrieval (Smith, 1996), the proposed "dual" representation can lead to different evaluations of the object depending on the context in which it is encountered (see Table 1). First, if positive or negative information about an object is learned in an initial Context A, and this information is subsequently challenged by evaluatively incongruent information in another Context B, encountering the object in Context A should activate the initially formed context-free representation. As a result, evaluative responses to the object in Context A reflect the valence of the initial attitudinal information. Adopting terminology from the animal learning literature, this outcome can be described as a case of *ABA renewal* (Bouton, 2004), in that an initial response that was acquired in Context A reemerges in the original learning Context A after a different response was learned in a different Context B (e.g., Bouton & Bolles, 1979; Bouton & Peck, 1989).

Second, if positive or negative information about an object is learned in an initial Context A, and this information is subsequently challenged by evaluatively incongruent information in another Context B, encountering the object in a novel Context C should also activate the initially formed context-free representation. As a result, evaluative responses to the object in Context C should reflect the valence of the initial attitudinal information. This outcome can be described as

a case of *ABC renewal* (Bouton, 2004), in that an initial response that was acquired in Context A reemerges in a novel Context C after a different response was learned in Context B (e.g., Bouton & Bolles, 1979; Bouton & Brooks, 1993).

Third, if positive or negative information about an object is learned in an initial Context A, and this information is subsequently challenged by evaluatively incongruent information in a second Context B, encountering the object in Context B should activate the contextualized representation, leading to an evaluative response that reflects the valence of the counterattitudinal information learned in Context B. Thus, in combination with the above cases of ABA and ABC renewal, the presence versus absence of Context B moderates the evaluative response that is elicited by the object. If Context B is present, the object will elicit an evaluative response that reflects the valence of the subsequently acquired counterattitudinal information. In contrast, if Context B is absent, the object will elicit an evaluative response that reflects the valence of the initially acquired attitudinal information.

To illustrate these patterns, imagine that a person formed a favorable first impression of a new colleague at work and this impression was later challenged by negative behavior of that person at the gym. In this case, the initial positive information would be stored in a context-free representation, whereas the subsequent negative information would be stored in a contextualized representation. As a result, evaluative responses to the new colleague should be negative only in contexts that are similar to the gym. In contrast, the initial positive information should influence evaluative responses in any other context, including the work context in which the initial impression had been formed and any other context in which the new colleague had not been encountered before (e.g., at a bookstore).

In line with this idea, Rydell and Gawronski (2009) provided the first evidence for contextualized changes in social attitudes. In their research, participants were first presented with either positive or negative information about a target person against a meaningless, colored background (e.g., a yellow screen). In a second block of the learning task, participants were presented with new information about the target person that was evaluatively opposite to the information provided in the first block, and this information was presented against a different colored background (e.g., a blue screen). After each of the two learning blocks, spontaneous evaluative responses to the target person were assessed with an affective priming task (Payne, Cheng, Govorun, & Stewart, 2005) in which the target person was presented against either: (a) the background of the first learning block (Context A), (b) the background of the second learning block (Context B), or (c) a novel background that was not part of the learning task (Context C). Results showed that evaluations of the target person changed in response to the counterattitudinal information only when the target was presented against the background of the second learning block (see Figure 1). Evaluations of the target person continued to reflect the valence of the initial attitudinal information when the target was presented against the background of the first learning block (ABA renewal) or a novel background that was not part of the learning task (ABC renewal).

To assess the reliability of contextualized change effects in Rydell and Gawronski's (2009) impression formation paradigm, Gawronski, Hu, Rydell, Vervliet, and De Houwer (2015) conducted a meta-analysis that included all studies from the authors' labs regardless of whether they did or did not replicate the original pattern of results. In their meta-analysis, ABA renewal was conceptualized as the difference in the evaluation of a given target in Context A and Context B after the learning of evaluatively incongruent information about the target in Contexts A and

B; ABC renewal was conceptualized as the difference in the evaluation of a given target in

Context C and Context B after the learning of evaluatively incongruent information about the

target in Contexts A and B. In both cases, renewal effects were coded such that higher scores

reflect a stronger influence of the initial attitudinal information compared to the subsequent

counterattitudinal information in Context A versus Context B (ABA renewal) and Context C

versus Context B (ABC renewal), respectively. Although average effect sizes were relatively

small with $d = .249$ for ABA renewal (30 studies, $N = 3,142$) and $d = .174$ for ABC renewal (27

studies, $N = 2,930$), both effect sizes were significantly different from zero. Together, these

results support the hypothesis that effects of counterattitudinal information can be limited to the

context in which this information was learned, such that initially formed attitudes continue to

shape evaluations in any other context.

Attention to Context

A central assumption of the representational theory is that exposure to expectancy-

violating information enhances attention to context (e.g., enhanced attention to context when an

initial positive impression of a new colleague is challenged by negative behavior), which leads to

an integration of the context into the representation of the expectancy-violating information (see

also Bouton, 2010; Pearce, George, & Redhead, 1998; Rosas & Callejas-Aguilera, 2007). Thus,

to the extent that attention to context is typically low during the learning of initial attitudinal

information (Gilbert & Malone, 1995; Jones, 1991), initial attitudinal information should be

stored in context-free representations, whereas counterattitudinal information should be stored in

contextualized representations.

Evidence for these assumptions comes from a study by Gawronski, Ye, Rydell, and De

Houwer (2014). To investigate attention to context during the learning of attitude-congruent and

attitude-incongruent information, participants were presented with 30 behavioral statements about a target individual that were displayed one-by-one against different background colors. The initial 20 statements suggested either a positive or a negative personality; the 21st statement was used as a target statement and described a behavior that was either congruent or incongruent with the valence of the initial 20 statements. The target statement was followed by nine distracter statements that matched the valence of the initial 20 statements. Participants' task was to form an impression of the target individual. After the impression formation task, participants completed a surprise recognition test, in which they had to identify the background color against which the target statement was presented during the impression formation task. Results showed that recognition memory for the background color of the target statement was at chance level when the target statement was evaluatively congruent with the initial impression of the target. However, recognition memory was substantially higher, and significantly different from chance level, when the target statement was evaluatively incongruent with the initial impression (see Figure 2). These results support the hypothesis that exposure to counterattitudinal information enhances attention to momentarily available context cues, which in turn leads to an integration of these cues into the representation of the counterattitudinal information.

Further evidence for the hypothesized role of attentional processes comes from a series of studies by Gawronski et al. (2010) that aimed to influence participants' attention to context cues during the learning of either (a) initial attitudinal information or (b) subsequent counterattitudinal information. In the scenarios described above, attention to context cues was assumed to be low during the learning of initial attitudinal information and enhanced by exposure to subsequent counterattitudinal information. As a result, context cues should be included in the representation of counterattitudinal information but not in the representation of initial attitudinal information. If,

on the other hand, attention to context cues is high for some reason during the learning of both initial attitudinal and subsequent counterattitudinal information, then the two pieces of information should be stored in two contextualized representations: one including the initial attitudinal information and the initial Context A, and the other one including the counterattitudinal information and the subsequent Context B. In such cases, encountering the object in a novel Context C 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. In other words, evaluations of the object in a novel Context C should be neutral (or ambivalent) instead of reflecting the valence of the initially acquired attitudinal information. Thus, a central prediction of the representational theory is that ABC renewal should be reduced if attention to context is enhanced during the learning of initial attitudinal information about an object.

Importantly, enhanced attention to context during the learning of initial attitudinal information should attenuate only ABC renewal, but not ABA renewal. In fact, ABA renewal should remain perfectly intact because evaluative responses in Context A should be driven by the contextualized representation of the initial attitudinal information. The same is true for evaluative responses in Context B. Because evaluative responses in Context B are driven by the contextualized representation of the subsequent counterattitudinal information, enhanced attention to context during the learning of initial attitudinal information should leave evaluative responses in Context B unaffected. That is, evaluations of the object in Context B should reflect the valence of counterattitudinal information that was learned in this context.

These predictions have been confirmed in a study by Gawronski et al. (2010, Experiment 1) that combined Rydell and Gawronski's (2009) impression formation paradigm with a priming manipulation designed to increase versus decrease attention to the background color during the

first block of the impression formation task. Toward this end, participants first completed a context priming task that involved the presentation of evaluative information about another individual (named Jim) before they learned about the target individual (named Bob). The presented information about Jim included both positive and negative information, which was randomly interspersed across trials. The information about Jim was presented against two background colors (i.e., brown, green). For half of the participants, there was a perfect contingency between background color and valence of information about Jim (e.g., positive-green; negative-brown). For the remaining half, there was no contingency between background color and valence. The first case was assumed to enhance attention to background color, because it was predictive of the valence of the information about Jim. The second case was assumed to reduce attention to background color, because background color was inconsequential in this case. Following the context priming task, participants completed the impression formation task by Rydell and Gawronski (2009), using blocked presentations of positive and negative information about the target individual Bob against two background colors that were different from the ones in the context priming task (i.e., yellow, blue).

Consistent with the predictions of the representational theory, ABC renewal was significantly reduced under conditions that aimed to enhance attention to context during the learning of initial attitudinal information about the target individual Bob; ABA renewal was unaffected by the attention priming manipulation (see Figure 3). Specifically, evaluative responses toward the target individual Bob were unaffected by the context priming manipulation when he was encountered against the background of the first learning block (Context A) or the second learning block (Context B). In either of these cases, evaluative responses toward Bob reflected the valence of the information learned in the respective context. In contrast, when the

target individual Bob was encountered against a novel background that was not part of the impression formation task (Context C), evaluative responses toward Bob reflected the valence of the initial attitudinal information when attention to context was primed to be low. When attention to context was primed to be high, evaluative responses were relatively neutral overall, presumably reflecting a mixture of the initial attitudinal and subsequent counterattitudinal information.

Another implication of the representational theory is that both ABA and ABC renewal should be attenuated when attention to context cues is reduced during the encoding of counterattitudinal information. In such cases, counterattitudinal information should be integrated into the initial, context-free representation, which should eliminate context effects altogether. That is, evaluative responses should reflect all of the available information about the object regardless of whether it is encountered in the initial Context A, the subsequent Context B, or a novel Context C. In other words, reduced attention to context during the encoding of counterattitudinal information should eliminate both ABA and ABC renewal, and thus the full pattern of contextualized attitude change.

Gawronski et al. (2010, Experiment 3) confirmed these predictions in a study that aimed to manipulate attention to context cues during the encoding of counterattitudinal information by presenting this information against either a single background color or multiple different background colors. The basic idea underlying this manipulation was that counterattitudinal information in multiple different contexts signals that the newly acquired information generalizes across contexts, which should reduce participants' attention to context. Consistent with these assumptions, Gawronski et al. found evidence for ABA and ABC renewal when counterattitudinal information was presented against a single background. In contrast, when

counterattitudinal information was presented against multiple different backgrounds, participants showed neutral evaluations regardless of whether the target was presented in the initial Context A, the subsequent Context B, or a novel Context C (see Figure 4).

Mere Attention or Causal Attribution?

The evidence reviewed thus far suggests at least two possible ways by which attentional processes may promote the formation of contextualized representations. First, one could argue that context cues are integrated into the representation of counterattitudinal information to the extent that these cues systematically covary with different kinds of information. For example, if Sally forms an initial positive impression of her new colleague Harry at work and later sees him being rude to another person at the gym, the gym context may be integrated into Sally's representation of Harry's negative behavior to the extent that Harry's behavior differs depending on whether he is in gym or somewhere else. This hypothesis resonates with covariation-based theories of causal attribution, suggesting that unexpected behaviors are attributed to factors that tend to covary with these behaviors (e.g., Försterling, 1989; Hilton & Slugoski, 1986; Kelley, 1973). Second, it is possible that enhanced attention to context is sufficient for an integration of the context in a contextualized representation regardless whether the context does or does not covary with the unexpected expected behavior. This scenario resembles attentional interpretations of illusory correlation effects (e.g., Hamilton & Gifford, 1976; Sherman, Kruschke, Sherman, Percy, Petrocelli, & Conrey, 2009), which suggest that attention to two co-occurring stimuli (e.g., Paul's negative behavior and the plane context) can create a link between these stimuli in memory even if they lack any objective contingency.

An important difference between the two accounts is that, in the former case, the context should be integrated into the representation of counterattitudinal information only when this

context differs from the context that was present during the learning of the initial attitudinal information (i.e., when context systematically covaries with valence). In the latter case, counterattitudinal information may become contextualized even when there is no difference in the context (i.e., even when context does not systematically covary with valence). Although the former hypothesis can account for the reviewed patterns of ABA and ABC renewal, it is unable to explain cases of AAB renewal (see Bouton & Ricker, 1994; Tamai & Nakajima, 2000). In AAB renewal, both initial attitudinal information and subsequent counterattitudinal information are learned in the same Context A, and counterattitudinal information effectively influences evaluations in that context. However, initial attitudinal information continues to influence evaluations in any other Context B (see Table 1). For example, if a person formed a favorable first impression of a new colleague at work and this impression was later challenged by negative behavior of that person in the same work context, evaluative responses to the new colleague should be negative only in contexts that are similar to the work context. In contrast, evaluative responses in any other context should reflect the initial positive impression of the new colleague.

An important aspect of AAB renewal is that there is no meaningful relation between context and valence, which implies that the context during the learning of counterattitudinal information does not systematically covary with the expectancy-violating behavior. Thus, although causal attributions might contribute to renewal effects when the contexts of initial attitudinal and subsequent counterattitudinal information differ, such attributions do not seem necessary to produce contextualized representations. Instead, renewal effects can occur as a result of mere attentional processes, such that enhanced attention to context cues may produce contextualized representations regardless of whether these cues do or do not systematically covary with the unexpected behavior.

The representational theory explains AAB renewal with the same set of assumptions that have been proposed to explain ABA and ABC renewal. That is, exposure to expectancy-violating counterattitudinal information enhances attention to context, which leads to an integration of the context into the representation of the counterattitudinal information. To the extent that attention to context is low during the learning of initial attitudinal information (Gilbert & Malone, 1995; Jones, 1991), the mental representation of the object acquires a "dual" nature, in that it comprises (a) a *context-free representation* of initially acquired attitudinal information, and (b) a *contextualized representation* of the subsequently acquired counterattitudinal information. Thus, if positive or negative information about an object is learned in an initial Context A, and this information is subsequently challenged by evaluatively incongruent information in the same Context A, encountering the object in a novel Context B should activate the initially formed context-free representation, eliciting an evaluative response that reflects the valence of the initial attitudinal information (*AAB renewal*). In contrast, if positive or negative information about an object is learned in an initial Context A, and this information is subsequently challenged by evaluatively incongruent information in the same Context A, encountering the object in Context A should activate the subsequently formed contextualized representation, eliciting an evaluative response that reflects the valence of the counterattitudinal information.

Evidence for AAB renewal in the formation and change of social attitudes comes from a study by Gawronski et al. (2010, Experiment 4). Using Rydell and Gawronski's (2009) impression formation paradigm, participants were first presented with either positive or negative information about a target person against a meaningless, colored background. In a second block, participants were presented with information that was evaluatively opposite to the information provided in the first block. However, different from Rydell and Gawronski's (2009) procedure,

this information was presented against the *same* colored background. Evaluative responses to the

target person were assessed with an affective priming task (Payne et al., 2005) in which the

target person was presented against either the background of the two blocks of the impression

formation task (Context A) or a novel background that was not shown during the impression

formation task (Context B). Results showed that evaluations of the target reflected the valence of

the newly learned information when the target was presented against the background that was

displayed during the two blocks of the impression formation task. However, evaluations of the

target reflected the valence of the information learned initially when the target was presented

against the novel background (*AAB renewal*).

Mental Representation of Contextualized Attitudes

Constrained versus Direct Activation of Valence

An important question within the representational theory concerns the mental structure of

contextualized representations of counterattitudinal information. On the one hand, it is possible

that context cues are stored in a manner such that they constrain which information is activated

by the target object. On the other hand, context cues might become directly associated with the

valence of the counterattitudinal experience, such that they influence evaluative responses over

and above the information that is available about the target (see Bouton, 2010; Vervliet,

Baeyens, Van den Bergh, & Hermans, 2013). To illustrate the difference between the two cases,

consider the earlier example in which a person formed a favorable first impression of a new

colleague at work, and this impression is later challenged by the colleague's negative behavior at

the gym. In this case, subsequent encounters with the colleague at the gym may elicit a negative

response because either (a) the gym context facilitates the activation of negative information

about the colleague and inhibits the activation of positive information about the colleague or (b)

the gym context directly elicits a negative response over and above the response that is elicited by the available information about the colleague.

The first interpretation is consistent with the idea that context cues are represented as modulatory nodes that determine whether the initial attitudinal information or the subsequent counterattitudinal information is activated in response to the attitude object (see Bouton, 1994). According to this account, contextualized representations constrain the spread of activation from the attitude object to the available evaluative information by virtue of inhibitory links (see Figure 5, left panel). If the context of the counterattitudinal experience is absent, activation of the object node is assumed to spread to the initial attitudinal information, which in turn inhibits the activation of the counterattitudinal information. In contrast, if the context of the counterattitudinal experience is present, activation of the context node is assumed to inhibit the link between the attitude object and the initial attitudinal information, thereby gating the spread of activation from the attitude object to the counterattitudinal information, which further inhibits the activation of the initial attitudinal information.[1] Applied to the above example of a new colleague engaging in expectancy-violating negative behavior at the gym, the gym context would inhibit the associative link between the colleague and the initial positive information. This inhibition would, in turn, promote the spread of activation from the colleague to the counterattitudinal negative information, which should further inhibit the activation of the initial positive information.

---

[1] Note that a similar idea can be implemented in episodic theories in which each experienced event is stored as a specific memory trace rather than in terms of abstract valence (e.g., Hintzman, 1986; Schmidt, De Houwer, & Rothermund, 2016). Within episodic models, context functions as a retrieval cue that increases the probability of retrieving specific episodic information previously encountered in that context.

The second interpretation is consistent with the idea that enhanced attention to context cues during the learning of counterattitudinal information may create a direct link between the mental representation of these cues and the valence of the counterattitudinal experience (see Rescorla & Wagner, 1972). From this perspective, the observed differences in evaluative responses may be driven by additive effects of independent excitatory links between (a) the attitude object and the available information about that object and (b) the context of the counterattitudinal information and the valence of the counterattitudinal experience (see Figure 5, right panel). As a result, evaluative responses should be more likely to reflect the valence of the counterattitudinal experience when the attitude object is encountered in the context in which this experience was made than when it is encountered in any other context. Applied to the above example of a new colleague engaging in expectancy-violating negative behavior at the gym, the gym context would directly elicit a negative response in addition to the negative response that arises from the spread of activation from the new colleague to the counterattitudinal negative information.

Using Rydell and Gawronski's (2009) impression formation paradigm, Gawronski et al. (2010) tried to differentiate between the two accounts by comparing evaluative responses to the target individual to those elicited by other unknown individuals within the same contexts. According to the constrained activation account, context effects on evaluative responses should be specific to the target individual and not generalize to other unknown individuals within the same context. In contrast, the direct activation account implies that context cues should influence evaluative responses independent of the target individual, because the context is directly linked to the valence of the counterattitudinal experience. Gawronski et al. (2010) found that context cues moderated evaluative responses to the target individual, but not evaluative responses to

other unknown individuals that were presented in the same contexts. These findings are

consistent with the hypothesis that context cues constrain the activation of evaluative information

in response to the target individual (see Figure 5, left panel), and inconsistent with the hypothesis

that context cues directly activate the valence of the counterattitudinal experience (see Figure 5,

right panel).

Although Gawronski et al.'s (2010) findings provide preliminary evidence against the

direct activation account, a potential objection is that the null effect of context cues in

influencing evaluative responses to unknown individuals could be due to incidental features of

the unknown individuals. Specifically, it is possible that incidental facial features of the

unknown individuals elicited a positive or negative response (e.g., Gawronski & Quinn, 2013),

which may dilute or override the simultaneous effects of the contexts. Such incidental effects

could undermine the detection of direct effects of the context cues, thereby leading to a

premature rejection of the direct activation account.

More compelling evidence regarding the two accounts was provided by Gawronski et al.

(2014). A central aspect of the direct activation account is that the formation of a direct link

between context and valence presupposes a systematic relation between the valence of

counterattitudinal experiences and the context in which these experiences were made. If there is

no such relation, there should be no direct links between context and valence, and therefore no

effects of the context on evaluative responses. This prediction is different from the one implied

by the constrained activation account, which suggests that context cues may constrain the

activation of evaluative information about an object irrespective of whether they are directly

associated with a particular valence.

To test the two competing hypotheses, Gawronski et al. (2014, Experiment 2) used a modified version of Rydell and Gawronski's (2009) paradigm that included evaluative information about two individuals rather than one. To avoid any systematic relation between context and valence, one of the two targets was described as positive in an initial Context A, whereas the other one was described as negative in the same Context A. In a subsequent block, the initially positive target was presented with negative information in a second Context B, while the initially negative target was presented with positive information in the same Context B. Finally, evaluations of the two targets were assessed in the initial Context A, the second Context B, and a novel Context C that was not part of the impression formation task. Consistent with the hypothesis that context cues are stored in a manner such that they constrain which information is activated by the target, evaluative responses were moderated by the presence versus absence of the second context despite the absence of any systematic relation between context and valence (see Figure 6). That is, evaluations of the two targets reflected the valence of the counterattitudinal information about the targets only when the targets were encountered in the context in which the counterattitudinal information had been acquired. In contrast, evaluations of the targets reflected the valence of the initial attitudinal information when the targets were encountered either in the context of the initial attitudinal information or in a novel context that had not been part of impression formation task. These results support the idea that context cues are stored in a manner such that they constrain which information is activated by the target (see Figure 5, left panel), and are inconsistent with the idea of direct links between the mental representation of these cues and the valence of the counterattitudinal experience (see Figure 5, right panel). According to the latter idea, the absence of a systematic relation between context and valence should eliminate contextual renewal effects, which was not the case.

Although Gawronski et al.'s (2014) findings support the idea that context cues are stored in a manner such that they constrain which information is activated by the target object, Gawronski et al.'s (2015) meta-analysis suggests that both mechanisms can contribute to renewal effects. Consistent with the idea of direct links between context and valence, effects sizes of both ABA and ABC renewal were significantly more pronounced when the study design included a systematic relation between context and valence (i.e., when the impression formation task included a single target individual) than when there was no such relation (i.e., when the impression formation task included two target individuals with opposite valence in the two blocks of the task). Yet, effect sizes of both types of renewal effects were still statistically significant when there was no systematic relation between context and valence. The finding that contextual renewal occurs in the absence of a systematic relation between context and valence supports the idea that context cues are stored in a manner such that they constrain which information is activated by the target object (see Figure 5, left panel); the finding that contextual renewal is more pronounced when there is a systematic relation between context and valence supports the idea that the contexts can additionally become associated with the valence of the counterattitudinal experience (see Figure 5, right panel). From this perspective, Gawronski et al.'s (2015) meta-analytic findings indicate that both mechanisms can jointly contribute to contextual renewal effects, although direct links between context and valence do not seem to be necessary for their emergence.

Further evidence against the necessity of direct links between context and valence was provided by Gawronski et al. (2014, Experiment 3). A central question of this study was whether context cues retain their modulating function when they become associated with an evaluative response that is opposite to the one elicited by the target object within that context. For example,

if negative experiences were made with a positively evaluated person in the context of a gym, will visual cues related to the gym context continue to activate a negative response toward the target when the gym context becomes directly associated with a positive response? Persistence in contextual modulation after such "counterconditioning" of the relevant context provides further evidence that the modulating function of context cues does not depend on direct links between context and valence (see De Houwer, Crombez, & Baeyens, 2005). Gawronski et al. addressed this question by repeatedly pairing context cues with positive or negative images after participants had completed an impression formation task similar to the one by Rydell and Gawronski (2009). On the basis of previous research on evaluative conditioning (see De Houwer et al., 2001), Gawronski et al. expected that repeated pairings of context cues with positive and negative images would influence evaluative responses to these cues in line with the valence of the images. More importantly, these newly acquired responses to the contexts were expected to leave their modulating function unaffected, such that the contexts should continue to moderate the evaluative response that is elicited by the target.

Consistent with this hypothesis, context cues continued to moderate the evaluative response toward a target person even when the context cues became subsequently associated with a particular valence by virtue of repeated pairings with positive or negative stimuli. Interestingly, the results showed that context cues had two distinct effects on evaluative responses when they independently acquired a particular valence (see also Urcelay & Miller, 2010). First, they continued to moderate the evaluative response that was elicited by the target person within that context independent of their conditioned valence (see Figure 7). Second, they directly elicited an evaluative response reflecting their conditioned valence independent of the evaluative response that was elicited by the target person within that context (see Figure 8). For

example, when participants learned positive information about a target in Context A and then counterattitudinal negative information about the same target in Context B, they continued to show a negative response toward the target in Context B (see Figure 7) even when Context B itself elicited a positive response as a result of repeated pairings with positive images (see Figure 8). Conversely, when participants learned negative information about a target in Context A and then counterattitudinal positive information about the same target in Context B, they continued to show a positive response toward the target in Context B (see Figure 7) even when Context B itself elicited a negative response as a result of repeated pairings with negative images (see Figure 8). Using the previous example, if new negative information was learned about an initially positively evaluated target in the context of a gym and the gym becomes subsequently associated with a positive response, visual cues related to the gym context will have two distinct effects when the target is encountered at the gym: (a) they will constrain the activation of available information about the target, leading to a negative response toward the target within the gym context, and (b) they will directly elicit a positive response despite the negative response that is elicited by the target within that context.

Abstract versus Specific Representation of Evaluative Information

An interesting question is whether "dual" representations of attitudinal and counterattitudinal information involve abstract information about the valence of an object or concrete information about specific attributes. This question is closely related to the debate between dispositional and constructivist conceptualizations of attitudes (see Gawronski, 2007). Dispositional accounts hypothesize that attitudes are stored in the form of abstract summary evaluations in long-term memory. According to Fazio (2007), these representations can be conceptualized as mental associations between an object and its overall valence. Constructivist

accounts reject the notion of enduring dispositional tendencies resulting from abstract

representations of valence, and instead argue that all evaluative responses are constructed on the

spot on the basis of momentarily accessible information (Schwarz, 2007).

Although the debate between the two competing views is notoriously difficult to resolve,

some of our findings lend support to the idea of abstract attitudinal representations. Using

Gawronski et al.'s (2014) expectancy-violation paradigm, Brannon, Sacchi, and Gawronski

(2017) presented participants with 30 behavioral statements about a target individual that were

displayed one-by-one against different background colors. The initial 20 statements about the

target were either positive or negative and varied independently on the trait dimensions of

warmth and competence. That is, participants learned that the target individual was either (a)

high on warmth, (b) low on warmth, (c) high on competence, or (d) low on competence. The 21st

statement was used as a target statement and described a behavior suggesting either high warmth,

low warmth, high competence, or low competence. The valence and dimension of the target

statement was independent of the valence and dimension of the initial impression statements.

Thus, the valence of target statement could either be congruent or incongruent with the valence

of the initial statements, and the dimension of the target statement could either match or

mismatch the dimension of the initial statements. The target statement was followed by 9

distracter statements that matched the dimension and valence of the initial 20 statements.

Participants' task was to form an impression of the target individual. After the impression

formation task, participants completed a surprise recognition test, in which they had to identify

the background color against which the target statement was presented during the impression

formation task.

Expanding on Gawronski et al.'s (2014) finding of a memory advantage for the background of expectancy-violating information, Brannon et al. (2017) found an expectancy-violation effect for behaviors that were incongruent with the valence of prior information, regardless of whether the dimension of the target statement matched the dimension of the initial statements. For example, when participants formed an initial impression of the target as being warm (cold), they showed an expectancy-violation effect when they were presented with new information suggesting that the target is incompetent (competent). Conversely, when participants formed an initial impression of the target as being competent (incompetent), they showed an expectancy-violation effect when they were presented with new information suggesting that the target is cold (warm). This pattern replicated across three independent studies with close to 2,000 total participants. Together, these results suggest that participants' representation of the target captured the abstract valence of the initially acquired information independently of more nuanced details about its dimensional content. This conclusion is consistent with the hypothesis that initial attitudinal information is stored in the form of general positivity versus general negativity (Fazio, 2007), but it seems difficult to reconcile with the constructivist argument that there are no abstract representations of valence (Schwarz, 2007).

An open question is whether contextualized representations of counterattitudinal information involve abstract representations of valence or concrete representations of specific information. Although Brannon et al.'s (2017) findings do not provide an answer to this question, their paradigm could be modified to gain deeper insights in this regard. Instead of asking participants to identify the background color against which a given statement was presented during the impression formation task, participants could be asked to identify the statement that was presented against a given background color. By using multiple distracter

statements that vary in terms of valence and dimensional content (e.g., warmth vs. competence),

participants' responses in the recognition task could be analyzed with advanced modeling

techniques to quantify participants' memory for (a) the actual statement, (b) the valence of the

statement, and (c) the dimensional content of the statement (see Klauer & Wegener, 1998). Data

along these lines would help to provide deeper insights into whether contextualized

representations of counterattitudinal information involve abstract representations of valence or

concrete representations of specific information.

Pertinent Questions

Is There Evidence for Valence Asymmetries?

Another interesting finding of Brannon et al.'s (2017) studies is that there was no

evidence for valence asymmetries in the violation of expectancies. That is, negative

counterattitudinal information was as expectancy-violating as positive counterattitudinal

information. Yet, based on previous evidence for a negativity bias in impression formation (for a

review, see Skowronski & Carlston, 1989), one might expect that negative counterattitudinal

information should have a greater impact than positive counterattitudinal information, and

therefore lead to stronger expectancy-violations. A more nuanced prediction could be derived

from previous research on the perceived diagnosticity of positive and negative information in

different content domains (see Reeder & Brewer, 1979; Skowronski & Carlston, 1989). This

research suggests that negative information related to warmth is perceived as more diagnostic of

an underlying trait than positive information, whereas positive information related to competence

is perceived as more diagnostic of an underlying trait than negative information. In line with this

view, negative information typically receives greater weight in impressions of warmth, whereas

positive information receives greater weight in impressions of competence. This conclusion

suggests that the pattern of valence asymmetries in expectancy-violation may depend on the trait

dimension. Specifically, new information about a person's warmth may elicit a stronger

expectancy-violation when new negative information conflicts with an initial positive impression

of that person's warmth than when new positive information conflicts with an initial negative

impression of that person's warmth (i.e., negativity bias). Conversely, new information about a

person's competence may elicit a stronger expectancy-violation when new positive information

conflicts with an initial negative impression of that person's competence than when new negative

information conflicts with an initial positive impression of that person's competence (i.e.,

positivity bias).

Counter to these predictions, Brannon et al. (2017) did not find any evidence for valence

asymmetries in the violation of expectancies regardless of whether the trait dimension involved

warmth or competence. This result replicated across three independent studies and in a combined

analysis that included the data from all three studies. Follow-up research by Brannon and

Gawronski (2017a) further demonstrated that the lack of valence asymmetries in expectancy-

violation is independent of the strength of the initial expectancy, such that expectancy violation

effects were equally pronounced regardless of whether participants received large or small

amounts of initial attitudinal information.

Consistent with Brannon et al.'s (2017) findings on the violation of expectancies, none of

our individual studies on contextualized attitude change revealed any evidence for valence

asymmetries (Gawronski et al., 2010, 2014; Rydell & Gawronski, 2009). That is, neither ABA

nor ABC renewal differed as a function of whether (a) initial positive information was

challenged by counterattitudinal negative information or (b) initial negative information was

challenged by counterattitudinal positive information. The only evidence for valence

asymmetries comes from Gawronski et al.'s (2015) meta-analysis, which revealed a marginally smaller effect size for ABC renewal when initial negative information was challenged by counterattitudinal positive information than when initial positive information was challenged by counterattitudinal negative information. Because ABC renewal is characterized by a weak influence of counterattitudinal information in novel contexts, these meta-analytic findings suggest that counterattitudinal positive information tends to be more impactful in influencing evaluations in novel contexts than counterattitudinal negative information. There was no evidence for valence asymmetries in ABA renewal.

It is worth noting that all studies in Gawronski et al.'s (2015) meta-analysis involved information along the warmth dimension. Therefore, the findings of the meta-analysis suggest a valence asymmetry that is *opposite* to the aforementioned prediction of a negativity bias along the warmth dimension. That is, ABC renewal was more pronounced when initial negative information along the warmth dimension was challenged by counterattitudinal positive information than when initial positive information was challenged by counterattitudinal negative information. Drawing on the assumptions of the representational theory, a potential explanation for the obtained 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. As a result, context 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 the representational 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 subsequent negative information. More research is needed,

however, to establish the reliability of the obtained valence asymmetry before exploring its

underlying mechanisms.

Which Context Features Determine the Activation of Contextualized Representations?

The studies reviewed thus far have focused on effects of relatively simple, one-

dimensional visual cues, such as the background color of the computer screen. These findings

suggest that context cues do not need to have any meaning that is semantically related to the

content of the counterattitudinal information in order to moderate the activation of this

information (cf. Ma, Correll, & Wittenbrink, 2016). This approach also allowed us to gain

valuable insights into the processes underlying contextualized attitude change by eliminating

potential influences of the valence of real-world contexts. Yet, two important questions that still

need to be addressed are: (a) do context effects resulting from these processes generalize to real-

world contexts with higher levels of complexity, and if so, (b) which features of complex real-

world contexts determine the activation of contextualized representations? For example, if

counterattitudinal information about an object was learned in the context of a seminar room, does

only the same seminar room activate the representation of the counterattitudinal information or

will other contexts that are similar to the seminar room have the same effect? If similar contexts

can have the same effect, in which particular sense do they have to resemble the context in which

the counterattitudinal information was learned? Would any seminar room have the same effect

even if it is perceptually dissimilar to the one in which the counterattitudinal information was

learned (e.g., a perceptually distinct seminar room in a different building)? Alternatively, would

a room that is perceptually similar to the seminar room have the same effect even if it is not a

seminar room (e.g., a dining hall that visually resembles the seminar room)?

Gawronski et al. (2014, Experiment 4) addressed these questions by investigating effects of real-world contexts that were either perceptually or conceptually similar to the context in which counterattitudinal information was learned. Toward this end, participants were presented with evaluatively incongruent information about a target individual against two different real-world backgrounds. Evaluative responses to the target were then measured against (a) the background of the initial attitudinal information, (b) the background of the counterattitudinal information, (c) a background that was perceptually similar to, but conceptually distinct from, the background of the counterattitudinal information, (d) a background that was conceptually similar to, but perceptually distinct from, the background of the counterattitudinal information, and (e) a background that was both perceptually and conceptually distinct from the background of the counterattitudinal information (see Figure 9). The central question was whether backgrounds that are either perceptually or conceptually similar to the background of the counterattitudinal information elicit evaluative responses in line with the valence of the counterattitudinal information.

The results suggest that contextualized representations of counterattitudinal information can be activated by either (a) contexts that are perceptually similar to, but conceptually distinct from, the context in which the counterattitudinal information was learned, or (b) contexts that are conceptually similar to, but perceptually distinct from, the context in which the counterattitudinal information was learned (see Figure 10). Both kinds of contexts produced evaluative responses to the target that reflected the valence of the counterattitudinal information. In contrast, contexts that were both perceptually and conceptually distinct from the context in which the counterattitudinal information had been learned produced a renewal effect, such that evaluative responses to the target reflected the valence of the initial attitudinal information (*ABC renewal*).

The same was true for the context of the initial attitudinal information, which elicited evaluative responses to the target that were in line with the initial attitudinal information (*ABA renewal*). Taken together, these results indicate that contextualized representations can be activated by contexts that are either perceptually or conceptually similar to the context in which the counterattitudinal information was acquired.

Are There Individual Differences?

Another pertinent question is whether there are individual differences in the tendency to form contextualized attitudes. Based on the core assumptions of the representational theory, there are a few potential candidates for individual difference effects. A central issue is how people respond to conflicting information about an attitude object. For example, individuals with a strong preference for consistency might experience a stronger expectancy-violation in response to counterattitudinal information compared to individuals with a weak preference for consistency (see Cialdini, Trost, & Newsom, 1995). A similar hypothesis could be derived for individual differences in the need for structure, such that individuals with a high need for structure might experience a stronger expectancy-violation in response to counterattitudinal information compared to individuals with a low for need for structure (see Neuberg & Newsom, 1993). In the domain of interpersonal attitudes, one could further argue that individuals with an entity theory of personality (i.e., a lay theory that a person's traits cannot change) might experience a stronger expectancy-violation in response to counterattitudinal information about another person compared to individuals with an incremental theory of personality (i.e., a lay theory that a person's traits are malleable) (see Chiu, Hong, & Dweck, 1997; Plaks, in press). Because expectancy-violation is assumed to be the driving force behind enhanced attention to context during the learning of counterattitudinal information, these considerations suggest that

individuals with a weak preference for consistency, a low need for structure, or an incremental theory of personality might be less likely to form contextualized representations of counterattitudinal information compared to individuals with a strong preference for consistency, a high need for structure, or an entity theory of personality, respectively.

Counter to these hypotheses, a study by Brannon and Gawronski (2017b) did not obtain any significant relations between individual differences along the three variables and the tendency to show ABA or ABC renewal effects, regardless of whether initial positive information was challenged by counterattitudinal negative information, or vice versa. These findings replicated in two independent studies with a Canadian and an American sample as well as in a combined analysis of the data from both samples (see Table 2). Nevertheless, both studies revealed the typical pattern of ABA and ABC renewal. Although it is possible that ABA and ABC renewal effects are systematically related to other kinds of individual differences, these findings contradict the hypotheses that individual differences in the preference for consistency, the need for structure, and lay theories of personality moderate the tendency to show contextualized attitude change.

Are There Cultural Differences?

Another interesting question concerns the universality of contextualized attitude change across cultures. A common assumption in the cross-cultural literature is that individuals from Eastern cultures show a greater sensitivity to contextual information than individuals from Western cultures. Expanding on this assumption, Ye and Gawronski (in press) discussed two alternative predictions regarding cultural differences in contextualized attitude change.

First, it is possible that individuals from Eastern cultures pay more attention to context cues during the learning of initial attitudinal information than individuals from Western cultures

(e.g., Chua, Boland, & Nisbett, 2005; Masuda & Nisbett, 2001). Hence, individuals from Eastern cultures may integrate conflicting information in two contextualized representations, one including initial attitudinal information and the other including counterattitudinal information (see Gawronski et al., 2010). In contrast, individuals from Western cultures may pay attention to context cues only when their expectancies are violated, leading to the formation of context-free representations of initial attitudinal information and contextualized representations of counterattitudinal information. These assumptions imply that Easterners and Westerners should differ specifically in their susceptibility to ABC renewal. That is, evaluations in a novel Context C should reflect the valence of initial attitudinal information for individuals from Western cultures, but this effect should be reduced for individuals from Eastern cultures. In contrast, evaluations in Context A should not differ across cultures, reflecting the valence of initial attitudinal information for individuals from both cultures. The same is true for evaluations in Context B, which should reflect the valence of counterattitudinal information regardless of culture.

Second, it is possible that individuals from Eastern cultures pay less attention to context cues during the learning of counterattitudinal information than individuals from Western cultures. A central assumption of Gawronski et al.'s (2010) representational theory is that attention to context cues during the learning of counterattitudinal information is enhanced by violated expectancies, specifically by the inconsistency between attitudinal and counterattitudinal information. A substantial body of research has shown that individuals from Eastern cultures tend to have a higher tolerance for inconsistency than individuals from Westerner cultures (Spencer-Rodgers, Williams, & Peng, 2010). This difference has been attributed to cultural differences in reasoning styles, in that dialectical thinking tends to be more prevalent among

Easterners, whereas analytical thinking tends to be more prevalent in among Westerners (Choi &

Nisbett, 2000; Peng & Nisbett, 1999). A fundamental difference between the two thinking styles

is that an object can have a particular attribute as well as the opposite attribute from a dialectical

point of view, which would be inconsistent from an analytic perspective.

Applied to the current question, these findings suggest that individuals from Eastern

cultures may pay less attention to context during the learning of counterattitudinal information

than individuals from Western cultures. Thus, individuals from Eastern cultures may integrate

conflicting information in a single context-free representation that includes both attitudinal and

counterattitudinal information. In contrast, individuals from Western cultures may experience a

strong expectancy-violation in response to counterattitudinal information, which should increase

attention to context. As a result, individuals from Western cultures should form context-free

representations of initial attitudinal information and contextualized representations of

counterattitudinal information. Together, these assumptions suggest ABA and ABC renewal

should occur for participants from Western cultures, but not for participants from Eastern

cultures. This hypothesis implies the counterintuitive yet interesting prediction that evaluative

responses of Eastern participants should be *less* sensitive to contextual influences than evaluative

responses of Western participants. Whereas evaluative responses of Western participants 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, evaluative

responses of Eastern participants may reflect a mixture of the available information regardless of

the context.

Although Gawronski et al.'s (2015) meta-analysis provides preliminary evidence for the

second hypothesis (i.e., eliminated ABA and ABC renewal in Eastern cultures), a reanalysis of

the available data by Ye, Tong, Chiu, and Gawronski (2017) suggests that the effects sizes for

ABA and ABC renewal in Eastern samples were suppressed by the inclusion of data obtained

with a measure that is known for its low reliability (see Gawronski & De Houwer, 2014).[2] When

these data were excluded, a combined analysis of all available studies with participants from

Eastern cultures (including a new study that was conducted after the publication of the meta-

analysis) revealed significant effects for both ABA and ABC renewal (Ye et al., 2017). Thus,

counter to the two competing predictions regarding cross-cultural differences in contextual

renewal (see Ye & Gawronski, in press), the currently available evidence suggests that

contextualized attitude change occurs in both Western and Eastern cultures. Using Gawronski et

al.'s (2014) expectancy-violation paradigm, a follow-up study by Ye et al. (2017) provided

further evidence for this conclusion, showing that individuals from Eastern and Western cultures

show the same expectancy-violation effect in response to counterattitudinal information (see

Figure 11).

Theoretical Challenges

Although there is considerable evidence supporting the predictions of the representational

theory of contextualized attitude change, there are a few findings that pose a challenge to the

theory. One example concerns the types of information that lead to context-independent changes

in spontaneous evaluations. Counter to the widespread assumption that spontaneous evaluations

are highly robust and difficult to change (e.g., Rydell, McConnell, Strain, Claypool, &

Hugenberg, 2007), two recent sets of studies suggest that spontaneous evaluations can change

---

[2] The critical measure was the evaluative priming task (Fazio, Jackson, Dunton, & Williams, 1995), which rarely shows Cronbach's α values higher than .50 (see Gawronski & De Houwer, 2014). Low reliability can undermine the detection of actually existing effects by reducing statistical power through large proportions of measurement error (LeBel & Paunonen, 2011).

rapidly in response to a single piece of information (Cone & Ferguson, 2015; Mann & Ferguson, 2015; for a review, see Cone, Mann, & Ferguson, in press). From the perspective of the representational theory, these findings raise the question of whether such changes generalize across contexts or instead are limited to the context in which the new information was learned.

First, Mann and Ferguson (2015) found that new information suggesting a reinterpretation of prior information led to rapid changes in spontaneous evaluations. In a series of studies, participants formed a negative impression of a target individual by reading a story that depicted the individual performing various negative actions (e.g., breaking into a house and taking precious things from the bedroom). After forming the impression, participants received additional information that suggested a positive interpretation of the target's negative actions (e.g., the individual broke into the house to save the family's kids from a fire). Counter to the hypothesis that changes in spontaneous evaluations require large amounts of counterattitudinal information (e.g., Rydell et al., 2007), participants showed a rapid revision of their spontaneous evaluations in response to the new information suggesting a reinterpretation of the initial attitudinal information. A follow-up study showed that such reinterpretations can reverse the effect of previously acquired attitudinal information even when there is a delay of several days between the learning of the initial attitudinal information and the information suggesting a reinterpretation (Mann & Ferguson, 2017).

Another set of experiments by Cone and Ferguson (2015) demonstrated the effectiveness of extreme "diagnostic" information in rapidly changing spontaneous evaluations. In their studies, participants formed an impression of a target individual via large amounts of positive information and then received one piece of either highly diagnostic negative information or neutral information. Again, counter to the hypothesis that changes in spontaneous evaluations

require large amounts of counterattitudinal information (e.g., Rydell et al., 2007), participants who received a single piece of diagnostic negative information showed a rapid revision of their spontaneous evaluations in response to the diagnostic negative information.

Expanding on the representational theory of contextualized attitude change, Brannon and Gawronski (2017c) investigated whether the two cases of rapid updating reflect context-dependent or context-independent changes in spontaneous evaluations. Drawing on the hypothesis that context-dependent change results from enhanced attention to the context during the encoding of expectancy-violating information, Brannon and Gawronski predicted that reinterpretation of earlier information should lead to *context-independent* changes in spontaneous evaluations. In contrast, diagnostic counterattitudinal information was predicted to produce *context-dependent* changes in spontaneous evaluations.

The first prediction was derived from the idea that information that sheds new light on old information simply provides additional cues regarding the meaning of the old information, which does not involve any kind of expectancy-violation. In such cases, the new information may not violate any expectancies, because it does not directly contradict the initial information. Because expectancy-violation is assumed to be a central determinant of attention to context (Gawronski et al., 2010), the new information should simply lead to a reinterpretation of the old information instead of being stored in a contextualized representation. As a result, it should lead to changes in spontaneous evaluations that generalize across contexts.

The second prediction was derived from the idea that diagnostic information that directly conflicts with initially acquired attitudinal information should violate perceivers' expectancies, and thereby enhance attention to the context during the learning of the counterattitudinal

information. As a result, diagnostic counterattitudinal information should be stored in a contextualized representation, leading to context-dependent changes in spontaneous evaluations.

To test these predictions, Brannon and Gawronski (2017c) conducted two experiments that adapted the paradigms by Mann and Ferguson (2015) and Cone and Ferguson (2015) and combined them with the contextualized-change paradigm by Rydell and Gawronski (2009). Consistent with the first hypothesis, their results showed that a single piece of information that shed a positive light on prior negative information led to context-independent changes in spontaneous evaluations (see Figure 12). Counter to the second hypothesis, however, a single piece of diagnostic counterattitudinal information also resulted in context-independent changes in spontaneous evaluations (see Figure 13). Because diagnostic counterattitudinal information should violate attitude-related expectancies, and thereby enhance attention to the context during the learning of the counterattitudinal information, the latter finding poses a challenge to the predictions derived from the representational theory.

There are at least two potential explanations for this discrepancy. First, it is possible that extreme counterattitudinal information is capturing attention to an extent that it directs attention away from the context.[3] In this case, the relation between expectancy-violation and contextual renewal may be curvilinear rather than linear. That is, for low to moderate levels of expectancy-violation, attention to context may continuously increase, leading to a parallel increase in contextual renewal effects. In contrast, for moderate to high levels of expectancy-violation, more attention may be devoted to the expectancy-violating information at the expense of attention to context, leading to a decrease in contextual renewal effects. From this perspective, contextual

[3] We thank Nicholas Patton for this idea.

renewal effects may be strongest for moderate levels of expectancy-violation, and weaker for

low and high levels of expectancy-violation (i.e., inverse U-shaped function). We are currently

conducting a study designed to test this hypothesis.

A second potential explanation is that, in the domain of impression formation, extreme

negative behavior is deemed indicative of core personality characteristics, resulting in

dispositional attributions that generalize across contexts (Fiske, 1980; Reeder, 1993). This may

be especially the case within the domain of (im)moral behavior (see Reeder & Brewer, 1979;

Skowronski & Carlston, 1989). Hence, when confronted with expectancy-violating behavior of

extreme negative valence, perceivers may engage in dispositional attributions that fully override

the effects of enhanced attention to the context. From this perspective, Brannon and Gawronski's

(2017c) results may not necessarily disprove the core assumptions of the representational theory.

Instead, they may suggest an important boundary condition that has to be reconciled with the

original theory by means of an additional assumption: effects of enhanced attention to context

can be overridden by dispositional attributions.

This conclusion also suggests a different interpretation of Gawronski et al.'s (2010)

finding that exposure to counterattitudinal information in multiple different contexts eliminates

both ABA and ABC renewal (see Figure 4). We initially interpreted this finding in terms of

attention to context during the learning of counterattitudinal information. Our central assumption

was that counterattitudinal information in multiple different contexts signals that the newly

acquired information generalizes across contexts, which was assumed to reduce attention to

context. Yet, this interpretation is difficult to reconcile with Gawronski et al.'s (2014) finding

that even a single piece of expectancy-violating information was sufficient to enhance attention

to context (see Figure 2). Hence, although attention to context may decline with increasing

amounts of counterattitudinal information when counterattitudinal information is learned in multiple contexts, the very first piece of counterattitudinal information should still be stored in a contextualized representation. Based on the findings by Brannon and Gawronski (2017c), a more plausible explanation is that counterattitudinal information in multiple contexts leads to dispositional attributions that fully override the effects of enhanced attention to context. That is, people may infer a context-independent disposition when counterattitudinal information is learned in multiple different contexts, and such dispositional attributions may counteract the influence of enhanced attention to context during the learning of the first piece of counterattitudinal information. Thus, although evidence for AAB renewal suggests that causal attributions to situational factors may not be necessary for contextualized attitude change, causal attributions to dispositional factors may have the potential to fully override the effects of enhanced attention to context during the learning of counterattitudinal information.

Although causal attribution to dispositional factors provides an integrative explanation for context-independent effects of counterattitudinal information that is either (a) highly diagnostic or (b) encountered in multiple different contexts, any such explanation additionally has to assume a retroactive interference with the integration of the context into the representation of the counterattitudinal information. In both cases, dispositional attribution would have to counteract the immediate effects of enhanced attention to context during exposure to expectancy-violating counterattitudinal information (see Brannon et al., 2017; Gawronski et al., 2014). One important factor in this regard might be the consolidation of a newly formed contextualized representation. That is, dispositional attribution may effectively prevent the integration of context cues into the representation of counterattitudinal information if this representation did not have a chance to consolidate. To the extent that a newly formed contextualized

representation had a chance to consolidate, however, dispositional attribution may be ineffective in retroactively erasing the context from the existing representation. This hypothesis has important implications for the effects of counterattitudinal information in multiple different contexts, because it suggests a boundary condition for the obtained elimination of contextual renewal effects (Gawronski et al., 2010). That is, to the extent that a newly formed contextualized representation had a chance to consolidate, delayed learning of additional counterattitudinal information in other contexts may be ineffective in eliminating ABA and ABC renewal. Future research testing this hypothesis would help to provide deeper insights into the potential role of attributional processes and memory consolidation in the elimination of contextual renewal effects.

Another challenge to the representational theory is an unexpected finding of Gawronski et al.'s (2015) meta-analysis, suggesting that enhanced attention to context during initial attitudinal learning does not reduce the average effect size for ABC renewal. If anything, the average effect size for ABC renewal was greater when attention to context during the learning of initial attitudinal information was enhanced. This finding stands in contrast to the predictions of the representational theory, which implies that enhanced attention to context during initial attitudinal learning should reduce ABC renewal. Although it is possible that the lack of a difference in the obtained effect sizes of ABC renewal is due to procedural confounds between different studies in the meta-analysis, it poses a major challenge to the predictions of the representational theory. Future research is needed to further investigate the hypothesized role of attention to context during initial attitudinal learning.

Future Directions

Spontaneous versus Deliberate Evaluations

To investigate contextualized changes in attitudes, all of the reviewed studies relied on measures of spontaneous evaluations. For example, a substantial proportion of our work (e.g., Brannon & Gawronski, 2017c; Gawronski et al., 2010; Rydell & Gawronski, 2009; Ye et al., 2017) has used the affect misattribution procedure (AMP; Payne et al., 2005), which capitalizes on the notion of sequential priming to measure unintentional responses to a prime stimulus by assessing its effects on judgments of a neutral target stimulus. In our studies on contextualized attitude change, participants were briefly presented with images of the impression target against (a) the background of initial attitudinal information, (b) the background of the counterattitudinal information, or (c) a novel background that was not presented during the impression formation task. Following the presentation of the impression target, participants were briefly presented with a neutral target stimulus (e.g., Chinese ideograph, polygon), which was replaced by a black-and-white pattern mask. Participants' task was to indicate whether they found the neutral target stimulus visually more pleasant or visually less pleasant than average. The modal finding in the AMP is that participants judge the neutral targets more favorably when they were primed with a positive stimulus than when they were primed with a negative stimulus (for a review, see Payne & Lundberg, 2014). Based on evidence that the primes influence participants' judgments of the targets despite their intention not to be influenced by the primes (e.g., Payne et al., 2005), evaluative responses measured by the AMP can be described as unintentional.

Although the dominant use of the AMP in our work may suggest that ABA and ABC renewal may be limited to unintentional evaluations, we have also conducted several studies that investigated the two kinds of renewal effects on measures of intentional evaluation (e.g.,

Gawronski et al., 2014). In these studies, participants were briefly presented with images of the

impression target against (a) the background of initial attitudinal information, (b) the background

of the counterattitudinal information, or (c) a novel background that was not presented during the

impression formation task. Following the presentation of the impression target, participants had

to make a forced-choice judgment by indicating whether their immediate "gut" response to the

presented stimulus was positive or negative. These judgments had to be made under time

pressure, such that participants were asked to provide their response within 1000 ms after the

onset of the impression target. Because participants' responses in this speeded evaluation task

involved a goal to evaluate the impression target, the measured evaluations can be described as

intentional. Yet, they were still spontaneous in the sense that participants were prevented from

deliberating about the presented stimulus or their evaluative response. These measures

replicated the patterns obtained on the AMP.

The dominant focus on spontaneous evaluations raises the question of whether the

reviewed patterns of contextualized attitude change generalize to deliberate evaluations in which

participants have ample time to think about the target object. Gawronski and Cesario (2013)

proposed that the context cues in ABA and ABC renewal function as retrieval cues, in that they

influence which information comes to mind most rapidly upon encountering a target object.

From this perspective, the reviewed patterns of contextualized change may be most pronounced

for immediate evaluative responses to the target object. With increasing delays, deliberate

processing may involve the retrieval of other target-related information, including information

that was learned in other contexts (see Cunningham, Zelazo, Packer, & Van Bavel, 2007;

Wojnowicz, Ferguson, Dale, & Spivey, 2009). In this case, the conflicting pieces of information

may give raise to a state of ambivalence, which may be reflected in people's deliberate

evaluative judgments. Alternatively, people may use their contextual knowledge to disambiguate the valence of the target. In this case, deliberate evaluations may be guided by counterattitudinal information when the target is encountered in the context in which this information was learned.

It is less clear how people would resolve the resulting conflict for deliberate evaluations in Context A and Context C for which there is no contextual information stored in memory. Drawing on research on ease-of-retrieval effects (Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka, & Simons, 1991), one possibility is that people attribute higher validity to information that comes to mind easily and discount the validity of information that requires cognitive effort to be retrieved from memory (Tormala, Petty, & Briñol, 2002). In this case, deliberate evaluations in Contexts A and C may be determined by the initial attitudinal information. Alternatively, it is possible that less accessible information is given equal weight in an integrated judgment that combines all available information regardless of how rapidly it comes to mind. In this case, the patterns obtained for spontaneous evaluations may not necessarily generalize to deliberate evaluations, which may instead reflect a state of ambivalence. Although the correspondence between spontaneous and deliberate evaluations can be moderated by a variety of other factors (for a review, see Hofmann, Gschwendner, Nosek, & Schmitt, 2005), future research may help to clarify the commonalities and differences between spontaneous and deliberate evaluations in their susceptibility to ABA and ABC renewal.[4]

---

[4] Gawronski et al.'s (2015) meta-analysis includes data from an AMP variant that used a longer delay between the presentation of the primes and the targets. The average effect sizes for ABA and ABC renewal revealed by this measure were non-significant and close to zero. Yet, counter to the conclusion that longer processing times reduce the two kinds of renewal effects, Klein and Ratliff (2014) found significant effects of ABA and ABC renewal when participants had unlimited time to provide their evaluative judgments of the target in different contexts.

Associative versus Propositional Learning

According to the representational theory, violation of attitude-related expectancies represents a crucial determinant of contextualized attitude change. Specifically, expectancy-violations are assumed to enhance attention to context, which leads to an integration of the context into the representation of expectancy-violating counterattitudinal information. Yet, different from the proposed role of conscious expectancies, some theories suggest that mere co-occurrences between a neutral conditioned stimulus (CS) and a positive or negative unconditioned stimulus (US) can influence evaluative representations via a process of automatic link formation in the absence of conscious expectancies about their co-occurrence (e.g., Gawronski & Bodenhausen, 2006; Rydell & McConnell, 2006). Whereas the formation of conscious beliefs about states of affairs has been described as *propositional learning*, the formation of associative links on the basis of mere co-occurrences has been described as *associative learning* (Gawronski & Bodenhausen, 2011). Although some researchers (including authors of this article) have rejected the idea of associative learning via automatic link formation, arguing that all forms of learning in humans are mediated by propositional processes (e.g., De Houwer, 2009; Mitchell, De Houwer, & Lovibond, 2009), it is an interesting question whether mere co-occurrences between a CS and a US produce the same patterns of contextualized attitude change. Preliminary findings from our lab suggest that ABA and ABC renewal effects do not occur for repeated CS-US pairings involving neutral images as CSs and pleasant and unpleasant images as USs. In these studies, counterattitudinal pairings effectively reversed the effects of initial attitudinal pairings regardless of the context. However, more evidence is needed to address potential confounds in these studies and to rule out alternative interpretations of the obtained discrepancies.

From Social to Non-social Objects

Another important issue for future research concerns the generalizability of our findings to non-social objects. In the reviewed studies, we primarily focused on social attitudes toward other people. It seems reasonable to assume that the identified principles generalize to non-social objects (e.g., consumer products). To the extent that counterattitudinal experiences with non-social objects can violate attitudinal expectancies about these objects, the content of the counterattitudinal experience should be stored in a contextualized representation, thereby leading to the same patterns of contextualized attitude change. Preliminary evidence for this assumption comes from clinical research on ABA and ABC renewal in the extinction and return of conditioned fear responses (for a review, see Vervliet, Craske, & Hermans, 2013). In line with our findings in the area of social attitudes, a considerable body of evidence suggests that extinction of a conditioned fear response is often limited to the context in which extinction occurred, in that the initial fear response returns in the initial conditioning context (ABA renewal) or a novel context in which the fear-eliciting stimulus had not been encountered before (ABC renewal). Similar parallels have been found for the moderators of ABA and ABC renewal. For example, in line with Gawronski et al.'s (2010) finding that both ABA and ABC renewal effects are eliminated when counterattitudinal information is learned in multiple different contexts, several clinical studies have shown that exposure treatment in multiple contexts attenuates the contextual renewal of fear responses after extinction (e.g., Gunther, Denniston, & Miller, 1998; Vansteenwegen, Vervliet, Iberico, Baeyens, Van den Bergh, & Hermans, 2007; but see Bouton, García-Gutiérrez, Zilski, & Moody, 2006). Despite these overlaps in empirical findings, more research is needed to investigate whether the principles identified for social objects generalize to non-social objects.

Implications

The reviewed evidence for ABA and ABC renewal effects has important implications for the effectiveness of attitude change manipulations in social psychology. A common question in research on attitudes is whether experimentally induced changes in attitudes reflect enduring long-term changes or ephemeral shifts that may dissipate over time (Petty & Cacioppo, 1986). To address this question, participants are often brought back into the lab several days or weeks after the experimental manipulation. To the extent that the initially observed change remains stable over time, it is assumed that the employed manipulation was effective in producing enduring long-term change (e.g., Chaiken, 1980; Devine, Forscher, Austin, & Cox, 2012; Hu, Antony, Creery, Vargas, Bodenhausen & Paller, 2015; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Olson & Fazio, 2006; Petty, 1977). However, the notion of contextualized attitude change suggests that, although the observed changes may be stable over time within the same context, they may not generalize to other contexts. After all, it is possible that effective attitude change in the lab does not generalize to other contexts outside of the lab even when the observed change in the lab is stable over time. Thus, to establish the effectiveness of experimental manipulations in producing enduring attitude change that generalizes across contexts, it is important to include not only delayed follow-up measurements, but also measurements in contexts that are different from the one in which the manipulation took place. These considerations seem particularly important for interventions designed to combat undesired or dysfunctional attitudes (e.g., prejudice).

At a broader level, this conclusion resonates with Mischel and Shoda's (1995) notion of IF-THEN conditionals reflecting idiosyncratic situation-behavior profiles, which implies that individuals may show behavioral consistency over time *within* a particular context, even if

behavioral consistency *across* contexts is low. From this perspective, a person may show inconsistent evaluations of the same object if the object is encountered in different contexts. This inconsistency across contexts does not imply, however, that the person's evaluations of the object are inconsistent over time. After all, evaluations of a given object may be highly consistent over time if they are measured within the same context.

Expanding on these ideas, the notion of contextual renewal offers specific predictions about the conditions under which evaluations should be consistent or inconsistent across contexts (see Gawronski & Cesario, 2013). First, if initial attitudinal information about an object is learned in Context A and then challenged by counterattitudinal information in Context B, evaluations of the object in Context A and Context B should be inconsistent. Whereas evaluations in Context A should reflect the initial attitudinal information, evaluations in Context B should reflect the counterattitudinal information. Second, if initial attitudinal information about an object is learned in Context A and then challenged by counterattitudinal information in another Context B, evaluations of the object in Context B and a novel Context C should also be inconsistent. Whereas evaluations in Context B should reflect the counterattitudinal information, evaluations in Context C should reflect the initial attitudinal information. Third, if initial attitudinal information about an object is learned in Context A and then challenged by counterattitudinal information in Context B, evaluations of the object in Context A and a novel Context C should be consistent. In this case, evaluations should reflect the initial attitudinal information regardless of whether the object is encountered in Context A or Context C. Finally, if initial attitudinal information about an object is learned in Context A and then challenged by counterattitudinal information in the same Context A, evaluations of the object in Context A and a novel Context B should be inconsistent. Whereas evaluations in Context A should reflect the

counterattitudinal information, evaluations in the novel Context B should reflect the initial attitudinal information.[5] According to the representation theory of contextualized attitude change, these predictions should hold unless attention to context is enhanced during the learning of initial attitudinal information or attention to context is reduced during the learning of counterattitudinal information.

Our social-psychological research program on contextualized attitude change also highlights interesting connections to other psychological disciplines. We already mentioned several parallels between our work on social attitudes and clinical research on renewal effects in the treatment of affective disorders (for reviews, see Vervliet, Baeyens, et al., 2013; Vervliet, Craske, et al., 2013). Similar parallels can be found to research on renewal effects in animal learning (for reviews, see Bouton, 2010; Gawronski & Cesario, 2013). The reviewed work on social attitudes provides valuable insights for these two lines of research by offering an integrative theoretical framework that relies on relatively basic, well-established principles in the processing of social information (e.g., enhanced attention to context in response to expectancy-violations). In addition, the current focus on attitudes and evaluation complements previous work on contextual renewal of fear responses, which has almost exclusively focused on physiological responses (see Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004). Conversely, previous work on renewal effects in animal learning can offer valuable insights for social-psychological work on attitudes through a rich body of evidence for boundary conditions and conceptually related phenomena. Finally, the rapidly growing body of work on renewal effects in the clinical domain highlights the broader real-world relevance of the general principles

---

[5] A potential qualification of this hypothesis is implied by evidence for spontaneous recovery, which suggests that delayed testing in Context A may yield evaluations that reflect the valence of the initial attitudinal information (for a review, see Rescorla, 2004).

identified by social-psychological work on attitudes. Thus, research on contextualized attitude change may function as a hub for enhanced cross-talk between areas, and thereby facilitate scientific progress through the unique contributions made by each area of inquiry.

Conclusion

The main goal of the current chapter was to review the findings of an ongoing research program on contextualized attitude change, suggesting that changes in attitudes can be limited to the context in which counterattitudinal information had been acquired. The reviewed findings suggest that, although counterattitudinal information may effectively influence evaluations in the context in which this information was learned, previously formed attitudes may continue to influence evaluations in any other context. According to the representational theory of contextualized attitude change, these effects occur because exposure to expectancy-violating information enhances attention to context, which leads to an integration of the context into the representation of expectancy-violating counterattitudinal information. As a result, counterattitudinal information influences evaluations only in the context in which this information was learned, whereas initial attitudinal information continues to influence evaluations in any other context. Although social-psychological research on contextualized attitude change is still in its infancy, the observation of corresponding patterns in other sub-disciplines corroborates the significance of context for understanding attitude change.

References

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: 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., García-Gutiérrez, A., Zilsik, J., & Moody, E. M. (2006). Extinction in multiple contexts does not necessarily make extinction less vulnerable to relapse. *Behaviour Research and Therapy, 44,* 983-994.

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.

Bouton, M. E., & Ricker, S. T. (1994). Renewal of extinguished responding in a second context. *Animal Learning and Behavior, 22,* 317-324.

Brannon, S. M., & Gawronski, B. (2017a). *In search of a negativity bias in the violation of expectancies*. Manuscript in preparation.

Brannon, S. M., & Gawronski, B. (2017b). *Are there individual differences in contextualized attitude change?* Manuscript in preparation.

Brannon, S. M., & Gawronski, B. (2017c). A second chance for first impressions? Exploring the context (in)dependent updating of implicit evaluations. *Social Psychological and Personality Science, 8,* 275-283.

Brannon, S. M., Sacchi, D. L. M., & Gawronski, B. (2017). (In)consistency in the eye of the beholder: The roles of warmth, competence, and valence in lay perceptions of inconsistency. *Journal of Experimental Social Psychology, 70,* 80-94.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39,* 752-756.

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford Press.

Chiu, C.-Y., Hong, Y.-Y., & Dweck, C. S. (1997). Lay dispositionism and implicit theories of personality. *Journal of Personality and Social Psychology, 73,* 19-30.

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 of the United States of America*, *102*, 12629-12633.

Cialdini, R. B., Trost, M. R., & Newsom, J. T. (1995). Preference for consistency: The

development of a valid measure and the discovery of surprising behavioral implications.

Journal of Personality and Social Psychology, 69, 318-328.

Cone, J., & Ferguson, M. J. (2015). He did what? The role of diagnosticity in revising implicit

evaluations. *Journal of Personality and Social Psychology, 108*, 37-57.

Cone, J., Mann, T. C., & Ferguson, M. J. (in press). Changing our implicit minds: How, when,

and why implicit evaluations can be rapidly revised. *Advances in Experimental Social*

Psychology

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. (2009). The propositional approach to learning as an alternative for association

formation models. *Learning and Behavior, 37,* 1-20.

De Houwer, J., Crombez, G., & Baeyens, F. (2005). Avoidance behavior can function as a

negative occasion setter. *Journal of Experimental Psychology: Animal Behavior*

Processes, 31, 101-106.

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.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. L. (2012). Long-term reduction in

implicit race bias: A prejudice habit-breaking intervention. *Journal of Experimental*

Social Psychology, 48, 1267-1278.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory*, *11*, 549-554.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Fort Worth, TX: Harcourt Brace Jovanovich.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75-109.

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.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*, 889-906.

Försterling, F. (1989). Models of covariation and attribution: How do they relate to the analogy of analysis of variance? *Journal of Personality and Social Psychology, 57,* 615-625.

Gawronski, B. (2007). Attitudes can be measured! But what is an attitude? *Social Cognition, 25,* 573-581.

Gawronski, B. (2012). Back to the future of dissonance theory: Cognitive consistency as a core motive. *Social Cognition, 30,* 652-668.

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., & Bodenhausen, G. V. (2011). The associative-propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44,* 59-127.

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: Cambridge University Press.

Gawronski, B., Hu, X., Rydell, R. J., Vervliet, B., & De Houwer, J. (2015). Generalization and contextualization in automatic evaluation revisited: A meta-analysis of successful and failed replications. *Journal of Experimental Psychology: General, 144,* e50-e64.

Gawronski, B., & Quinn, K. A. (2013). Guilty by mere similarity: Assimilative effects of facial resemblance on automatic evaluation. *Journal of Experimental Social Psychology, 49,* 120-125.

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

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117,*

    21-38.

Gunter, L. M., Denniston, J. C., & Miller, R. R. (1998). Conducting exposure treatment in

    multiple contexts can prevent relapse. *Behaviour Research and Therapy, 36,* 75-91.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A

    cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12,*

    392-407.

Hilton, D. J., & Slugoski, B. R. (1986). Knowledge-based causal attribution: The abnormal

    conditions focus model. *Psychological Review, 93,* 75-88.

Hintzman, D. L. (1986). ''Schema abstraction" in a multiple-trace memory model. *Psychological*

    Review, 93, 411-428.

Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (2005). What moderates implicit-

    explicit consistency? *European Review of Social Psychology, 16,* 335-390.

Hu, X., Antony, J. W., Creery, J. D., Vargas, I. M., Bodenhausen, G. V., & Paller, K. A. (2015).

    Unlearning implicit social biases during sleep. *Science, 348,* 1013-1015.

Jones, E. E. (1991). *Interpersonal perception*. New York: Freeman.

Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to

    stereotyping): Effects of training in the negation of stereotypic associations on stereotypic

    activation. *Journal of Personality and Social Psychology, 78,* 871-888.

Kelley, H. H. (1973). The process of causal attribution. *American Psychologist, 28,* 107-128.

Klauer, K. C. & Wegener, I. (1998). Unraveling social categorization in the "Who said what?"

    paradigm. *Journal of Personality and Social Psychology, 75,* 1155-1178.

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: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin, 37,* 570-583.

Ma, D. S., Correll, J., & Wittenbrink, B. (2016). Context dependency at recall: Decoupling context and targets at encoding. *Social Cognition, 34,* 119-132.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology, 108*, 823-849.

Mann, T. C., & Ferguson, M. J. (2017). Reversing implicit first impressions through reinterpretation after a two-day delay. *Journal of Experimental Social Psychology, 68,* 122-127.

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.

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102,* 246-268.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*, 183-198.

Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology, 65,* 113-131.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32,* 421-433.

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.

Payne, B. K. & Lundberg, K. B. (2014). The Affect Misattribution Procedure: Ten years of evidence on reliability, validity, and mechanisms. *Social and Personality Psychology Compass, 8,* 672-686.

Pearce, J. M., George, D. N., & Redhead, E. S. (1998). The role of attention in the solution of conditional discriminations. In N. A. Schmajuk, & P. C. Holland (1998), *Occasion setting: Associative learning and cognition in animals* (pp. 249-275). Washington, DC: American Psychological Association.

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

Perloff, R. M. (2003). *The dynamics of persuasion: Communication and attitudes in the 21st century*. Mahwah, NJ: Erlbaum.

Petty, R. E. (1977). The importance of cognitive responses in persuasion. *Advances in Consumer Research*, *4*, 357-362.

Petty, R. E., & Cacioppo, J. T. (1986): The elaboration likelihood model of persuasion. In L.

Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205).

New York: Academic Press.

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.

Plaks, J. E. (in press). Implicit theories: Assumptions that shape social and moral cognition.

Advances in Experimental Social Psychology.

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.

Reeder, G. D. (1993). Trait-behavior relations and dispositional inference. *Personality and

Social Psychology Bulletin, 19*, 586-593.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in

interpersonal perception. *Psychological Review, 86,* 61-79.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning & Memory, 11,* 501-509.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the

effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy

(Eds.), *Classical conditioning II* (pp. 64-99)*.* New York, NY: Appleton Century Crofts.

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: Guilford Press.

Rosas, J. M., & Callejas-Aguilera, J. E. (2007). Acquisition of a conditioned taste aversion

    becomes context dependent when it is learned after extinction. *The Quarterly Journal of*

    Experimental Psychology, 60, 9-15.

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.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A

    systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995-

    1008.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007).

    Implicit and explicit attitudes respond differently to increasing amounts of

    counterattitudinal information. *European Journal of Social Psychology, 37,* 867-878.

Schmidt, J. R., De Houwer, J., & Rothermund, K. (2016). The Parallel Episodic Processing

    (PEP) model 2.0: A single computational model of stimulus-response binding,

    contingency learning, power curves, and mixing costs. *Cognitive Psychology, 91,* 82-108.

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition, 25,* 638-656.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991).

    Ease of retrieval as information: Another look at the availability heuristic. *Journal of*

    Personality and Social Psychology, 61, 195-202.

Sherman, J. W., Kruschke, J. K., Sherman, S. J., Percy, E. J., Petrocelli, J. V., & Conrey, F. R.

    (2009). Attentional processes in stereotype formation: A common model for category

    accentuation and illusory correlation. *Journal of Personality and Social Psychology*, *96*,

    305-323.

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.

Spencer-Rodgers, J., Williams, M. J., & Peng, K. (2010). Cultural differences in expectations of change and tolerance for contradiction: A decade of empirical research. *Personality and Social Psychology Review*, *14*, 296-312.

Stiff, J. B., & Mongeau, P. A. (2003). *Persuasive communication* (2nd ed.). New York: Guilford Press.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220-247.

Tamai, N., & Nakajima, S. (2000). Renewal of formerly conditioned fear in rats after extensive extinction training. *International Journal of Comparative Psychology, 13,* 137-147.

Tormala, Z. L., Petty, R. E., & Briñol, P. (2002). Ease of retrieval effects in persuasion: A self-validation analysis. *Personality and Social Psychology Bulletin*, *28*, 1700-1712.

Urcelay, G. P., & Miller, R. R. (2010). Two roles of the context in Pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes, 36,* 268-280.

Vansteenwegen, D., Vervliet, B., Iberico, C., Baeyens, F., Van den Bergh, O., & Hermans, D. (2007). The repeated confrontation with videotapes of spiders in multiple contexts attenuates renewal of fear in spider-anxious students. *Behaviour Research and Therapy, 45,* 1169-1179.

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.

Vervliet, B., Craske, M. G., & Hermans, D. (2013). Fear extinction and relapse: State of the art. *Annual Review of Clinical Psychology, 9,* 215-248.

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: Oxford University Press.

Ye, Y., Tong, Y.-Y., Chiu, C.-Y., & Gawronski, B. (2017). Attention to context during evaluative learning and context-dependent automatic evaluation: A cross-cultural analysis. *Journal of Experimental Social Psychology, 70,* 1-7.

Table 1. Different kinds of renewal effects in the learning of attitudinal and counterattitudinal information.

Effect	Description
ABA Renewal	Learning of initial attitudinal information in Context A
	Learning of counterattitudinal information in Context B
	Initial attitudinal information influences evaluations in Context A
ABC Renewal	Learning of initial attitudinal information in Context A
	Learning of counterattitudinal information in Context B
	Initial attitudinal information influences evaluations in Context C
AAB Renewal	Learning of initial attitudinal information in Context A
	Learning of counterattitudinal information in Context A
	Initial attitudinal information influences evaluations in Context B

Table 2. Means, standard deviations, and correlations to individual difference variables for ABA and ABC renewal effects as a function of valence-order (positive-negative, negative-positive).

	Renewal Effect		Preference for Consistency	Personal Need for Structure	Lay Theory of Personality
	M	*SD*	*r*	*r*	*r*
ABA Renewal					
positive-negative	0.073	0.312	.02 *ns*	.02 *ns*	-.08 *ns*
negative-positive	0.117	0.320	.07 *ns*	.04 *ns*	.07 *ns*
ABC Renewal					
positive-negative	0.056	0.314	.09 *ns*	.03 *ns*	-.09 *ns*
negative-positive	0.048	0.276	-.06 *ns*	.00 *ns*	.01 *ns*

Note. Higher scores on the individual difference measures indicate a stronger preference for consistency, a higher need for structure, and a stronger endorsement of an entity versus incremental theory of personality, respectively ($N = 270$).

Figure Captions

Figure 1. Evaluative responses toward target person as a function of context during the measurement of evaluative responses (context of first learning block vs. context of second learning block vs. novel context) and time (after first learning block vs. after second learning block). The upper panel depicts evaluations when participants learned positive information in the first learning block and negative information in the second learning block (Rydell & Gawronski, 2009, Experiment 3); the lower panel depicts evaluations when participants learned negative information in the first learning block and positive information in the second learning block (Rydell & Gawronski, 2009, Experiment 4). Error bars depict standard errors. Figures adapted from Rydell and Gawronski (2009); reprinted with permission.

Figure 2. Proportions of correct memory for the context (background color) of a target statement as a function of valence of initial information (positive vs. negative) and valence of target statement (positive vs. negative). The dotted line depicts chance-level performance of 10% correct memory judgments. The upper panel shows the results of Gawronski et al. (2014, Experiment 1a); the lower panel shows the results of Gawronski et al. (2014, Experiment 1b). Figures adapted from Gawronski, Ye, Rydell, and De Houwer (2014); reprinted with permission.

Figure 3. Evaluative responses toward target person as a function of valence order in impression formation (positive-negative vs. negative-positive), priming of context salience before the first learning episode (low salience vs. high salience), and context during the measurement of evaluative responses (first context vs. second context vs. novel context). Error bars depict standard errors. Figure adapted from Gawronski, Rydell, Vervliet, and De Houwer (2010); reprinted with permission.
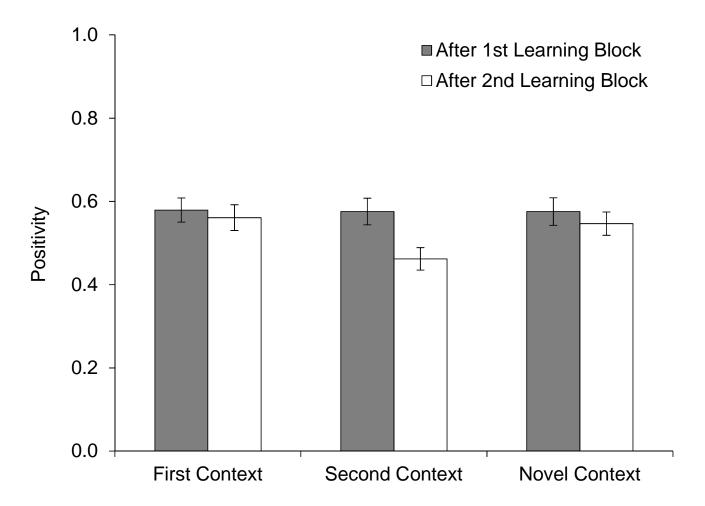
Figure 4. Evaluative responses toward target person as a function of valence order in impression formation (positive-negative vs. negative-positive), number of contexts in the second learning block (single context vs. multiple contexts), and context during the measurement of evaluative responses (first context vs. second context vs. novel context). Error bars depict standard errors. Figure adapted from Gawronski, Rydell, Vervliet, and De Houwer (2010); reprinted with permission.

Figure 5. Two associative network models to explain contextual renewal effects on evaluative responses. Arrows indicate excitatory associations; blocked lines indicate inhibitory associations; open circles indicate memory nodes or representations. In the network model on the left, context leads to renewal effects by constraining the activation of valence in response to the target object. In the network model on the right, context leads to renewal effects by directly activating valence over and above the activation of valence by the target object. The network model on the left was adapted from a figure in Bouton (1994); reprinted with permission.

Figure 6. Evaluative responses toward target individuals as a function of valence order in impression formation (positive-negative vs. negative-positive) and context during the measurement of evaluative responses (first context vs. second context vs. third context). Error bars depict standard errors. Figure adapted from Gawronski, Ye, Rydell, and De Houwer (2014); reprinted with permission.

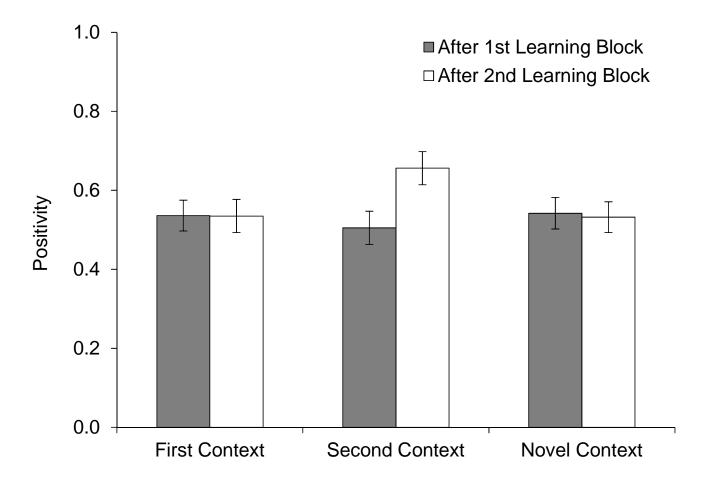**Figure 7.** Evaluative responses toward target individuals as a function of valence order in impression formation task (positive-negative vs. negative-positive), context during the measurement of evaluative responses (first context vs. second context), and time of measurement (before context conditioning vs. after context conditioning). Error bars depict standard errors. Figure adapted from Gawronski, Ye, Rydell, and De Houwer (2014); reprinted with permission.

Figure 8. Evaluative responses as a function of context conditioning (first-positive, second-negative vs. first-negative, second-positive), context during the measurement of evaluative responses (first context vs. second context), and time of measurement (before context conditioning vs. after context conditioning). Error bars depict standard errors. Figure adapted from Gawronski, Ye, Rydell, and De Houwer (2014); reprinted with permission.

Figure 9. Images used to manipulate perceptual versus conceptual similarity between contexts. Images shown in the same row are perceptually similar but conceptually distinct; images in the same column are conceptually similar but perceptually distinct; images displayed diagonally are both perceptually and conceptually distinct. Figure adapted from Gawronski, Ye, Rydell, and De Houwer (2014); reprinted with permission.

Figure 10. Evaluative responses toward target person as a function of valence order in impression formation (positive-negative vs. negative-positive) and context during the measurement of evaluative responses (first context vs. second context vs. perceptually similar to second context vs. conceptually similar to second context vs. distinct from second context). Error bars depict standard errors. Figure adapted from Gawronski, Ye, Rydell, and De Houwer (2014); reprinted with permission.

Figure 11. Proportion of correct memory for the context (background color) of a target statement as a function of target statement (expectancy-congruent vs. expectancy-incongruent) and country (Canada vs. Singapore). The dotted line depicts chance-level performance of 10% correct memory judgments. Error bars depict standard errors. Figure adapted from Ye, Tong, Chiu, and Gawronski (2017); reprinted with permission.

Figure 12. Evaluative responses toward target person as a function of new information (supportive vs. reinterpretation), measurement time (Time 1 vs. Time 2), and measurement
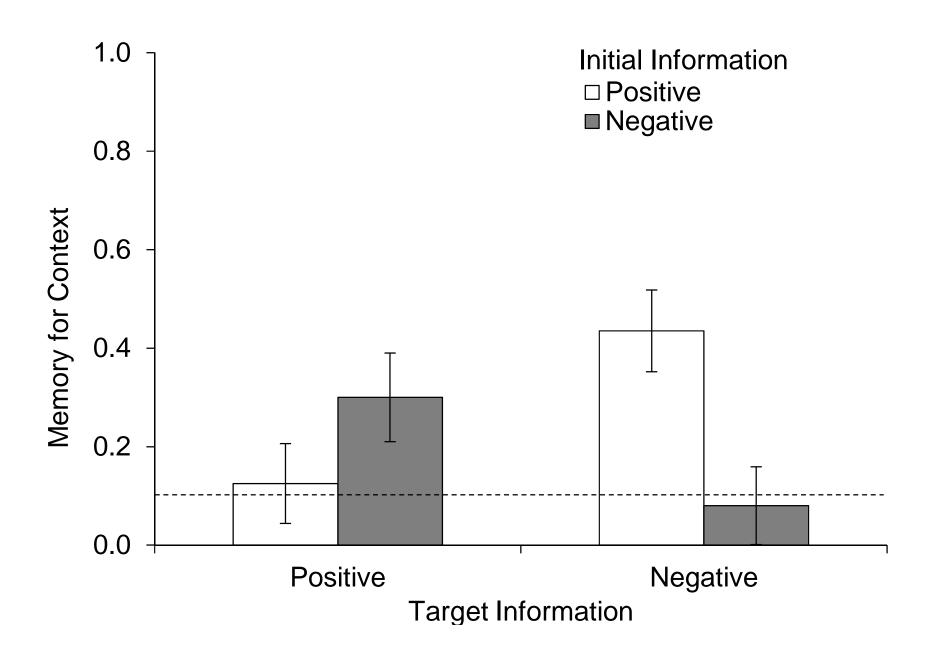
context (first learning context vs. second learning context vs. novel context). Error bars depict

standard errors. Figure adapted from Brannon and Gawronski (2017c); reprinted with
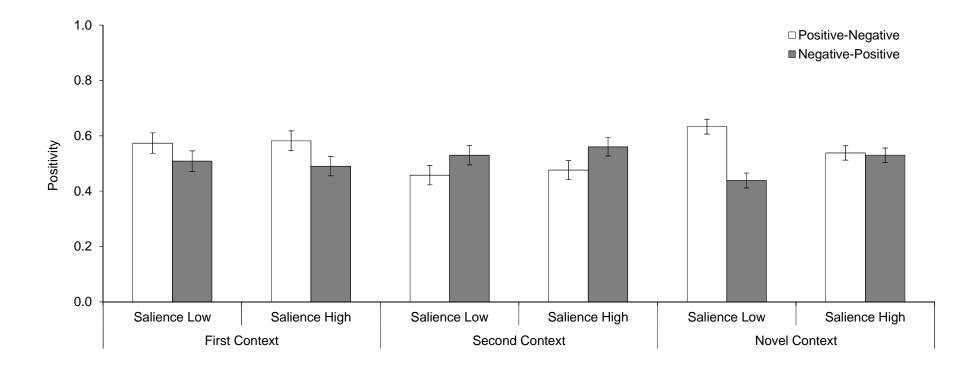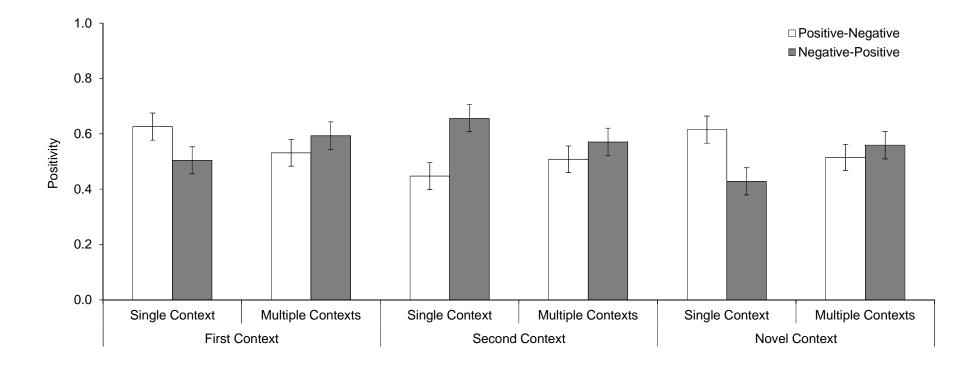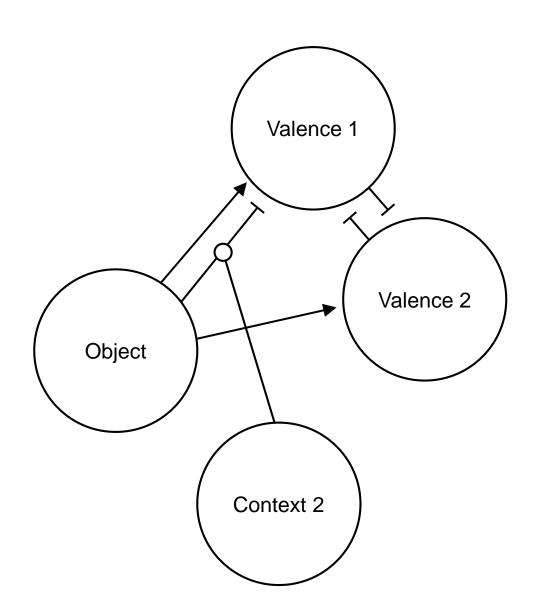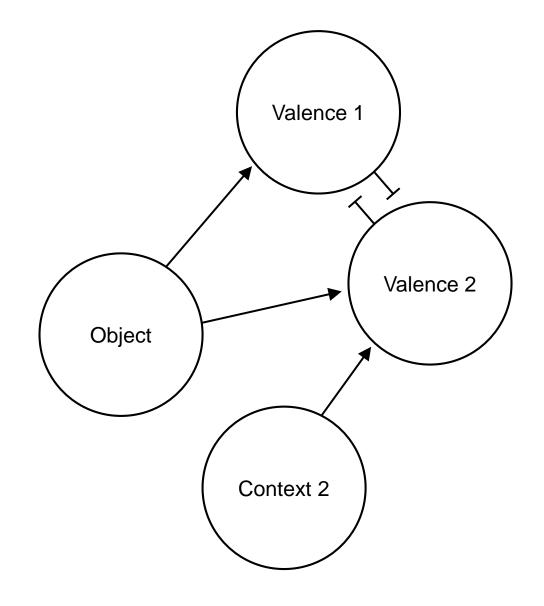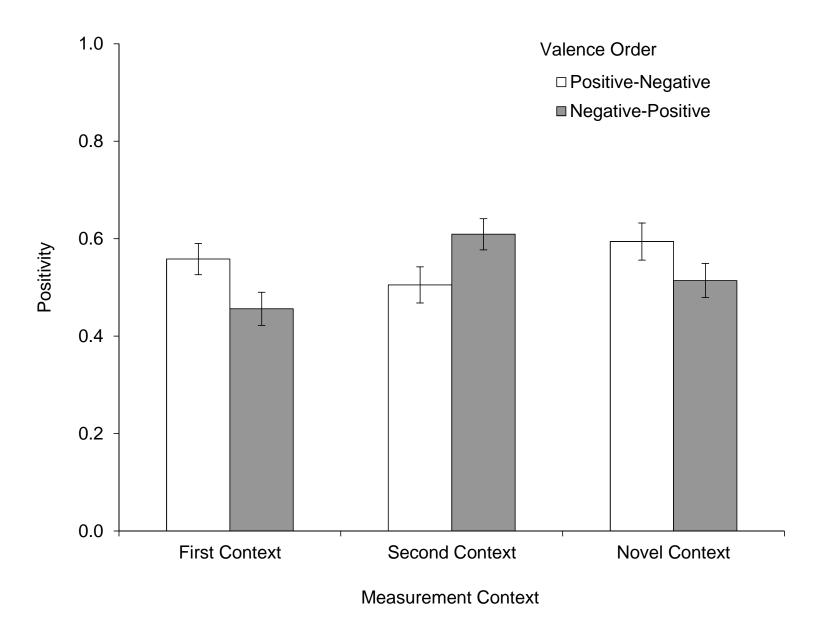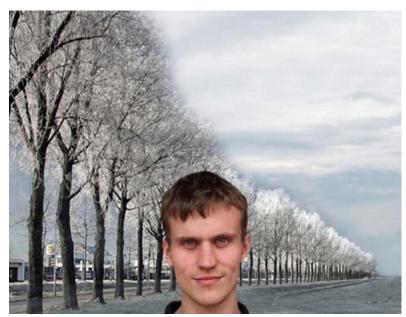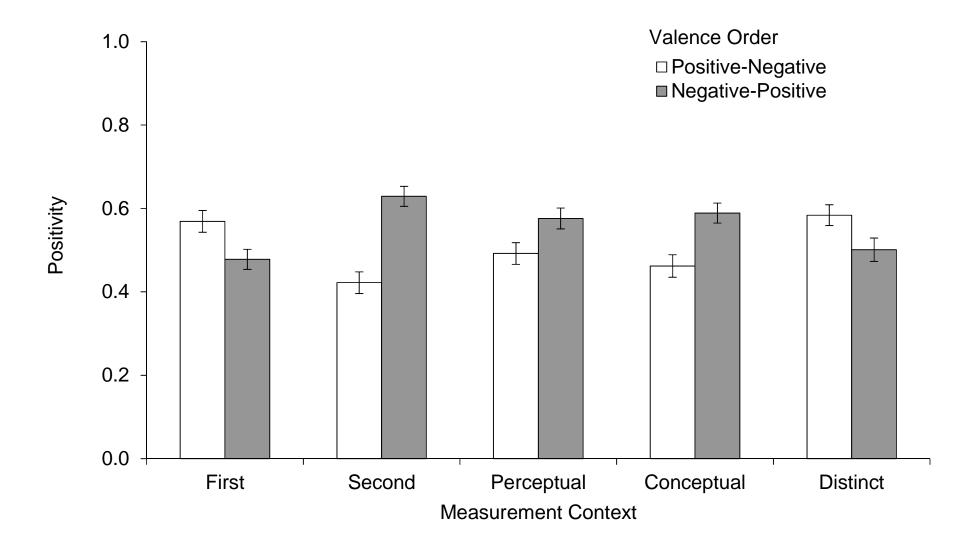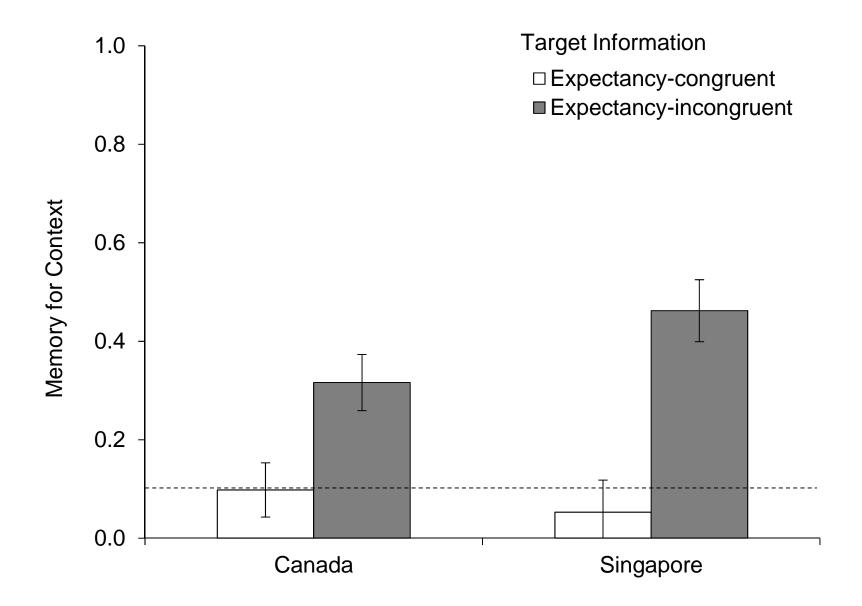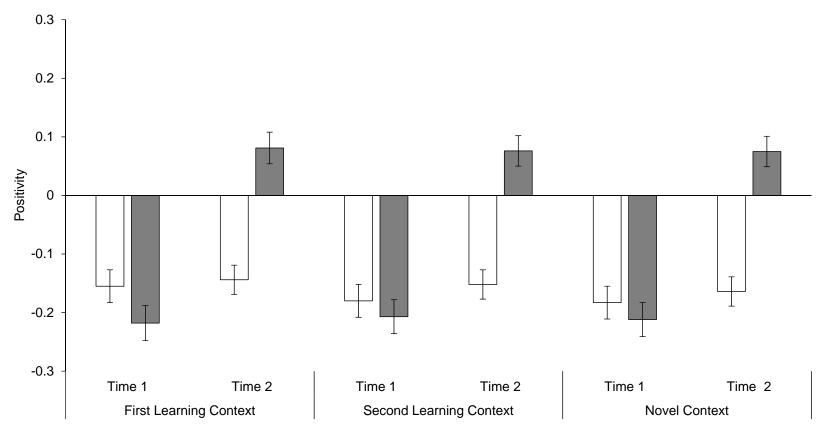
permission.

Figure 13. Evaluative responses toward target person as a function of new information

(diagnostic vs. neutral), measurement time (Time 1 vs. Time 2), and measurement context (first

learning context vs. second learning context vs. novel context). Error bars depict standard errors.

Figure adapted from Brannon and Gawronski (2017c); reprinted with permission.