CHAPTER TWO

The Associative–Propositional Evaluation Model: Theory, Evidence, and Open Questions

Bertram Gawronski* *and* Galen V. Bodenhausen†

Contents

* Department of Psychology, The University of Western Ontario, Social Science Centre, London, Ontario, Canada
† Department of Psychology, Northwestern University, Chicago, Illinois, USA

Advances in Experimental Social Psychology, Volume 44
ISSN 0065-2601, DOI: 10.1016/B978-0-12-385522-0.00002-0

Abstract

A central theme in contemporary psychology is the distinction between implicit and explicit evaluations. Research has shown various dissociations between the two kinds of evaluations, including different antecedents, different consequences, and discrepant evaluations of the same object. The associative–propositional evaluation (APE) model accounts for these dissociations by conceptualizing implicit and explicit evaluations as the outcomes of two qualitatively distinct processes. Whereas implicit evaluations are described as the outcome of associative processes, explicit evaluations represent the outcome of propositional processes. Associative processes are further specified as the activation of mental associations on the basis of feature similarity and spatiotemporal contiguity; propositional processes are defined as the validation of activated information on the basis of logical consistency. The APE model includes specific assumptions about the mutual interplay between associative and propositional processes, implying a wide range of predictions about symmetric and asymmetric changes in implicit and explicit evaluations. The current chapter reviews the conceptual and empirical assumptions of the APE model and evidence in support of its predictions. In addition, we discuss conceptual and empirical challenges for the APE model and various directions for future research on implicit and explicit evaluation.

1. Introduction

The human mind has a ubiquitous tendency to evaluate objects and events in the environment. Such evaluations can be the result of thoughtful, deliberate processes that aim at integrating various stimulus characteristics in a coherent evaluative judgment. At the same time, environmental stimuli may elicit evaluative responses instantaneously and in the absence of a conscious intention to evaluate these stimuli. In many situations, the evaluations derived from these two kinds of responses may be similar. However, in other situations, the two kinds of responses may be in conflict. For instance, we may sometimes experience a spontaneous feeling of attraction for another person even though we know that this person is not a good match; or we may experience spontaneous negative reactions to members of stigmatized groups even though we have a strong aspiration to be

unprejudiced. In social psychology, such differences between spontaneous and deliberate evaluations are captured by the distinction between implicit and explicit evaluations. Whereas *explicit evaluations* are typically equated with self-reported evaluative judgments, *implicit evaluations* are inferred from an individual's performance on indirect measurement procedures, such as the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998), sequential priming tasks (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Payne, Cheng, Govorun, & Stewart, 2005; Wittenbrink, Judd, & Park, 1997), or various other kinds of performance-based measures (for overviews, see Gawronski, 2009; Gawronski, Deutsch, & Banse, in press).[1]

An important question in this line of research concerns the mental processes that underlie explicit and implicit evaluations and the causal factors that lead to changes in the two kinds of responses. In the current chapter, we provide an overview of our associative–propositional evaluation (APE) model (Gawronski & Bodenhausen, 2006a,b, 2007a), which was particularly designed to answer these questions. Toward this end, we first outline the core assumptions of the APE model and then review the implications of the APE model for changes in explicit and implicit evaluations. Expanding on this review, we address particular aspects of our model that have sometimes led to misunderstandings. Finally, Section 6 discusses some challenges for the APE model and open questions for future research.

2. Associative and Propositional Processes Underlying Evaluation

The central notion of the APE model is that implicit and explicit evaluations represent the behavioral outcomes of two qualitatively distinct mental processes. Whereas implicit evaluations are the behavioral outcome of *associative processes*, explicit evaluations represent the behavioral outcome of *propositional processes*. In general terms, associative processes are defined as the *activation* of mental associations in memory, which is assumed to be driven by spatiotemporal contiguity between stimuli and the similarity between the features of input stimuli and available memory representations.

[1] Following terminological suggestions by De Houwer (2006), we use the terms *direct* and *indirect* to describe characteristics of measurement procedures; the terms *explicit* and *implicit* are used to describe different kinds of evaluative responses. Whereas the term *explicit evaluation* is used to refer to self-reported evaluative judgments assessed by direct measurement procedures, the term *implicit evaluation* is used to refer to evaluative responses assessed by performance-based, indirect measurement procedures. Note that in the current chapter the terms *explicit* and *implicit* are meant to describe responses on different types of measurement procedure instead of the conscious versus unconscious nature of these responses. Claims about the consciousness of different kinds of evaluative responses are discussed in more detail in Section 3.

In contrast, propositional processes are defined as the *validation* of the information that is implied by activated associations, which is assumed to be guided by the principles of logical consistency.[2] Thus, the most important feature that distinguishes between associative and propositional processes is their (in)dependency of subjective truth or falsity. Whereas associations can be activated in memory regardless of whether the information implied by these associations is considered accurate or inaccurate, propositional processes are inherently concerned with the validity of activated information.

According to the APE model, which associations will be activated in response to a particular object depends on (a) the preexisting structure of associations in memory and (b) the overall set of input stimuli. This assumption resembles the notion of pattern activation in connectionist models (see Smith, 1996), which refers to the idea that association activation is not an all-or-none process, such that encountering a given object activates each and every mental association related to that object. Instead, objects tend to activate only a limited subset of all object-related associations that are available in memory. Which subset of associations is activated in response to a given object is assumed to depend on the overall configuration of input stimuli. For example, encountering a Black person in a jazz bar may activate the stereotypical attribute *musical*, whereas the same Black person may activate the stereotypical attribute *criminal* if that person is encountered in a dark alley (e.g., Barden, Maddux, Petty, & Brewer, 2004; Wittenbrink, Judd, & Park, 2001). These considerations imply that the same attitude object may activate different patterns of associations in memory depending on the particular context in which the object is encountered (see Barsalou, 1982). Yet, the activation of associations is not entirely context-driven, as associative processes are constrained by the preexisting structure of associations in memory. After all, different contexts can modulate the activation of concepts in response to a given object only if these concepts are part of the associative representation of that object.

Applied to the distinction between implicit and explicit evaluations, we argue that the overall valence of the concepts that are activated in response to a given object determines the evaluative quality of an individual's affective gut reaction to that object, which in turn drives responses on measures of *implicit evaluations*. In many cases, people may use their affective gut reaction to an object as a basis for an endorsed evaluative judgment about that object, such that they may simply report the evaluative quality of their gut response on measures of *explicit evaluations*. Specifically, we argue that affective gut reactions are translated into the format of a propositional

[2] Note that the term *logical consistency* is intended to refer more broadly to subjective consistency resulting from any kind of inferential rule that a person considers valid, rather than to strict logical consistency in terms of normative syllogistic rules.

statement (e.g., a negative affective reaction toward object X is transformed into propositional statements such as "I dislike X" or "X is bad"). To the extent that this proposition is consistent with other propositional beliefs that are considered relevant for an evaluative judgment, it may be endorsed in a verbally reported explicit evaluation. If, however, the propositional evaluation implied by the affective gut response is inconsistent with other salient propositions that are considered relevant, the inconsistency has to be resolved to avoid aversive feelings of cognitive dissonance (Festinger, 1957). In such cases, consistency may be restored either by rejecting one of the involved propositions (i.e., reversing the subjective truth value of that proposition) or by searching for an additional proposition that resolves the inconsistency.[3] In the former case, people may sometimes reject the propositional evaluation implied by their affective gut response. As the mere rejection of an affective gut response does not necessarily deactivate the mental associations that gave rise to this response, the common outcome is a dissociation between implicit and explicit evaluations. In such cases, implicit evaluations will reflect the evaluative quality of the original gut response, whereas explicit evaluations will reflect whatever evaluation is implied by the accepted set of propositions. If, however, consistency is restored by other means—for instance by rejecting one of the other propositions or by identifying a new proposition that resolves the inconsistency—the affective gut response may still serve as a basis for an endorsed evaluative judgment, thereby leading to corresponding implicit and explicit evaluations.

2.1. Interactions between associative and propositional processes

According to the APE model, associative and propositional processes do not operate in isolation but mutually interact with each other (Gawronski & Bodenhausen, 2006a). On one hand, associative processes tend to influence propositional processes, given that processes of propositional validation generally operate on the information that is implied by momentarily activated associations. On the other hand, propositional processes can influence associative processes, given that processes of propositional reasoning may create or activate new associations in the course of validating activated information. Nevertheless, associative and propositional processes can lead to different behavioral outcomes when cognitive inconsistency leads to a rejection of the affective gut response resulting from activated associations.

[3] The first strategy—rejecting one of the involved propositions—refers to Festinger's (1957) general notion of changing a cognitive element (e.g., attitude change, behavior change); the second strategy—searching for an additional proposition—represents the notion of adding a cognitive element (e.g., search for consonant information, trivialization).

2.1.1. "Bottom-up" influences of associative on propositional processes

According to the APE model, a central determinant of whether the information implied by activated associations is accepted in the process of propositional validation is the consistency of that information with other salient propositional beliefs that are considered relevant for a judgment (Gawronski & Bodenhausen, 2006a). To illustrate the role of consistency in the process of propositional validation, consider a case in which the activation of negative associations related to the social stereotype of African Americans elicits a negative gut reaction in response to Black people. According to the APE model, this gut response may be translated into a corresponding propositional evaluation (e.g., "I dislike Black people"), which may be assessed for its validity on the basis of its consistency with other propositional beliefs that are considered relevant for an evaluative judgment (Gawronski, Peters, Brochu, & Strack, 2008). To the extent that the propositional evaluation implied by the affective gut response is consistent with these beliefs, it may be used as a basis for a verbally reported judgment, implying a negative response for both implicit and explicit evaluations. If, however, the propositional evaluation implied by the affective gut response is inconsistent with other relevant propositional beliefs, this inconsistency will have to be resolved to avoid aversive feelings of cognitive dissonance (Festinger, 1957). In general, propositional evaluations of a given object may be assessed for their consistency with (a) nonevaluative beliefs about the world and (b) propositional evaluations of other attitude objects (Jones & Gerard, 1967). In the current example, these two kinds of propositions may include propositional beliefs about the prevalence of racial discrimination and propositional evaluations of discriminatory behavior (Gawronski, Peters, Brochu, et al., 2008). More specifically, the overall set of judgment-relevant elements may include the following three propositions:

1. "I dislike Black people."
2. "Black people represent a disadvantaged group."
3. "Negative evaluations of disadvantaged groups are wrong."

Taken together, these three propositions are inconsistent with each other in that they cannot be endorsed at the same time without violating the basic notion of cognitive consistency (see Fig. 2.1A). Proposition 1 is inconsistent with the joint implication of Propositions 2 and 3; Proposition 2 is inconsistent with the joint implication of Propositions 1 and 3; and Proposition 3 is inconsistent with the joint implication of Propositions 1 and 2. To the extent that consistency is achieved through a rejection of either Proposition 2 (see Fig. 2.1C) or Proposition 3 (see Fig. 2.1B), the negative evaluation of Proposition 1 may be endorsed in a verbally reported evaluative judgment. In these cases, implicit and explicit evaluations should reveal

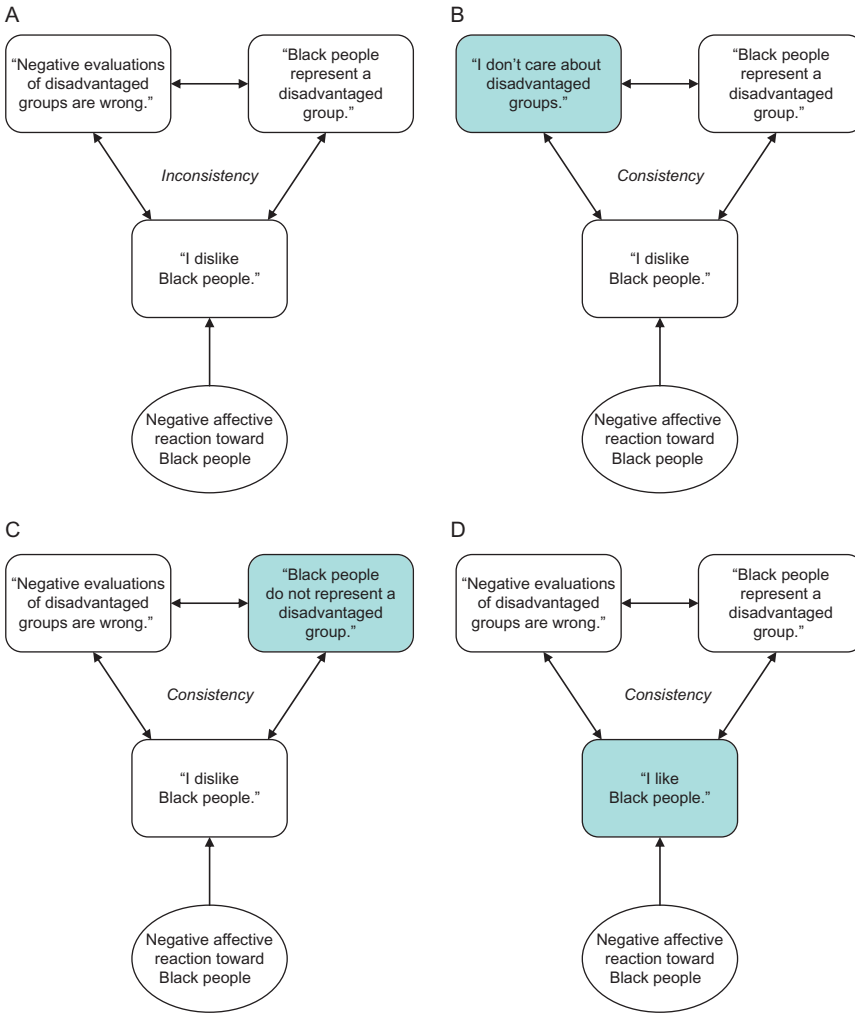

Figure 2.1 Interplay between affective reactions (circles) and propositional beliefs (squares) in racial prejudice against Black people. (A) depicts the case of an inconsistent belief system; (B–D) depict consistent belief systems, involving either a reliance on affective reactions for evaluative judgments (B, C) or a rejection of affective reactions for evaluative judgments (D). Adapted from Gawronski, Peters, Brochu, et al. (2008), reprinted with permission.

corresponding responses, such that both reflect the negativity of the affective gut response resulting from activated associations. If, however, consistency is achieved through a rejection of Proposition 1 (see Fig. 2.1D), people may endorse a neutral or positive evaluation in their verbally reported judgments. Importantly, merely reversing the subjective truth

value of Proposition 1 does not necessarily deactivate the associations that gave rise to the affective gut response that built the foundation for this proposition. As a result, a rejection of Proposition 1 may lead to a dissociation between implicit and explicit evaluations, such that implicit evaluations may still reflect the negativity of the affective gut response, whereas explicit evaluations may reflect the neutral or positive evaluation that is inferred in the propositional validation process.

Evidence for these assumptions comes from a series of studies by Gawronski, Peters, Brochu, et al. (2008). To investigate the role of cognitive consistency in prejudice-related belief systems, participants were asked to complete measures of implicit and explicit evaluations of Black people as well as self-report measures of perceived discrimination and evaluative beliefs about discriminatory behavior. In line with the claim that the reliance on affective gut reactions in making evaluative judgments depends on the consistency of the evaluation implied by the affective response with other relevant information, correlations between implicit and explicit evaluations were moderated by the interaction of perceptions of discrimination and anti-discriminatory beliefs. Specifically, implicit and explicit negativity toward Black people showed significant positive correlations when (a) perceptions of discrimination were high and anti-discriminatory beliefs were weak (see Fig. 2.1B) and (b) perceptions of discrimination were low and anti-discriminatory beliefs were strong (see Fig. 2.1C). However, implicit and explicit evaluations showed tendencies for negative correlations when (c) perceptions of discrimination were high and anti-discriminatory beliefs were strong (see Fig. 2.1D). In other words, whether or not participants relied on their affective gut reactions toward Black people in making evaluative judgments about Black people depended on the consistency of the evaluation implied by their gut response with other relevant beliefs, in this case perceptions of discrimination and anti-discriminatory beliefs (for corresponding findings regarding weight prejudice, see Brochu, Gawronski, & Esses, in press).

To avoid a common misunderstanding, it is important to note that the APE model does not propose a separate storage of propositions in memory. Instead, all information is assumed to be stored in the form of associations (Gawronski & Bodenhausen, 2006a; Footnote 3). For instance, in the above example the proposition "Black people represent a disadvantaged group" may be based on an association between the concepts *Black people* and *disadvantaged group*. Similarly, the proposition "Negative evaluations of disadvantaged groups are wrong" may be based on negative associations related to the behavioral concept *negative evaluations of disadvantaged groups*. If an association between two concepts is activated, its content will be regarded as valid unless it is inconsistent with other information that is momentarily activated. In other words, the default mode of propositional reasoning is the acceptance of activated associations (Gilbert, 1991). If the

acceptance of all activated associations produces a set of propositions that is inconsistent (e.g., Fig. 2.1A), people will try to resolve this inconsistency either by rejecting one of the involved propositions as false or by searching for a new proposition that resolves the inconsistency (Gawronski & Strack, 2004).

2.1.2. "Top-down" influences of propositional on associative processes

Even though propositional processes are assumed to operate on the information implied by activated associations, propositional processes can also influence associative processes, such that processes of propositional reasoning may create or activate new associations in the course of validating activated information (Gawronski & Bodenhausen, 2006a). An important aspect in this regard is the distinction between affirmation versus negation (see Gilbert, 1991). Specifically, we argue that merely negating a particular proposition (i.e., reversing its truth value) is insufficient to deactivate the association(s) underlying this proposition. On the contrary, negations often lead to ironic effects, such that the activation level of the underlying association is enhanced rather than reduced (Wegner, 1994). For instance, negating the proposition "old people are bad drivers" may enhance the association between the concepts *old people* and *bad drivers*, thereby leading to a dissociation between the outcomes of associative and propositional processes (e.g., Deutsch, Gawronski, & Strack, 2006). Thus, cognitive limits in processing negations can be interpreted as the critical proximal factor underlying inconsistency-related dissociations between implicit and explicit evaluations, given that inconsistency-related rejections of affective gut responses usually operate through a negation of these responses (Gawronski, Strack, & Bodenhausen, 2009).

This situation is different if processes of propositional reasoning involve an affirmation of new information. In such cases, propositional processes may influence associative processes through the activation or creation of associations. For instance, affirming the proposition "old people are good drivers" may enhance the association between the concepts *old people* and *good drivers*, thereby increasing the correspondence between the responses resulting from associative and propositional processes (e.g., Deutsch et al., 2006). Thus, the critical factor that determines the impact of propositional on associative processes is whether propositional processes involve an affirmation or negation of evaluative information. Whereas affirmations lead to influences in line with the inferred propositional conclusion, negations often lead to ironic effects, such that negations of a particular proposition tend to enhance rather than reduce the activation level of the underlying association.

Evidence for these assumptions comes from a study by Gawronski, Deutsch, Mbirkou, Seibt, and Strack (2008) who investigated the differential effectiveness of affirmation versus negation training on implicit

evaluations (see Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000). In their study, participants were presented with Black and White faces that were paired with trait words related to either the negative stereotype of Black people or the positive stereotype of White people. Half of the participants were asked to press a key labeled "*NO*" each time they saw a face-trait combination that was consistent with the cultural stereotype of Blacks and Whites (i.e., negation of stereotype). The remaining half were asked to press a key labeled "YES" each time they saw a face-trait combination that was inconsistent with the cultural stereotype of Blacks and Whites (i.e., affirmation of counterstereotype). Consistent with our assumptions, extended training in the affirmation of counterstereotypes led to a reduction in implicit preferences for Whites over Blacks. In contrast, extended training in the negation of stereotypes enhanced rather than reduced implicit preferences for Whites over Blacks (see Fig. 2.2).

Despite the generality of our claims about affirmation and negation, it is important to note two exceptions under which negations may not necessarily produce a dissociation between implicit and explicit evaluations. First, dissociations between implicit and explicit evaluations may not occur if the evaluative meaning of a particular negation is highly overlearned through frequent processing. In such cases, the negated concept may be stored as a separate unit in associative memory with a direct link to its evaluative meaning. For instance, frequent processing of the phrase *no problem* may

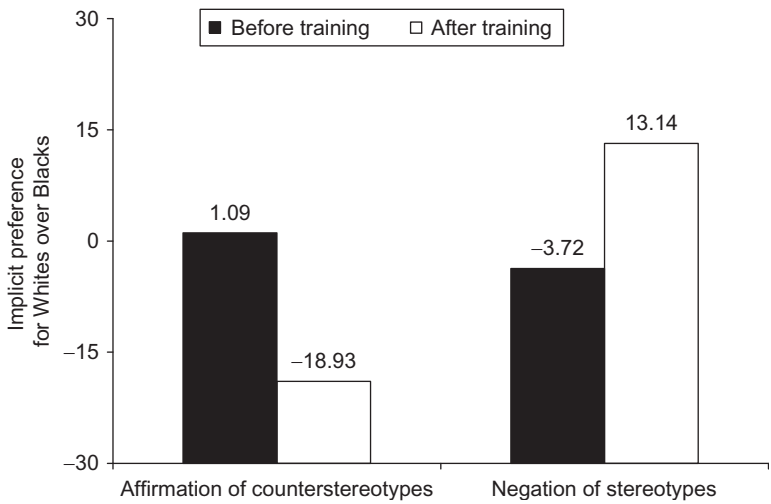

Figure 2.2 Implicit preferences for Whites over Blacks as a function of training task (affirmation of counterstereotypes vs. negation of stereotypes) and time of measurement (before vs. after training). Adapted from Gawronski, Deutsch, Mbirkou, et al. (2008), reprinted with permission.

produce a corresponding unit in associative memory over time, and this unit may be directly linked to the positive meaning implied by the negation. Note, however, that such cases of instance learning (Logan, 1988) do not imply a generalization to other negations, as might be expected for cases of procedural learning (Anderson, 1993). Applied to the current example, this assumption implies that processing the highly overlearned phrase *no problem* may activate positive associations, but processing the less frequent phrase *no cockroach* may still activate negative associations (e.g., Deutsch et al., 2006; Rozin, Markwith, & Ross, 1990).

Second, dissociations between implicit and explicit evaluations may not occur if a particular negation directly activates a specific referent that affirmatively represents the meaning of the negated term (e.g., the phrase *no war* may directly activate the specific referent *peace*; e.g., Mayo, Schul, & Burnstein, 2004). In such cases, processing the negation may activate the specific referent representing the negated meaning, thereby leading to corresponding responses for explicit and implicit evaluations. Note, however, that the activation of specific referents is again limited to specific instances and therefore does not generalize to other negations that are not directly associated with a specific referent. Thus, even though processing the term *no war* may activate positive associations by virtue of its specific referent *peace*, processing the term *no cockroach* may still activate negative associations, as it does not have a specific referent that is directly associated with that term.

2.2. Cognitive elaboration

A central variable in virtually all dual-process theories of attitudes and evaluation is the degree of cognitive elaboration (e.g., Chaiken, Liberman, & Eagly, 1989; Fazio, 1990; Petty & Cacioppo, 1986; for a review, see Gawronski & Creighton, in press), which is defined as the amount of thought that is devoted to an attitude object. The APE model agrees with the contention that cognitive elaboration represents a critical determinant of evaluative responses. However, the APE model makes a number of additional assumptions about the role of cognitive elaboration that deviate from earlier theorizing.

A first issue concerns cognitive elaboration as a moderator of the correspondence between implicit and explicit evaluations. Several attitude theories suggest that implicit and explicit evaluations should show stronger relations when cognitive elaboration during the generation of an evaluative judgment is low rather than high (e.g., Fazio, 2007; Wilson, Lindsey, & Schooler, 2000) and the available evidence is mostly consistent with this assumption (e.g., Florack, Scarabis, & Bless, 2001; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Koole, Dijksterhuis, & van Knippenberg, 2001; LeBel, 2010; Ranganath, Smith, & Nosek, 2008). The APE model agrees with this general claim, albeit with some important

qualifications. In our view, the correspondence between explicit and implicit evaluations primarily depends on the (in)consistency of the evaluation implied by one's affective gut response with other judgment-relevant information. To the extent that inconsistency becomes more likely as a function of the amount of information that is considered for the judgment, higher degrees of cognitive elaboration will often reduce the correspondence between implicit and explicit evaluations. However, what ultimately reduces the correspondence between implicit and explicit evaluations is not cognitive elaboration *per se*, but the inconsistency of the affective gut response with other momentarily considered information. Thus, if enhanced levels of cognitive elaboration lead to a recruitment of information that is consistent with the affective gut response, the correspondence between implicit and explicit evaluations may be unaffected. Moreover, if extensive elaboration helps to identify information that supports the validity of an affective gut response (e.g., Galdi, Gawronski, Arcuri, & Friese, 2010), enhanced elaboration may in fact increase rather than decrease the correspondence between implicit and explicit evaluations. In other words, what moderates the relation between implicit and explicit evaluations is not cognitive elaboration *per se*, but the consistency of the additionally recruited information with the evaluation implied by one's affective gut response. These predictions were confirmed in a direct test by Whitfield (2009), who demonstrated that enhanced levels of cognitive elaboration arising from rational analysis reduced the correspondence between implicit and explicit evaluations only when the generated thoughts were inconsistent, but not when they were consistent, with the implicit evaluation.

Another issue in this context concerns the role of cognitive elaboration during the encoding of new information. Several theories of attitude formation and change argue that cognitive elaboration during the encoding of evaluative information determines the effectiveness of different types of information in influencing attitudes, and ultimately the stability of the newly formed attitudes over time (e.g., Chaiken et al., 1989; Petty & Cacioppo, 1986). Specifically, it is assumed that more extensive elaboration during the encoding of evaluative information produces attitudes that are relatively strong, less susceptible to counterpersuasion, and more predictive of behavior (e.g., Petty, Cacioppo, & Schumann, 1983). Even though we did not address the role of cognitive elaboration during the encoding of new information in the original presentation of the APE model (Gawronski & Bodenhausen, 2006a), it is important to distinguish between encoding-related and judgment-related effects of cognitive elaboration. As outlined above, enhanced elaboration during the expression of evaluative judgments may often decrease the correspondence between implicit and explicit evaluations through the consideration of additional information that may be inconsistent with the evaluation implied by one's affective gut response. Yet, the role of cognitive elaboration during the encoding of evaluative information is quite different, in that it may influence (a) the

generation of arguments in favor and/or against newly acquired propositional information (Greenwald, 1968) and (b) the strength of the newly formed associations resulting from this validation process (Craik & Lockhart, 1972). As the two factors influence explicit and implicit evaluations in the same direction, enhanced elaboration during the encoding of evaluative information may therefore enhance rather than reduce the correspondence between explicit and implicit evaluations, at least for newly acquired propositional information.

2.3. Motivational influences

Similar to the role of cognitive elaboration, it is often assumed that the relation between implicit and explicit evaluations is moderated by motivational factors. For instance, it has been argued that implicit evaluations reveal thoughts, opinions, or feelings that people are motivated to conceal in explicit evaluations because of self-presentation or social desirability concerns (for a critical discussion, see Gawronski, LeBel, & Peters, 2007). Supporting this assumption, several studies have shown that the relation between implicit and explicit evaluations of minority groups is moderated by individual differences in the motivation to control prejudiced reactions (e.g., Degner & Wentura, 2008; Dunton & Fazio, 1997; Gawronski, Geschke, & Banse, 2003). In these studies, individuals with a low motivation to control prejudiced reactions revealed positive correlations between implicit and explicit evaluations of minority groups, whereas individuals with a high motivation to control prejudiced reactions showed either no or negative correlations.

According to the APE model, motivational factors can certainly play an important role in this regard. Yet, their impact is more distal and mediated by cognitive processes, in particular by the consistency between propositionally represented goals and propositional evaluations implied by affective gut responses. To illustrate this assumption, consider the previous example of racial prejudice against Black people (see Gawronski, Peters, Brochu, et al., 2008). In this example, we assumed that the activation of negative stereotypical associations elicits a negative gut response to Black people, which in turn may be translated into the propositional evaluation "I dislike Black people." As outlined above, this proposition may be inconsistent with other propositions that are considered relevant for an evaluative judgment, such as "Black people represent a disadvantaged group" and "Negative evaluations of disadvantaged groups are wrong" (see Fig. 2.1A). From the perspective of the APE model, the latter proposition may be interpreted as a propositionally represented goal to the extent that an individual is committed to the action implied by the evaluation (i.e., "I don't want to evaluate disadvantaged groups negatively"). Yet, it may not necessarily lead to a rejection of the negative gut response as a basis for an evaluative judgment, if the inconsistency between the three propositions is reduced by denying the continued discrimination of Black people (see Fig. 2.1C; cf. McConahay, 1986). In line with

this assumption, we repeatedly found that high levels in the motivation to control prejudiced reactions reduced the correlation between implicit and explicit evaluations of minority groups only when these groups were perceived to be a target of discrimination (for a meta-analysis, see Gawronski, Brochu, Sritharan, & Strack, in press). However, individuals with a high motivation to control prejudiced reactions showed a positive correlation between explicit and implicit evaluations when perceptions of discrimination were low (see also Franco & Maass, 1999). These results suggest that the moderating effect of motivational factors—such as the motivation to control prejudiced reactions—on the relation between implicit and evaluations is more distal, in that their impact depends on the consistency of all information that is momentarily considered for an evaluative judgment, including propositionally represented goals, propositional evaluations implied by affective gut responses, and any other evaluative or declarative information that may be regarded as relevant.

Another important aspect related to motivational influences concerns motivated reasoning. In terms of the APE model, motivated reasoning can be understood as a particular influence on the process of propositional validation, such that people may sometimes be motivated to believe in the validity of particular propositions (i.e., directional motivated cognition). In such cases, inconsistency will usually be resolved by rejecting any other propositions that are inconsistent with the desired proposition or by searching for information that confirms the validity of the desired proposition (Ditto & Lopez, 1992). However, such cases of motivated reasoning are still constrained by the principles of cognitive consistency, such that the overall set of accepted propositions needs to form a consistent set of beliefs (Kunda, 1990). In fact, inconsistency can often be resolved in multiple ways and this flexibility provides room for influences of motivated reasoning, such that people may affirm the validity of desirable propositions and negate the validity of undesirable propositions.

Even though these considerations may suggest that the influence of motivated reasoning is limited to explicit evaluations, it is important to note implicit evaluations may also be affected to the extent that people activate new information in the course of the validation process. For instance, if someone is motivated to hold a positive impression of a particular social group but experiences negative affective reactions toward the members of this group, the individual may engage in a directed memory search to retrieve positive information about the group, which should promote a positive evaluation for both explicit and implicit evaluations (see Blair, Ma, & Lenton, 2001; Gawronski & Bodenhausen, 2005; Peters & Gawronski, 2011a). If, however, the positive impression is maintained by merely negating the negative evaluation implied by the affective gut response (i.e., without retrieving supportive positive information), explicit and implicit evaluations should show a dissociation, such that explicit

evaluations reflect the desired positive evaluation, whereas implicit evaluations should reflect the original negative response.

3. Operating Principles and Operating Conditions

An important question is how the distinction between associative and propositional processes relates to the distinction between automatic and controlled processes. This relation has been a common source of confusion, such that associative processes have sometimes been equated with automatic processes while propositional processes have been equated with controlled processes (e.g., Mitchell, De Houwer, & Lovibond, 2009). In the APE model, we draw a sharp line between the two dichotomies. Whereas the associative–propositional distinction describes the operating principles of different processes, the automatic-controlled distinction refers to the conditions under which a given process is operating (Gawronski & Bodenhausen, 2009). Put differently, statements about the *operating principles* of a given process represent definitions of *what* a particular process is doing (e.g., activation, validation). In contrast, statements about *operating conditions* represent empirical claims about *when* that process is operating (e.g., Does the process operate when cognitive resources are limited? Does the process operate when there is an intention to alter or stop the process?). In the APE model, associative processes are conceptualized as the activation of associations on the basis of feature similarity and spatiotemporal contiguity; propositional processes are conceptualized as the validation of activated information on the basis of logical consistency. Importantly, there is no one-to-one mapping between operating principles and operating conditions, such that associative processes would operate automatically, whereas propositional processes operate in a controlled fashion (Gawronski & Bodenhausen, 2007a, 2009). Instead, both associative and propositional processes have automatic and controlled aspects. Moreover, propositional validation processes involve different subcomponents, some of which may operate automatically while others operate in a controlled fashion. In the following sections, we discuss the operating conditions of associative and propositional processes separately for each of the "four horsemen" of automaticity: awareness, intentionality, efficiency, and controllability (Bargh, 1994).[4]

[4] Another common dichotomy is the distinction between implicit and explicit processes. As outlined in Footnote 1, we use the terms *implicit* and *explicit* to describe the evaluative responses assessed by different kinds of measurement procedures. This terminology is based on the fact that the evaluative connotation of a particular response is implicit in responses assessed by indirect measurement procedures, but explicit in direct self-report measures. However, the implicit–explicit distinction has also been used interchangeably with the distinctions between conscious versus unconscious processes (Greenwald & Banaji, 1995) and automatic versus controlled processes (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009). These features are addressed in detail in the following discussion of the four characteristics of automaticity.

3.1. Awareness

The awareness criterion is commonly defined in terms of introspective access to mental processes or mental contents. Empirically, lack of introspective access can be established through participants' inability to verbally report a mental process or a mental content. For instance, in the literature on implicit evaluation, it is often assumed that indirect measurement procedures tap into "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (Greenwald & Banaji, 1995, p. 8). To be sure, indirect measurement procedures are usually performance-based and therefore do not require introspective access for the assessment of an evaluation. However, as we argued elsewhere (Gawronski & Bodenhausen, 2007b), this does not imply that the evaluative responses assessed by these measures are indeed unconscious in the sense that participants are generally unable to verbally report these evaluations. Any such claim represents an empirical assumption that has to be tested as such, and the currently available evidence clearly speaks against this assumption (for reviews, see Gawronski, Hofmann, & Wilbur, 2006; Gawronski et al., 2007).

In the APE model, we argue that people usually have *experiential access* to their affective gut reactions resulting from associative processes, and that they often rely upon these reactions in making propositional evaluative judgments. Still, people also sometimes reject their affective gut reactions as a basis for an evaluative judgment when these reactions are inconsistent with other momentarily considered information (Gawronski, Peters, Brochu, et al., 2008; Gawronski & Strack, 2004). However, such dissociations between affective gut reactions and evaluative judgments do not imply that the affective gut reactions are introspectively inaccessible. Consistent with these assumptions, several studies showed that correlations between implicit and explicit evaluations increase when participants are instructed to focus on their feelings for the attitude object (e.g., Banse, Seise, & Zerbes, 2001; Gawronski & LeBel, 2008; Grumm, Nestler, & von Collani, 2009; Ranganath et al., 2008; Smith & Nosek, in press). Such findings are difficult to reconcile with the assumption that the evaluations assessed by indirect measurement procedures are unconscious. Yet, it is line with the core assumptions of the APE model, according to which the correspondence between implicit and explicit evaluations should depend on the reliance on affective gut responses in the course of making an evaluative judgment.

Note, however, that even though people may be experientially aware of the affective gut reactions resulting from activated associations, they may sometimes be unaware of the associative processes that gave rise to these reactions (Nisbett & Wilson, 1977). For instance, people may show a positive or negative gut response toward an unfamiliar individual on the

basis of that person's similarity to a known individual; yet they may not be able to identify the similarity between the two individuals as the cause of their affective gut response (e.g., Lewicki, 1985; Verosky & Todorov, 2010). In terms of the APE model, such effects can be explained by the principle of feature similarity that characterizes the process of association activation. In the current example, the feature similarity between the two individuals may activate evaluative associations related to the known individual, thereby eliciting an affective gut response that is in line with the valence of these associations. Yet, the particular content of these associations may remain unconscious even when people are experientially aware of the affective gut reaction resulting from these associations (Bargh, Litt, Pratto, & Spielman, 1989; Greenwald, 1992). In other words, people may be experientially aware of their affective gut reactions to a person or object, but they may sometimes be unaware of the particular associations that are responsible for these reactions.

As for propositional processes, conscious awareness may not be required for the default process of affirming the validity of activated information, even though people may sometimes engage a conscious reassessment of the validity of that information. Similar considerations apply to the process of monitoring the consistency of momentarily activated information. In many cases, this monitoring process may operate outside of conscious awareness, even though people can certainly monitor their belief systems consciously to identify potential inconsistencies. However, inconsistency between activated information is assumed to raise conscious awareness, which in turn supports the effective resolution of inconsistency (Morsella, Zarolia, & Gazzaley, in press). In such cases, the necessary reassessment of the activated information involves conscious awareness of the involved processing steps, such as the negation (i.e., reversal of the truth value) of a particular proposition or the search for an additional proposition that resolves the inconsistency. Ignoring ambiguous cases of altered states of consciousness (e.g., hypnosis), the behavioral process of reporting an evaluative judgment generally occurs under conscious awareness.

3.2. Intentionality

Another common assumption is that the responses assessed by indirect measurement procedures are unintentional. In general terms, a process can be described as unintentional when it is instigated in the absence of a person's intention to start that process (Bargh, 1994). Applied to evaluative responses, it has been argued that the activation of evaluative associations—and thus the affective gut reactions resulting from these associations—occurs regardless of the intention to evaluate an object (e.g., Bargh, Chaiken, Raymond, & Hymes, 1996). This assumption is

consistent with evidence from social-cognitive neuroscience suggesting that certain kinds of evaluative responses are indeed independent of the intention to evaluate a given target object (e.g., Cunningham, Raye, & Johnson, 2004). Based on these findings, it appears that the activation of evaluative associations in memory can indeed occur unintentionally, thereby meeting the second criterion of automaticity.

Even though this conclusion seems straightforward and in line with a widespread assumption in the literature, it is important to note that evaluative associations can also be activated intentionally. For instance, an individual may intentionally search for particular information in memory, and the evaluative associations that are activated through this search may influence this person's affective responses to the relevant target object. In line with this assumption, several studies have shown that instructing participants to think about social group members with certain characteristics influences implicit evaluations of the relevant groups (e.g., Gawronski & Bodenhausen, 2005; see also Blair et al., 2001; Peters & Gawronski, 2011a). Thus, even though the activation of evaluative associations—and thus the affective gut reactions resulting from these associations—does not require intention, evaluative associations can also be activated intentionally.

As for propositional processes, we argue that intention is not required for the default process of affirming the validity of activated information, even though individuals may sometimes engage in an intentional reassessment of the validity of that information. Similar to our assumptions about awareness, we assume that such intentional reassessments will occur when the overall set of activated information is inconsistent. In such cases, people will intentionally restore consistency by either negating (i.e., reversing the truth value of) a particular proposition or searching for an additional proposition that resolves the inconsistency. In addition, we argue that the monitoring of (in)consistency by the anterior cingulate cortex occurs mostly unintentionally (Morsella et al., in press), even though people may sometimes engage in an intentional assessment of the consistency of their beliefs. The behavioral process of reporting an evaluative judgment is usually intentional, even though people may construe their behavior differently afterward (e.g., "I didn't mean to say that.").

3.3. Efficiency

In general terms, a process can be described as efficient if it operates despite conditions of restricted cognitive resources. Applied to evaluative responses, it is commonly assumed that evaluative associations are activated in response to a particular stimulus even when cognitive resources are scarce. Although there seem to be limits in the processing of evaluative information under conditions of restricted resources (e.g., Mitchell, Nakic, Pine, & Blair, 2007), the APE model agrees with the contention that associative processes

are highly efficient. However, this efficiency does not imply that evaluative associations cannot be activated in an effortful manner. Even though evaluative associations—and thus the affective gut reactions resulting from these associations—do not require cognitive effort to become activated, they can also be activated through the effortful retrieval of evaluative information from memory (e.g., Blair et al., 2001; Gawronski & Bodenhausen, 2005; Peters & Gawronski, 2011a).

As for propositional processes, we again argue that the default process of affirming the validity of activated information is highly efficient in the sense that it occurs even under conditions of limited cognitive resources. However, the situation is different for the monitoring and the resolution of inconsistency. In many situations, the monitoring of momentarily activated information may occur efficiently through the unconscious and unintentional operation of the conflict monitoring system in the anterior cingulate cortex (Botvinick, Cohen, & Carter, 2004). However, when people engage in a conscious and intentional assessment of particular pieces of information, limits in working memory capacity will constrain (a) how much information they can hold simultaneously in memory and (b) the complexity of syllogistic inferences they can perform to assess the consistency of this information. To the extent that the amount of relevant information is rather small and the complexity of the required inferences is low, the identification and resolution of inconsistency will require little amounts of cognitive resources (e.g., Richter, Schroeder, & Wöhrmann, 2009). If, however, the amount of relevant information is large or the required inferences are rather complex, the cognitive resources that are required for the identification and resolution of inconsistency will be more substantial (e.g., Martinie, Olive, & Milland, 2010; see also Wilson, Lindsey, & Schooler, 2000). In other words, we do not assume that propositional processes are cognitively effortful *per se*, as some researchers have argued (e.g., Strack & Deutsch, 2004). Rather, processes of propositional reasoning can be more or less effortful depending on the complexity of the inferences that are involved (Gawronski & Bodenhausen, 2006a, 2007a). The behavioral process of reporting the outcome of these validation processes usually affords little cognitive resources. What requires more cognitive resources is the mental process of reaching an evaluative conclusion, not the behavioral process of reporting that conclusion.

3.4. Controllability

Another important characteristic that has been used to describe implicit evaluations is the notion of controllability. In technical terms, a process can be described as uncontrollable if it operates despite a person's intention to terminate that process. Thus, whereas the intentionality criterion refers to the goal of starting a process, the controllability criterion refers to the goal of altering or stopping a process (Bargh, 1994). Applied to the current question,

one could argue that the activation of evaluative associations—and thus of the affective gut reactions resulting from these association—is uncontrollable if this activation process cannot be altered or terminated. We argue that the activation of evaluative associations is controllable to some extent. Yet, the overall success in controlling the activation of evaluative associations depends on the nature of the adopted control strategy. According to the APE model, the most critical factor in this regard is whether the adopted control strategy implies a negation of an already activated evaluation or an affirmation of a new evaluation. As outlined above, negating the validity of an affective gut reaction will reduce the influence of this reaction on evaluative judgments. However, it may not eliminate the affective gut reaction *per se*, as mere negations do not necessarily deactivate the associations that gave rise to this reaction. In contrast, affirming new evaluative information typically activates new associations in memory, which tend to influence the affective gut reactions resulting from activated associations in the intended direction. These assumptions are consistent with research on emotion regulation, showing that deliberate attempts to suppress affective reactions (negation) usually leave these reactions unaffected, whereas attempts to attribute a different meaning to the response-eliciting stimulus via reappraisal (affirmation) are capable of modifying affective reactions (e.g., Gross, 1998).

As for propositional processes, we argue that the default process of affirming the validity of activated information is generally controllable, as activated information can always be negated. However, the process of monitoring the (in)consistency of activated information is most likely uncontrollable, in that it cannot be altered or terminated. Of course, *responses* to inconsistency can technically be altered through the resolution of inconsistency (e.g., activation of the anterior cingulate cortex may be reduced once inconsistency is resolved). However, that does not imply that the *process* of monitoring inconsistency can be altered or stopped (Morsella et al., in press). Finally, the process of inconsistency resolution is generally controllable, given that people may change the preferred strategy to resolve inconsistency. For instance, instead of rejecting one of the involved propositions as false, a person may search for a new proposition that resolves the inconsistency (or the other way round). The behavioral process of reporting an evaluative judgment is generally controllable, as a person can always report a different judgment voluntarily.

4. Changes in Implicit and Explicit Evaluations

Even though the APE model has been applied to a wide range of questions within and beyond social psychology, its original purpose was to integrate the mixed evidence regarding changes in implicit and explicit

evaluations (Gawronski & Bodenhausen, 2006a).[5] Using a variety of different manipulations, some studies showed changes in explicit, but not implicit, evaluations (e.g., Gawronski & Strack, 2004; Gregg, Seibt, & Banaji, 2006), whereas others found changes in implicit, but not explicit, evaluations (e.g., Karpinski & Hilton, 2001; Olson & Fazio, 2006). Yet, other studies found corresponding changes in both implicit and explicit evaluations (e.g., Olson & Fazio, 2001; Richeson & Nussbaum, 2004). The APE model not only explains the processes underlying these divergent patterns, but it also includes specific predictions regarding when each of these patterns should occur.

According to the APE model, whether a given factor leads to changes in explicit but not implicit evaluations, implicit but not explicit evaluations, or corresponding changes in both explicit and implicit evaluations depends on (a) which of the two kinds of processes—associative or propositional—is *directly* influenced in the first place, and (b) whether changes in one process lead to *indirect* changes in the other process. On the basis of these two principles, one can construct four basic cases of how a given factor may influence implicit and explicit evaluations of a particular object. The first case involves a direct influence on the activation of associations in memory, with the evaluation implied by these associations being accepted by the propositional validation process (see Fig. 2.3A). Such cases should lead to corresponding changes in implicit and explicit evaluations, with changes in explicit evaluations being mediated by changes in implicit evaluations (described as Case 1 by Gawronski & Bodenhausen, 2006a). The second case involves a direct influence on the activation of associations in memory, with the evaluation implied by these associations being rejected by the propositional validation process (see Fig. 2.3B). According to the APE model, such cases should lead to changes in implicit, but not explicit, evaluations (described as Case 2 by Gawronski & Bodenhausen, 2006a). The third case involves a direct influence on the propositional validation process, with the activation of associations being unaffected by propositional reasoning (see Fig. 2.3C). According to the APE model, such cases should lead to changes in explicit, but not implicit, evaluations (described as Case 3 by Gawronski & Bodenhausen, 2006a). Finally, the fourth case involves a direct effect on the propositional validation process, with the activation associations being influenced in line with the outcome of propositional reasoning (see Fig. 2.3D). Such cases should lead to corresponding changes in implicit and explicit evaluations, with changes in implicit evaluations

[5] Note that we use the term *change* to refer to changes in behavioral responses (i.e., evaluation). In terms of traditional attitude theories, such behavioral changes may reflect either the formation of an attitude toward an unfamiliar object or the change of an existing attitude toward a familiar object. Even though the APE model acknowledges prior knowledge as an important moderator of associative and propositional processes (e.g., the likelihood of propositional inconsistency is reduced for novel attitude objects), it does not assume differences in the basic mechanisms per se.

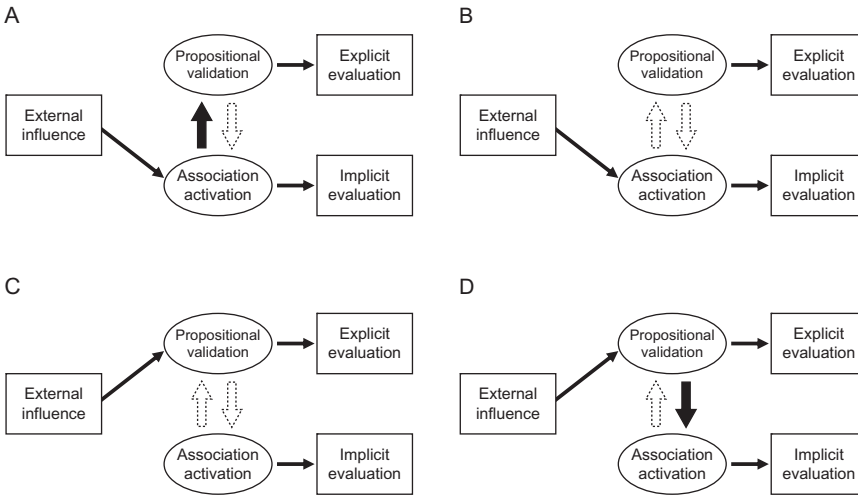

Figure 2.3 Potential direct and indirect influences of an external factor on associative and propositional processes underlying changes in implicit and explicit evaluations. Thin arrows depict direct effects of an external factor on either associative or propositional processes and influences of the two processes on implicit and explicit evaluations; fat arrows depict mutual influences between associative and propositional processes, with solid arrows depicting the presence of an effect and open arrows the absence of an effect. Adapted from Gawronski and Bodenhausen (2006a), reprinted with permission.

being mediated by changes in explicit evaluations (described as Case 4 by Gawronski & Bodenhausen, 2006a). In the following sections, we first discuss the nature of direct influences on associative and propositional processes. Expanding on this discussion, we then outline the conditions under which direct influences on one type of process lead to indirect effects on the other process.

4.1. Direct influences on associative processes

According to the APE model, the activation of associations depends on the relative fit between (a) the preexisting structure of associations in memory and (b) the overall set of input stimuli. Thus, the associations that are activated in response to a given stimulus may vary as a function of (a) changes in the underlying associative structure or (b) changes in the overall set of input stimuli. Limiting the focus to *direct* influences on the two components, the first case involves the formation of new associations on the basis of mere co-occurrences between stimuli, whereas the second case involves the context-dependent activation of existing associations.

4.1.1. Direct formation of new associations

Changes in the underlying associative structure usually occur through the formation of new associations in memory. In our original presentation of the APE model, we discussed evaluative conditioning (EC) effects as a prototypical example of direct influences on the underlying associative structure (for a review, see De Houwer, Thomas, & Baeyens, 2001). In a typical EC study, a conditioned stimulus (CS) is repeatedly paired with a positive or a negative unconditioned stimulus (US). The common finding is that subsequent evaluations of the CS reflect the valence of the US it has been paired with (for a meta-analysis, see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010). From the perspective of the APE model, being exposed to a CS that repeatedly appears in spatiotemporal proximity of a US creates a mental association between the CS and the US in memory. As a result, activating the CS in memory activates the representation of the US, which in turn produces an evaluative response to the CS that is in line with the valence of the US. Consistent with these assumptions, a large body of research has shown that repeated pairings of an attitude object with positive or negative stimuli influence implicit evaluations of the attitude object (e.g., Dijksterhuis, 2004; Gawronski & LeBel, 2008; Gibson, 2008; Grumm et al., 2009; Hermans, Baeyens, Lamote, Spruyt, & Eelen, 2005; Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Houben, Havermans, & Wiers, 2010; Karpinski & Hilton, 2001; Olson & Fazio, 2001, 2006; for a review, see Gawronski & Sritharan, 2010). In real-world settings, examples of such direct formation of associations include commercials in which a product is paired with a positive stimulus (e.g., an attractive person) even though the pairing does not bear any descriptive information about the product that would be deemed valid by the perceiver.

4.1.2. Context-dependent activation of existing associations

Mental representations of attitude objects can contain a variety of diverse associations, and these associations need not be evaluatively homogeneous (e.g., Quinn, Hugenberg, & Bodenhausen, 2004). Yet, encountering an attitude object is unlikely to result in the activation of every piece of information that is stored in memory. Instead, activation tends to be limited to particular subsets of associations that are cued by relevant contextual factors. Thus, if an attitude object is associated with both positive and negative information in memory, contextual cues may influence which of these associations become activated in response to the object, thereby producing different implicit evaluations of the object as a function of the context (e.g., Rydell & Gawronski, 2009). Consistent with these assumptions, several studies have shown that implicit evaluations of evaluatively ambiguous objects tend to vary as a function of the context in which these object are encountered (e.g., Barden et al., 2004; Dasgupta & Greenwald,

2001; Roefs,et al., 2006; Rudman & Lee, 2002; Wittenbrink et al., 2001; for a review, see Gawronski & Sritharan, 2010). An illustrative example in this regard is a study by Wittenbrink et al. (2001) who found that the same Black person elicited more favorable evaluations when that person was encountered in the context of a family barbeque than in the context of a graffiti wall.

4.2. Direct influences on propositional processes

In the APE model, we define propositional processes as the validation of momentarily activated information on the basis of consistency principles. As such, the outcome of propositional processes may vary as a function of (a) changes in the set of propositional information that is considered for an evaluative judgment and (b) changes in the strategy to achieve consistency within the set of considered information.

4.2.1. Changes in the set of considered propositions

In the original presentation of the APE model, we discussed two kinds of external factors that may directly influence the set of propositional information that is considered for an evaluative judgment. Whereas the first one involves direct exposure to new propositional information, the second one involves the retrieval of previously acquired information. First, an individual may be exposed to novel propositional information that is relevant for a particular judgment. A prototypical example is persuasive communication, in which recipients are exposed to persuasive arguments supporting either a positive or negative evaluation of an attitude object (Chaiken et al., 1989; Petty & Cacioppo, 1986). In many cases, being exposed to a persuasive message provides new information that may be inconsistent with an individual's current set of propositional beliefs. In such cases, consistency needs to be restored to avoid aversive feelings of cognitive dissonance (Festinger, 1957), for instance through the rejection of a previously endorsed evaluation. Second, individuals may sometimes be confronted with situations that lead them to retrieve previously stored information from memory that has not yet been activated. To the extent that this information is inconsistent with the set of momentarily considered propositions, the resulting inconsistency will have to be resolved (Festinger, 1957), which may involve the rejection of a previously endorsed evaluation. An illustrative example is research on hypocrisy (for a review, see Stone, in press), in which participants are initially induced to report a favorable evaluation of a particular behavior (e.g., a positive evaluation of physical exercise) and then reminded of instances when they failed to engage in that behavior (e.g., failures to engage in physical exercise). Although the most common outcome in this paradigm is a change in relevant behaviors (e.g., enhanced tendencies to

exercise), the resulting inconsistency could also be resolved by a rejection of the initial evaluation (e.g., discounting the benefits of physical exercise).

4.2.2. Changes in the strategy employed to achieve consistency

In some cases, people may consider the same set of inconsistent propositional information, but an external factor may lead them to adopt a different strategy to restore consistency. For instance, in the above example on racial prejudice, an individual may consider the three inconsistent propositions: (1) "I dislike Black people." (2) "Black people represent a disadvantaged minority group." (3) "Negative evaluations of disadvantaged minority groups are wrong." As we outlined above, consistency may be restored by rejecting any one of these propositions (see Fig. 2.1), and external factors may lead an individual to reject different propositions in different situations. Note, however, that such changes in the strategy to achieve consistency rarely occur without the consideration of additional information. In most cases, changes in the strategy to achieve consistency are driven by the consideration of additional propositions that either support or challenge one of the originally considered propositions (e.g., through exposure to information that racial minorities are still disadvantaged). Thus, even though changes in the strategy used to achieve consistency are theoretically possible without the consideration of additional information, most of these changes may be driven by the consideration of additional information, which makes them conceptually equivalent to the above case of changes in the considered set of propositions.

4.3. Mutual interactions and indirect influences

A central assumption of the APE model is that associative and propositional processes do not operate in isolation, but instead interact. This assumption implies that direct effects on one process may indirectly influence the respective other. However, such indirect effects are constrained by the operating principles of the two processes. In fact, our assumptions about the interplay between associative and propositional processes allows us to derive specific predictions about the conditions under which a given factor should lead to (a) changes in implicit but not explicit evaluations, (b) changes in explicit but not implicit evaluations, or (c) corresponding changes in both implicit and explicit evaluations.

4.3.1. Indirect influences on propositional processes

According to the APE model, the affective reactions resulting from activated associations are used as a basis for evaluative judgments unless the propositional evaluation implied by the affective response is inconsistent with other information that is considered to be relevant for that judgment. Thus, a constraining factor that moderates whether a direct effect on

associative processes will indirectly influence the outcome of propositional processes is whether the affective reaction resulting from the newly activated associations is consistent with other relevant beliefs. If the affective response is consistent with other relevant beliefs, it should be used as a basis for evaluative judgments, thereby leading to corresponding effects on implicit and explicit evaluations with the effect on explicit evaluations being mediated by the effect on implicit evaluations (see Fig. 2.3A). If, however, the affective response is rejected as a basis for an evaluative judgment because of its inconsistency with other relevant beliefs, changes should emerge only for implicit, but not explicit evaluations (see Fig. 2.3B).

Evidence for these assumptions comes from several studies that compared EC effects on implicit and explicit evaluations. As we outlined above, repeated pairings of a CS with a positive or negative US may create new associations in memory, which in turn may influence the affective response that is elicited by the CS. To the extent that the affective response is consistent with all other information that is considered to be relevant, it should be used as a basis for an evaluative judgment, implying corresponding EC effects on implicit and explicit evaluations with EC effects on explicit evaluations being mediated by EC effects on implicit evaluations. If, however, the affective response resulting from the newly created associations is inconsistent with other relevant information, it may be rejected as a basis for an evaluative judgment. In this case, EC effects should emerge only for implicit, but not explicit, evaluations.

In a preliminary test of these predictions, a reanalysis of data by Olson and Fazio (2001) revealed that EC effects on explicit evaluations were fully mediated by EC effects on implicit evaluations, such that EC effects on explicit evaluations decreased to nonsignificance after controlling for implicit evaluations (Gawronski & Bodenhausen, 2006a). The reverse mediation was not statistically significant. A more stringent test of our predictions was conducted by Gawronski and LeBel (2008) who experimentally manipulated the proposed indirect path from associative to propositional processes (see Spencer, Zanna, & Fong, 2005). In their study, participants were repeatedly presented with pairings of the words Europe and Asia, such that one of them was repeatedly paired with positive adjectives whereas the other one was repeatedly paired with negative adjectives. After the EC manipulation, half of the participants were asked to write down how they feel about Europe and Asia; the remaining half were asked to write down what they know about Europe and Asia. It was expected that a focus on feelings would lead participants to consider the newly created affective responses to Europe and Asia a valid basis for an evaluative judgment about the two continents. In contrast, retrieving knowledge from memory should lead participants to base their judgments on the retrieved information, which may or may not be consistent with their newly created affective responses. In line with these predictions,

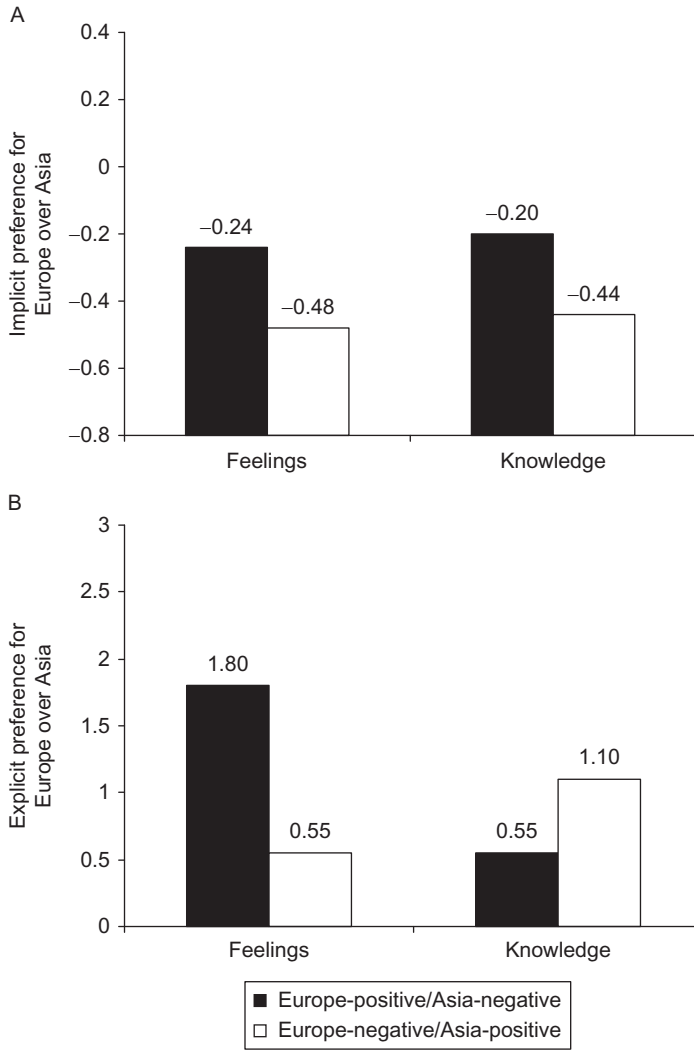

Figure 2.4 Implicit preferences (A) and explicit preferences (B) for Europe over Asia as a function of pairings with positive and negative stimuli (Europe-positive/Asia-negative vs. Europe-negative/Asia-positive) and introspection focus (feelings vs. knowledge). Adapted from Gawronski and LeBel (2008), reprinted with permission.

implicit evaluations showed significant EC effects regardless of whether participants focused on their feelings or on their knowledge (see Fig. 2.4A). In contrast, explicit evaluations showed a significant EC effect only when participants focused on their feelings but not when they focused on their knowledge (see Fig. 2.4B). Moreover, implicit and explicit

evaluations showed a significant positive correlation when participants focused on their feelings, but a tendency for a negative correlation when participants focused on their knowledge. This pattern was replicated in a study by Grumm et al. (2009) who tested the same predictions in the domain of self-evaluations.

Our assumptions about the reliance on the affective reactions resulting from newly created associations also explain an interesting asymmetry in the literature on EC effects. Several studies that compared EC effects on implicit and explicit evaluations used novel attitude objects as CS. These studies typically showed corresponding changes in both implicit and explicit evaluations (e.g., Hermans et al., 2002, 2005; Olson & Fazio, 2001). However, the majority of studies that used familiar attitude objects as CS found EC effects only on implicit, but not explicit, evaluations (e.g., Dijksterhuis, 2004; Gibson, 2008; Houben et al., 2010; Karpinski & Hilton, 2001; Olson & Fazio, 2006). To our knowledge, the only two exceptions are the reviewed studies by Gawronski and LeBel (2008) and Grumm et al. (2009) who found EC effects on both implicit and explicit evaluations when participants were instructed to focus on their feelings toward the familiar objects. From the perspective of the APE model, one could argue that prior knowledge about attitude objects may interfere with the emergence of EC effects on explicit evaluations to the extent that this knowledge is inconsistent with the affective response resulting from the newly created associations. For novel objects, however, there is no prior knowledge that could be inconsistent with the newly created affective response, thereby leading to EC effects on both implicit and explicit evaluations (see Footnote 5).

Similar considerations apply to context-dependent changes in the activation of associations, such that the affective reactions resulting from momentarily activated associations may or may not be used for an evaluative judgment depending on their consistency with other relevant beliefs. There is a large body of studies showing that the presence of context cues can modulate implicit evaluations of a given attitude object (for a review, see Gawronski & Sritharan, 2010). For instance, in the abovementioned study by Wittenbrink et al. (2001), implicit evaluations of Black people were more favorable when they were presented in the context of a family barbecue than when they were presented in front of a graffiti wall. Interestingly, although some studies showed corresponding effects on both implicit and explicit evaluations (e.g., Barden et al., 2004), others found context effects only on implicit, but not explicit, evaluations (e.g., Dasgupta & Greenwald, 2001). Again, these findings are consistent with our interpretation that contextual cues may influence the associations that are activated in response to a given object, and the affective reactions resulting from these associations may or may not be used as a basis for evaluative judgments depending on the consistency of the affective reaction with other momentarily considered information.

4.3.2. Indirect influences on associative processes

As for indirect effects on associative processes that are mediated by propositional processes, it is important to consider the two ways in which the activation of associations can be influenced in the first place (a) through the formation of new associations in memory and (b) through momentary changes in the activation of existing associations. Applied to the current question, the first case involves the acquisition of new propositional information that may be stored in associative memory, whereas the second case involves the momentary activation of existing associations through selective retrieval of information that is consistent with a particular proposition. In addition, we have argued that direct effects on propositional processes may indirectly influence the momentary activation of existing associations when these processes involve a recategorization of an attitude object in terms of an alternative category.

The first example of indirect effects on associative processes is the formation of new associations through the acquisition of new propositional information. Even though we discussed mere co-occurrences of an attitude object with positive or negative stimuli as the prototypical case of *direct* changes in the underlying associative structure, it is important to note that such changes can also occur *indirectly* through propositional processes. After all, encoding novel propositional information about an attitude object (e.g., persuasive arguments) is also capable creating new associations in memory that reflect the information entailed in these propositions. From the perspective of the APE model, the critical difference between the two cases is that mere co-occurrences of an attitude object with positive or negative stimuli creates new associations in a direct manner, whereas the acquisition of novel propositional information creates mental associations indirectly through processes of propositional reasoning. For the sake of conceptual clarity, we therefore distinguish between *associative learning* and *propositional learning* as two conceptually distinct learning processes. Both of these processes represent instances of learning, in that they involve the formation of new traces in memory. However, the two learning processes differ from each other in that associative learning involves the formation of new associations on the basis of mere co-occurrences between an attitude object and positive or negative stimuli regardless of whether these co-occurrences are considered to entail valid information about the valence of the attitude object. Propositional learning, in contrast, involves the formation of new associations on the basis of evaluative information that is considered valid. Thus, whereas associative learning is characterized by a direct effect on implicit evaluations that may indirectly influence explicit evaluations, propositional learning is characterized by a direct effect on explicit evaluations that may indirectly influence implicit evaluations.

A stringent test of these predictions was conducted by Whitfield and Jordan (2009), who combined an impression formation task with a standard

EC paradigm to study the distinct effects of associative and propositional learning on implicit and explicit evaluations. In one block of their study, participants were presented with positive or negative statements about an impression formation target; in another block, participants were repeatedly presented with pairings of the impression formation target and positive or negative stimuli. Results showed that the two manipulations influenced both implicit and explicit evaluations. However, whereas the EC manipulation produced a direct effect on implicit evaluations and an indirect effect on explicit evaluations that was mediated by implicit evaluations, the impression formation manipulation produced a direct effect on explicit evaluations and an indirect effect on implicit evaluations that was mediated by explicit evaluations.

An important question in this context concerns the conditions under which a direct effect on propositional processes does or does not lead to corresponding changes at the associative level (see Fig. 2.3C and D). As we outlined above, the critical factor that moderates such "top-down" influences is whether propositional reasoning processes involve an affirmation or negation of propositions. Whereas affirmation is capable of creating or activating new associations, negation typically leaves the associative components underlying the negated proposition unaffected (or in some cases even enhances rather than reduces the activation of these associations). To the extent that inconsistency represents the primary factor that leads to a rejection—or negation—of activated information, external factors that involve inconsistency-related rejections represent the prime candidate for cases in which a direct effect on propositional processes fails to produce corresponding effects at the associative level.

Evidence for these assumptions comes from a study by Gregg et al. (2006). Participants in their study were presented with positive or negative information about two hypothetical groups, and then asked to complete measures of explicit and implicit evaluations toward the two groups. After participants had completed the two measures, the experimenter explained that the particular pairing of positive and negative statements about the two groups was intended to be counterbalanced across participants, and that the participant was unfortunately run in the wrong condition. The experimenter then asked the participant to imagine a reversal of the positive and negative statements about the two groups, and to complete the two evaluation measures a second time. Whereas explicit evaluations showed the expected reversal between the two measurement occasions, implicit evaluations remained unaffected. In terms of the APE model, one could argue that participants rejected their initially created affective response toward the two groups in the course of making an evaluative judgment. However, merely rejecting this response did not erase or deactivate the associations that gave rise to the affective response, thereby leading to a dissociation between implicit and explicit evaluations.

More direct evidence for the proposed role of consistency processes comes from a study by Gawronski and Strack (2004) that was particularly designed to test the impact of inconsistency on explicit and implicit evaluations. Adopting Festinger and Carlsmith's (1959) induced compliance paradigm, participants were asked to endorse a counterattitudinal view under conditions of either high or low situational pressure, and then completed measures of implicit and explicit evaluations of the relevant subject. Replicating earlier findings on dissonance-related attitude change, explicit evaluations were more favorable toward the initially counterattitudinal subject when situational pressure was low than when it was high. Implicit evaluations, however, remained unaffected by the dissonance manipulation (for similar findings, see Wilson et al., 2000). Moreover, implicit and explicit evaluations showed a significant positive correlation when situational pressure was high, but a tendency for a negative correlation when situational pressure was low. According to Gawronski and Strack (2004), the inconsistency produced by counterattitudinal behavior under conditions of low situational pressure led participants to reject the evaluation implied by their preexisting affective response. Under conditions of high situational pressure, however, the resulting inconsistency could be resolved by means of a situational explanation for the counterattitudinal behavior. As a result, participants based their evaluative judgments on their preexisting affective response when situational pressure was high, but not when situational pressure was low.

The second example of indirect effects on associative processes involves momentary changes in the activation of existing associations. In the original presentation of the APE model, we discussed recategorization as the prototypical example of momentary changes in the activation of associations that are driven by propositional processes. The central idea underlying this assumption is that categorizing an attitude object in terms of a particular category will activate evaluative attributes that are associated with that category. Thus, to the extent that an external factor (e.g., differences in category salience, see Taylor, Fiske, Etcoff, & Ruderman, 1978) influences how a given attitude object is categorized, this factor may indirectly influence the activation of associations through the propositional categorization that object. For instance, categorizing Michael Jordan as an athlete may activate different evaluative associations compared with a categorization as African American. Consistent with this assumption, a study by Mitchell, Nosek, and Banaji (2003) found that participants showed more favorable implicit evaluations of Michael Jordan when they were forced to categorize him in terms of his occupation than when they were forced to categorize him in terms of his race (for related findings, see Fujita & Han, 2009; Hofmann, Deutsch, Lancaster, & Banaji, 2010; Wheeler & Fiske, 2005). Given that the application of a category label to an object represents a propositional statement that can be true or false (e.g., "Michael Jordan is

an athlete"), recategorization represents a propositional process that may indirectly influence associative processes. However, counter to previous studies showing that categorization in terms of a particular category inhibits the activation of alternative category-associations (Macrae, Bodenhausen, & Milne, 1995), recent evidence suggests that recategorization may not be as powerful as it is commonly assumed in preventing the activation of associations related to alternative categories (e.g., Gawronski, Cunningham, LeBel, & Deutsch, 2010). We will return to this issue in our discussion of challenges for the APE model.

A third example of indirect influences on associative processes that we did not explicitly discuss in the original presentation of the APE model is the selective retrieval of information that is consistent with a particular proposition. As previously noted, people may sometimes have a desire to believe in the validity of a particular proposition (Kunda, 1990). In such cases, people may engage in a directed memory search to retrieve information that confirms this proposition (see Trope & Liberman, 1996). From the perspective of the APE model, such processes of confirmatory hypothesis testing are characterized by a direct effect on propositional processes (i.e., a desired proposition is set as valid) that may indirectly influence associative processes (i.e., selective activation of information that supports the desired proposition).

Even though the three examples of indirect influences on associative processes (i.e., propositional learning, recategorization, selective retrieval in confirmatory hypothesis testing) may seem rather distinct, all of them involve factors that directly influence the consideration of particular propositions (i.e., acquisition of new propositional information, application of an alternative category label, affirming the validity of a desired proposition), which in turn affect associative processes in a proactive manner either through the formation of new associations or through the activation of preexisting associations (*affirmation*). According to the APE model, all of these cases should be characterized by corresponding effects on both explicit and implicit evaluations, with the effect on implicit evaluations being mediated by explicit evaluations (see Fig. 2.3D). If, however, an external factor leads to a reactive rejection of the propositional evaluation implied by an association-related affective reaction (*negation*), this factor should leave the activation of associations unaffected. In such cases, the relevant factor should produce changes in explicit, but not implicit, evaluations (see Fig. 2.3C).

4.4. Combined effects

In addition to the four basic cases, there may be circumstances under which evaluative responses are influenced by multiple distinct factors. Such cases may involve various combinations of the four patterns described in the last section. What these combined patterns have in common is that (a) one

factor directly influences associative processes and a second factor directly influences propositional processes and (b) the respective direct influences may or may not lead to a corresponding indirect influence on the respective other process.

A first example involves a direct influence on propositional processes that leaves the activation of associations unaffected and a direct influence on associative processes that does not pass the process of propositional validation (see Fig. 2.5A). This case can be described as a combination of Case 2 and Case 3—including their respective boundary conditions—leading to two independent main effects on implicit and explicit evaluations (described as Case 5 by Gawronski & Bodenhausen, 2006a). Empirical evidence for this pattern was provided by Rydell, McConnell, Mackie, and Strain (2006). In their study, participants were presented with a picture of a target person named Bob and an equal number of positive and negative statements about Bob. For each individual statement, participants were asked to guess whether that statement was an accurate or inaccurate description of Bob. Participants were provided feedback on their guesses, such that for half of the participants the positive statements were true and the negative statements were false; for the remaining half the negative statements were true

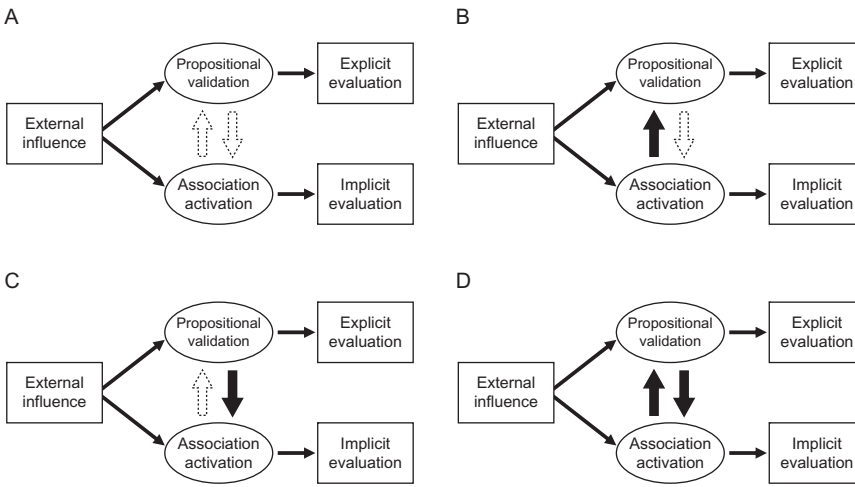

Figure 2.5 Combined effects of multiple direct and indirect influences of external factors on associative and propositional processes underlying changes in implicit and explicit evaluations. Thin arrows depict direct effects of an external factor on either associative or propositional processes and influences of the two processes on implicit and explicit evaluations; fat arrows depict mutual influences between associative and propositional processes, with solid arrows depicting the presence of an effect and open arrows the absence of an effect. Adapted from Gawronski and Bodenhausen (2006a), reprinted with permission.

and the positive statements were false. Immediately before the picture of Bob appeared on the screen, participants were subliminally presented with a prime word whose valence was opposite to the valence implied by the valid statements about Bob. Results showed that implicit evaluations were uniquely influenced by the valence of the subliminal words, whereas explicit evaluations were uniquely influenced the valence implied by the validity information about the evaluative statements.

Another example of multiple influences involves a direct influence on propositional processes that leaves the activation of associations unaffected and a direct influence on associative processes that passes the process of propositional validation (see Fig. 2.5B). This case represents a combination of Case 1 and Case 3, such that one factor will show a significant effect on explicit, but not implicit, evaluations, whereas the other factor will show corresponding effects on both explicit and implicit evaluations with the effect on explicit evaluations being mediated by implicit evaluations (described as Case 6 by Gawronski & Bodenhausen, 2006a). A potential example of this pattern is the spreading-of-alternatives effect, which describes the phenomenon that choosing between two equally attractive alternatives leads to more favorable evaluations of the chosen alternative and less favorable evaluations of the rejected alternative (Brehm, 1956). A common explanation of the spreading-of-alternatives effect is that people experience an aversive feeling of post-decisional dissonance when they recognize either (a) that the rejected alternative has positive features that the chosen alternative does not have, or (b) that the chosen alternative has negative features that are not present in the rejected alternative (Festinger, 1957). To reduce this aversive feeling, people are assumed to emphasize positive characteristics of the chosen alternative and negative characteristics of the rejected alternative, which leads to more favorable evaluations of the chosen alternative and to less favorable evaluations of the rejected alternative. To the extent that such processes of dissonance reduction involve a propositional negation of the relevant attributes (Gawronski et al., 2009), the spreading-of-alternatives effect may emerge only for explicit, but not implicit, evaluations.[6] However, Gawronski, Bodenhausen, and Becker (2007) identified a second, associative mechanism that may produce spreading-of-alternatives effects for implicit evaluations even in the absence of cognitive dissonance: associative self-anchoring. Specifically, Gawronski et al. argued that merely owning an object may be sufficient to create a mental association between the object and the self in memory, leading to an associative transfer of implicit self-evaluations to the owned object. To the

[6] Note that post-decisional dissonance may also be reduced by selective information search affirming a particular valence. In this case, the implied processes of confirmatory hypothesis testing may involve an indirect effect on implicit evaluations that is mediated by explicit evaluations (for a more detailed discussion, see Gawronski et al., 2009).

extent that most people's implicit self-evaluation is highly positive (Bosson, Swann, & Pennebaker, 2000; Greenwald & Farnham, 2000; Koole et al., 2001), this process may lead to postdecisional enhancement of implicit evaluations of owned objects without requiring the involvement of propositional processes (for related evidence, see Gawronski & LeBel, 2008; Prestwich, Perugini, Hurling, & Richetin, 2010; Zhang & Chan, 2009). Thus, choice decisions between two equally attractive alternatives may lead to spreading-of-alternatives effects via two independent mechanisms: (a) by a propositional process of postdecisional dissonance reduction that influences explicit, but not implicit, evaluations and (b) by a process of associative self-anchoring that influences both implicit and explicit evaluations, with the effect on explicit evaluations being mediated by implicit evaluations.[7]

A third potential combination involves a direct influence on propositional processes that indirectly influences the activation of associations and a direct influence on associative processes that does not pass the process of propositional validation (see Fig. 2.5C). This case can be described as a combination of Case 2 and Case 4, such that one factor will show a significant effect on implicit, but not explicit, evaluations, whereas the other factor will show corresponding effects on both explicit and implicit evaluations with the effect on implicit evaluations being mediated by explicit evaluations (described as Case 7 by Gawronski & Bodenhausen, 2006a). In the original presentation of the APE model, we considered this pattern as rather unlikely as the partial mediation pattern implied by this case would involve a rejection of association-related affective reactions even though processes of propositional reasoning influence affective reactions in exactly the same direction. Note, however, that these considerations about partial mediation apply only to cases in which a *single* factor simultaneously influences both associative and propositional processes in a direct fashion. Yet, there may be cases in which *two* factors operate independently in a given situation. For instance, if a factor of the kind discussed under Case 2 is combined with another factor of the kind discussed under Case 4, the two factors may indeed produce the pattern that we described as Case 7. However, such cases deviate from the notion of partial mediation emphasized in the original presentation of the APE model, which implies that the same factor may directly influence both associative and propositional processes. As we will discuss in the following section, there is preliminary evidence that such simultaneous direct influences may in fact be unlikely.

[7] Processes of associative self-anchoring may also account for the emergence of the spreading-of-alternatives effect in participants who have no explicit memory for their choice (e.g., Lieberman, Ochsner, Gilbert, & Schacter, 2001), which may seem difficult to reconcile with the notion of inconsistency resolution as a process that involves conscious awareness (for a more detailed discussion, see Gawronski et al., 2009).

A fourth potential combination involves a direct influence on propositional processes that indirectly influences the activation of associations and a direct influence on associative processes that passes the process of propositional validation (see Fig. 2.5D). This case can be described as a combination of Case 1 and Case 4, such that one factor will show corresponding effects on both explicit and implicit evaluations with the effect on implicit evaluations being mediated by explicit evaluations, whereas the other factor will show corresponding effects on both explicit and implicit evaluations with the effect on explicit evaluations being mediated by implicit evaluations (described as Case 8 by Gawronski & Bodenhausen, 2006a). In the original presentation of the APE model, we argued that such patterns may sometimes occur as a result of a single factor, such that this factor may directly influence associative processes and, at the same time, directly influence propositional processes, with the two kinds of direct effects leading to indirect effects on the respective other process. However, as we noted in the context of Case 7, such simultaneous direct influences may in fact be unlikely. Instead, parallel effects of the kind implied by Case 8 are more likely to be the result of two factors that operate independently, such as in the reviewed study by Whitfield and Jordan (2009) which combined a standard EC manipulation with an impression formation task involving verbal descriptions.

4.5. Multiple effects of a single factor or unique effects of multiple factors?

The four cases that we described as basic (see Fig. 2.3) all involve a single factor that is assumed to have a direct influence on either associative or propositional processes, and this direct effect may or may not produce an additional indirect effect on the respective other process. However, the question of how to describe the four combined patterns (see Fig. 2.5) is a little more ambiguous. On one hand, it is possible that a single factor directly influences both associative and propositional processes. On the other hand, the four combined patterns may also be interpreted as stemming from a joint operation of two independent factors. The emphasis on partial mediation in the original presentation of the APE model resonates with the first interpretation, implying that a single factor may simultaneously influence both associative and propositional processes. Counter to this interpretation, however, recent evidence suggests that such patterns of simultaneous direct influences may in fact be unlikely, and that the combined patterns depicted in Fig. 2.5 are more likely the result of two independent factors.

In a study that was particularly designed to test the simultaneous operation of two independent learning mechanisms on the basis of the same information, Peters and Gawronski (2011b) presented participants with pictures of four impression formation targets and written information

about these individuals. For two of the targets, 75% of the statements were positive and the remaining 25% were negative; for the other two targets, 75% of the statements were negative and the remaining 25% were positive. Participants' task was to guess the accuracy of the individual statements, and they were provided with feedback on their guesses, such that for half of the participants the majority information was true and the minority information was false; for the remaining half, the majority information was false and the minority information was true. It was expected that the object-valence contingencies implied by the differential proportions of positive and negative statements would influence implicit evaluations regardless of the validity of these statements. Explicit evaluations, in contrast, were expected to be influenced by the valence implied by the validity of the statements irrespective of the underlying object-valence contingencies. To the extent that these predictions could be confirmed, the results would provide evidence for the assumption that associative and propositional learning mechanisms can operate *simultaneously* on the basis of the *same* information. Counter to these predictions, however, validity information qualified both explicit and implicit evaluations when validity information was available during the encoding of the valence information. The expected dissociation occurred only when the presentation of validity information was delayed. In this case, validity information still influenced explicit evaluations, while its impact on implicit evaluations was significantly reduced (see also Gregg et al., 2006).

The reduced impact of validity information under delayed presentation conditions is consistent with the notion of *expression-related* dissociations, such that activated associations that have been formed at an earlier occasion may be rejected as invalid in the course of generating an evaluative judgment. However, the fact that both explicit and implicit evaluations were qualified by validity information when validity information was available during the encoding of evaluative information challenges the notion of *learning-related* dissociations that may stem from the simultaneous operation of two independent learning mechanisms on the basis of the same information. Note that this does not necessarily challenge the idea that dissociations between explicit and implicit evaluations may sometimes be due to learning-related processes. For instance, in the reviewed study by Rydell et al. (2006) implicit evaluations were uniquely influenced by the subliminal primes that preceded the presentation of the target individual Bob, whereas the valence implied by the valid behavioral descriptions uniquely influenced explicit evaluations. However, Rydell et al.'s (2006) paradigm differs from the one employed by Peters and Gawronski (2011b), in that it involved two independent sources of valence information (i.e., subliminal primes, behavioral descriptions) rather than a single source that was either validated or invalidated. Drawing on these considerations, it seems more likely that combined patterns of the kind depicted in Fig. 2.5 will be produced by the operation of two independent factors that directly influence either

associative or propositional processes (e.g., Rydell et al., 2006; Whitfield & Jordan, 2009) rather than the operation of a single factor that simultaneously influences both associative and propositional processes.

4.6. Robustness and stability

A common assumption in the attitude literature is that implicit evaluations are the product of highly robust representations that have their roots in long-term socialization experiences (e.g., Rudman, 2004; Wilson et al., 2000). This assumption has been challenged by research showing that implicit evaluations are sometimes easier to change than explicit evaluations (e.g., Gawronski & LeBel, 2008; Gibson, 2008; Grumm et al., 2009; Houben et al., 2010; Karpinski & Hilton, 2001; Olson & Fazio, 2006). In addition, implicit evaluations have been shown to be highly context-sensitive (e.g., Barden et al., 2004; Roefs et al., 2006; Rudman & Lee, 2002; Wittenbrink et al., 2001), and this sensitivity to contextual cues does not always generalize to explicit evaluations (e.g., Dasgupta & Greenwald, 2001).

From the perspective of the APE model, the question of relative robustness cannot be answered without addressing the question: "robust against what influence?" As we outlined in the preceding sections, either implicit or explicit evaluations can be comparatively more robust, depending on the nature of the influencing factor. For instance, factors that directly influence propositional processes may produce changes in explicit, but not implicit, evaluations (see Fig. 2.3B), suggesting that implicit evaluations are more robust than explicit evaluations (e.g., Gawronski & Strack, 2004; Gregg et al., 2006). In contrast, factors that directly influence associative processes may produce changes in implicit, but not explicit, evaluations (see Fig. 2.3C), suggesting that explicit evaluations are more robust than implicit evaluations (e.g., Gawronski & LeBel, 2008; Gibson, 2008; Grumm et al., 2009; Houben et al., 2010; Karpinski & Hilton, 2001; Olson & Fazio, 2006). In other words, the relative robustness of the two kinds of evaluations is not determined by particular features of the evaluations *per se*. Rather, their relative robustness is an interactive product of their underlying mental processes and the particular nature of the influencing factor.

A related question concerns the temporal and cross-situational stability of changes in implicit and explicit evaluations. Drawing on the abovementioned claim about the robustness of implicit evaluations, it is often assumed that implicit evaluations show higher levels of temporal and cross-situational stability, whereas explicit evaluations tend to fluctuate over time and as result of contextual factors that influence verbally reported evaluations (e.g., Wilson et al., 2000). The APE model turns this assumption on its head by arguing that the activation of associations in response to a given object depends on the overall set of input stimuli, including both the target object and momentarily present context cues. Yet, the resulting

fluctuations in affective gut responses may not be reflected in explicit evaluations if these responses are rejected as a valid basis for an evaluative judgment. If explicit evaluations of a given attitude object are consistently based on the same information, then changes in momentary affective reactions may not matter much for evaluative judgments formed in different contexts (Gawronski & Bodenhausen, 2007a). As such, implicit evaluations may often show higher rather than lower sensitivity to contextual influences compared with explicit evaluations (e.g., Dasgupta & Greenwald, 2001). However, this does not necessarily imply that implicit evaluations show lower levels of temporal stability than explicit evaluations. If the presence of contextual cues varies across measurement occasions, this may certainly be the case. However, to the extent that the presence of contextual cues is kept constant across measurement occasions, implicit evaluations may show relatively high levels of temporal stability, given that the same contexts should activate the same patterns of associations. Consistent with this assumption, Gschwendner, Hofmann, and Schmitt (2008) found that implicit evaluations showed higher levels of temporal stability when the measure included object-relevant contextual cues than when the attitude object was presented in isolation. The bottom line is that contextual cues modulate evaluative responses through the activation of associations in memory. As such, implicit evaluations may often show lower levels of temporal and cross-situational stability than explicit evaluations, at least when the fluctuating affective responses resulting from activated associations are rejected as a basis for evaluative judgments (Gawronski & Bodenhausen, 2007a).

4.7. Anything goes?

A potential concern about the APE model is that it can incorporate every possible outcome regarding changes in implicit and explicit evaluations (see Figs. 2.3 and 2.5), which seems to make it immune against disconfirming evidence. This concern would certainly be valid if our model did not specify the conditions under which each of the different cases should occur. However, as we outlined in the preceding sections, the APE model includes very specific assumptions about how associative and propositional processes are affected by external influences and how the two processes interact with each other. These assumptions imply precise predictions about the conditions under which one can expect changes in implicit, but not explicit, evaluations; changes in explicit, but not implicit, evaluations; or corresponding changes in both implicit and explicit evaluations. As for the third outcome, the APE model also includes specific assumptions about the conditions under which changes in implicit evaluations should be mediated by changes in explicit evaluations or the other way round. Of course, if any of these predictions is disconfirmed, one might be tempted to relate the unexpected outcome to its matching pattern in Figs. 2.3 and 2.5. However,

such *post hoc* integrations would require substantial revisions of the APE model's assumptions about the conditions under which the different patterns are supposed to occur. In fact, even though a significant body of research has supported novel predictions of the APE model in an *a priori* fashion (e.g., Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008; Gawronski & LeBel, 2008; Grumm et al., 2009; Whitfield & Jordan, 2009), the results of a small number studies are difficult to reconcile with the APE model without revising some of its original assumptions. One example is Peters and Gawronski's (2011b) research on the simultaneous operation of associative and propositional learning processes on the basis of the same information; another example is Gawronski, Cunningham, et al.'s (2010) research on the limited power of recategorization processes, which we will discuss in more detail in Section 6.3. Thus, even though the APE model may seem immune to disconfirming evidence because of the multitude of patterns that can occur according to the model, its assumptions about when each of these patterns should occur imply highly specific predictions, all of which have the potential for empirical disconfirmation.

5. Common Questions of Specific Issues

Over and above its hypotheses about changes in implicit and explicit evaluations, the APE model includes specific assumptions about various issues that play a central role in research and theorizing on evaluation. These issues include (a) the difference between propositional consistency and spreading activation, (b) the cross-cultural universality of consistency principles, and (c) the difference between processes, systems, and representations.

5.1. Propositional consistency versus spreading activation

According to the APE model, cognitive consistency is an inherently propositional phenomenon. This conceptualization is based on the fact that (in)consistency is not defined without reference to truth values and logical principles, which constitute the core concepts of propositional reasoning (Gawronski & Strack, 2004). The notion of syllogistic relations has already been inherent in Festinger's (1957) formal definition of inconsistency, according to which "*x* and *y* are dissonant if not-*x* follows from *y*" (p. 13). Moreover, the proposed role of truth values resonates with Festinger's (1957) argument that "a person does not hold an opinion unless he thinks it is correct" and "the same is true for beliefs, values, or attitudes" (p. 10).

Counter to this conceptualization, however, some researchers have argued that consistency principles tend to shape implicit, but not explicit, evaluations. For instance, drawing on their unified theory of attitudes,

stereotypes, self-esteem, and self-concept, Greenwald et al. (2002) showed that people's implicit evaluations of their ingroup (e.g., male-good), their implicit self-concept as a member of this group (e.g., me-male), and their implicit self-evaluations (e.g., me-good) are related in a manner such that individual differences in one type of association are predicted by the interaction of the other two. For instance, to the extent that participants strongly associated themselves with their gender and, at the same time, held positive associations with their gender, they also showed positive self-associations. Interestingly, such *balanced identities* consistently emerged at the level of implicit responses (assessed with IAT measures), but not for explicit responses (assessed with corresponding self-report measures). Drawing on these findings, Greenwald et al. (2002) concluded that consistency principles—most notably the principles of cognitive balance proposed by Heider (1958)—shape responses at the implicit level, and that the impact of cognitive consistency at the explicit level is often attenuated when people are unwilling or unable to accurately report their thoughts and feelings.

Even though we consider our disagreement with Greenwald et al.'s (2002) arguments largely terminological, we believe that it is conceptually more precise to describe balanced identity effects as the outcome of spreading activation processes in associative networks rather than as a reflection of consistency principles (Gawronski et al., 2009). Specifically, we argue that the activation of a particular concept in memory (e.g., self) may activate concepts that are associatively linked with this concept (e.g., ingroup), which in turn may activate evaluative associations related to that concept. As a result, the valence of one concept may transfer to other concepts that are associatively linked to that concept. Importantly, this transfer of evaluations may be driven by associative processes of spreading activation without requiring any involvement of propositional inferences regarding the validity of these associations. Thus, even though the three kinds of associations (e.g., self-ingroup, ingroup-evaluation, self-evaluation) may serve as the basis for three corresponding propositions (e.g., "I am male"; "male is good"; "I am good"), the associative process that gives raise to balanced identities can operate without the involvement of propositional inferences. Of course, balanced identities may also emerge at the explicit level when people base their judgments on the propositional implications of their activated associations. However, because the three propositions involved in balanced identity designs may reflect only a subset of all propositions that people consider for the corresponding judgments, the three components may sometimes seem imbalanced at the propositional level even though they are balanced at the associative level. Importantly, such "imbalanced" identities at the propositional level may not represent a genuine inconsistency if the full set of momentarily considered propositions is taken into account (Wellens & Thistlewaite, 1971; Wiest, 1965; for a review, see Insko, 1984). For these reasons, we argue that balanced identity effects may be better described as

the outcome of associative processes of spreading activation rather than as a reflection of consistency principles operating at the implicit level.

A useful example to illustrate the conceptual difference between spreading activation and cognitive balance is a study by Langer, Walther, Gawronski, and Blank (2009). In their study, participants were first presented with evaluative information about various individuals to create either positive or negative attitudes toward these individuals (e.g., participants learned that Peter is a likable person). Participants then learned that each of these "source" individuals liked or disliked another "target" individual (e.g., participants learned that Peter likes Mike). Afterward, participants' initial attitudes toward the sources were changed by means of evaluative information that was opposite to the initially induced attitude (e.g., participants learned that Peter is actually a dislikable person). In a control condition, participants were presented with neutral information about the sources to keep the initially induced source attitudes unqualified. The primary-dependent measure was participants' evaluation of the target (e.g., Mike). Results in the control condition showed that the initially formed source attitudes and the available information about source–target relationships led to evaluations of the target that were consistent with the notion of cognitive balance (Heider, 1958). That is, participants showed more favorable evaluations of targets that were liked by a positive source compared with targets that were liked by a negative source. Conversely, participants showed less favorable evaluations of targets that were disliked by a positive source compared with targets that were disliked by a negative source (see Fig. 2.6, left panel; see also Aronson & Cope, 1968; Gawronski, Walther, & Blank, 2005). The critical question in Langer et al.'s (2009) study was whether the crossover interaction obtained in the control condition would be reversed when participants' initial source attitudes were changed into the opposite direction, as predicted by the notion of cognitive balance. Interestingly, this was not the case. Instead, target evaluations directly matched participants' new attitudes toward the source that was paired with a given target, irrespective of the relation between the two (see Fig. 2.6, right panel). That is, targets that were paired with initially positive, now negative, sources were evaluated less favorably than targets that were paired with initially negative, now positive, sources. Importantly, this effect occurred regardless of whether the target was liked or disliked by the source. In other words, the new valence of the sources associatively spread to the targets that have previously been linked with these sources, and this associative transfer of valence undermined the emergence of cognitive balance.

Applied to present question, this finding is important, as it clarifies the conceptual difference between spreading activation and cognitive balance. Specifically, one could argue that learning about the relationship between two individuals creates a mental association between the two individuals regardless of whether the two individuals like or dislike each other

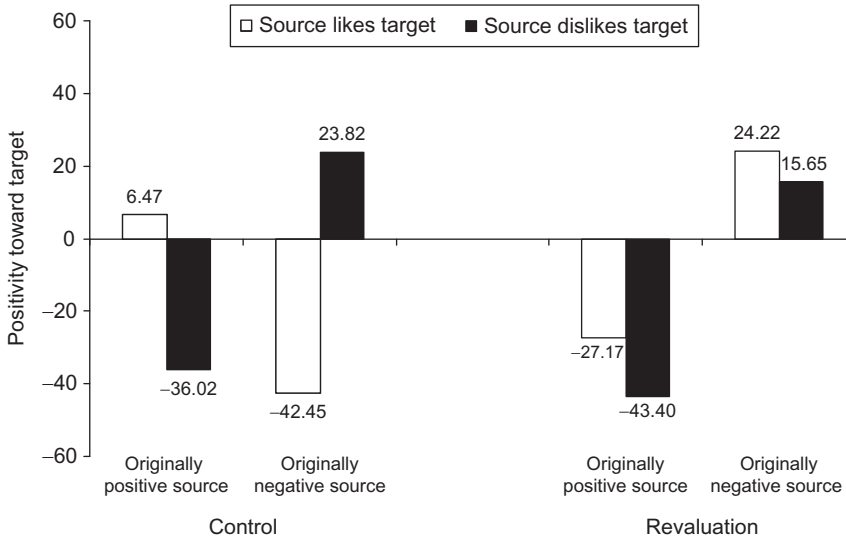

Figure 2.6 Target evaluations as a function of original source–valence (positive vs. negative), revaluation of source (control vs. revaluation), and source–target relation (source likes target vs. source dislikes target). Adapted from Langer et al. (2009), reprinted with permission.

(i.e., source–target association). To the extent that one of the two individuals is linked with positive or negative valence (i.e., source–valence association), the two associations may lead to an associative transfer of that valence to the other individual (i.e., target–valence association). Even though this transfer seems limited to conditions in which the relevant source–valence association is created after the acquisition of relationship information (see also Gawronski et al., 2005), these processes of spreading activation can lead to outcomes that contradict the basic notion of cognitive balance (e.g., a favorable evaluation of Mike, even though he is disliked by likable Peter). In other words, a person can become "guilty by mere association" with another individual even if the two people dislike each other. On the basis of these considerations, we regard it as important not to conflate associative processes of spreading activation with consistency-related processes, which we conceptualize as inherently propositional.

5.2. Culture and consistency

The central role of consistency principles in propositional reasoning has raised questions about potential limitations of the APE model in cross-cultural contexts. Even though Festinger (1957) argued that the need for cognitive consistency is as basic as hunger and thirst, the universality of

dissonance processes has been challenged by cross-cultural researchers who argued that the need for cognitive consistency may be limited to Western, individualist societies and less prevalent in Eastern, collectivist societies (e.g., Markus & Kitayama, 1991). In line with this contention, several studies have shown that effects of dissonance-related attitude change obtained in Western cultures do not necessarily generalize to Eastern cultures (e.g., Heine & Lehman, 1997; Hoshino-Browne et al., 2005; Kitayama, Snibbe, Markus, & Suzuki, 2004). On the basis of these findings, it is sometimes concluded that the need for cognitive consistency is a cultural phenomenon that is limited to Western societies, and that individuals from Eastern cultures do not necessarily have the same need to maintain cognitive consistency. If this conclusion is correct, it would certainly pose a challenge to the APE model, which attributes a fundamental role to cognitive consistency in the process of propositional validation.

In evaluating claims about cross-cultural differences in the need for consistency, it is useful to take a closer look at the different stages that are involved in dissonance-related processes, which can be described as (1) the identification of inconsistency, (2) the elicitation of dissonance experiences, (3) the resolution of inconsistency (see Fig. 2.7). Echoing Festinger's (1957) claim about the universality of consistency principles, we argue that this *sequence* of inconsistency processes is indeed universal, even though cultural worldviews may shape the *outcome* of inconsistency processes at each of the three stages (Gawronski, Peters, & Strack, 2008).

According to the APE model, both the identification and the resolution of inconsistency depend on the assignment of truth values to propositions, such that two propositions are inconsistent with each other if both are regarded as true and one follows from the opposite of the other (Gawronski & Strack, 2004). Thus, a relatively trivial source of cross-cultural differences resides in diverging opinions about the same state of affairs. For instance, the propositions "I love my wife" and "I am attracted to another woman" will be inconsistent if an individual endorses a culturally

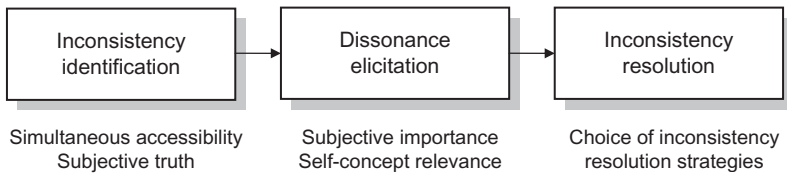

Figure 2.7 Three stages of inconsistency processes. Identification of inconsistency within one's system of beliefs is assumed to elicit aversive feelings of dissonance, which in turn motivate agents to resolve the inconsistency that has led to these feelings. Factors that influence the three processing steps are depicted below the respective boxes. Adapted from Gawronski, Peters, and Strack (2008), reprinted with permission.

transmitted norm of monogamous relationships, but it will be consistent if the individual endorses a culturally transmitted norm of polygamous relationships. Hence, what may appear as inconsistent from the perspective of one culture may be perfectly consistent from the perspective of another culture. In other words, endorsing the two propositions does not necessarily mean that the person lacks a need for consistency. Thus, in assessing the significance of cognitive consistency in different cultures, it is important to take culture-specific beliefs into account (Peng & Nisbett, 1999).

A less trivial source of cross-cultural differences has its roots in context-dependent assignments of truth values. A common finding in cross-cultural research is that individuals in Eastern cultures tend to put more emphasis on situational contexts than individuals in Western cultures (Choi, Nisbett, & Norenzayan, 1999; Markus & Kitayama, 1991; Spencer-Rodgers, Williams, & Peng, 2010). In East Asian cultures, whether or not a given statement about an object is regarded as true often depends on the particular context of that object. This interpretation stands in contrast to the predominantly decontextualized way of thinking in Western cultures. According to this view, the truth or falsity of a given statement about an object is determined by the intrinsic properties of that object, rather than by the particular context in which it is encountered. For instance, from a Western decontextualized point of view, endorsing a negative evaluation of abortion may be perceived as inconsistent with any kind of behavior favoring abortion. Yet, the same may not be true from an Eastern contextualized point of view, which may involve positive evaluations of abortions in some contexts and negative evaluations in others. From this perspective, decontextualized assignments of truth values have a higher likelihood to produce inconsistency than contextualized assignments of truth values. However, this does not imply that inconsistency would be irrelevant for propositional reasoning if the assignment of truth values is contextualized.

As for the elicitation of aversive feelings of dissonance, it is important to note that the magnitude of these feelings varies as a function of the subjective importance of the involved cognitions (Festinger, 1957). Needless to say, the relative importance of particular issues may vary substantially across cultures. In line with this contention, Hoshino-Browne et al. (2005) found that European Canadians tended to rationalize their choices more when these choices were made for themselves than when they were made for a friend. Conversely, Asian Canadians tended to rationalize their choices more when they were made for a friend than when they were made for themselves. These results can be explained with the higher importance of social relations compared to individual needs in East Asian cultures, which tends to be the opposite in North American cultures.

Finally, there may be cultural differences in the preferred strategy to resolve inconsistency. As outlined in the context of motivated reasoning, inconsistency can often be resolved in multiple ways, which leaves room for cultural differences in inconsistency resolution. For instance, several studies

on causal attribution have shown that Easterners have a stronger tendency to explain behavior in situational terms than Westerners (e.g., Miller, 1984; Morris & Peng, 1994). Thus, Easterners may explain their counterattitudinal behavior in induced compliance situations (see Festinger & Carlsmith, 1959) by means of situational demands regardless of whether situational pressure is high or low (e.g., "the experimenter asked me to do it"), whereas Westerners may explain their counterattitudinal behavior in situational terms only when situational pressure is high (e.g., "I got a lot of money for doing it"), but not when situational pressure is low.

Taken together, these considerations imply that the general sequence of consistency processes may indeed be universal (see Fig. 2.7). Still, cross-cultural differences may influence the respective outcomes of the three processing stages, thereby producing cross-cultural differences in the classic dissonance paradigms. Importantly, these content-related differences do not imply that consistency principles are irrelevant in the propositional reasoning processes of individuals from Eastern, collectivist cultures.

5.3. Processes, systems, and representations

Another common question about the APE model concerns its status as a dual-process, dual-system, or dual-representation theory. Whereas dual-process theories limit their assumptions to the distinction between two kinds of mental processes (see Gawronski & Creighton, in press), dual-system theories propose two distinct processing systems that operate on the basis of different mental or neurological structures (e.g., Lieberman, 2003; Strack & Deutsch, 2004). A particular subset of dual-system theories are dual-representation theories, which propose the storage of two distinct memory representations of the same object (e.g., Rydell & McConnell, 2006; Wilson et al., 2000). Even though the APE model has sometimes been interpreted as a dual-system or dual-representation theory, the emphasis in our model is on two distinct processes, rather than systems or representations. As we outlined in the context of mutual interactions between associative and propositional processes, the APE model does not assume a separate storage of propositions in memory. Instead, all information is stored in the form of associations, which may or may not pass a propositional assessment of validity. Thus, the APE model rejects the notion of two distinct memory representations, as it is implied by dual-representation theories (e.g., Rydell & McConnell, 2006; Wilson et al., 2000). In addition, the APE model remains agnostic about whether associative and propositional processes operate in one or two systems. Despite the recent trend from dual-process to dual-system theorizing (see Gawronski & Creighton, in press), we doubt whether claiming that the two processes operate in different systems provides any additional prediction over and above the ones that are already implied by the dual-process distinction. Moreover, even

though the notion of dual systems may provide a useful link to basic concepts in neuroscience, we remain skeptical as to whether the brain can be meaningfully divided into two systems, considering that the brain includes multiple specialized regions that mutually interact with other to produce a particular behavioral outcome.

6. Challenges and Open Questions

Even though the APE model has received strong empirical support in a substantial number of independent studies from different labs, there are a few conceptual and empirical issues that pose a challenge for the model. In the final sections of this chapter, we discuss these issues in more detail and outline various directions for future research.

6.1. Affect–cognition interface

A first issue concerns the interplay between affect and cognition in implicit evaluation. A central assumption of the APE model is that the overall valence of the associations that are activated in response to a given stimulus determines the evaluative quality of one's affective gut reaction to that stimulus. For instance, activation of the category *African American* may activate the stereotypical attribute *hostile*, which in turn leads to a negative affective response that is driven by the valence of the activated attribute. Considering that evaluative themes are known to be of fundamental importance in semantic concepts (Osgood, Suci, & Tannenbaum, 1957), it may seem uncontroversial to assert that such associations can shape immediate affective reactions. However, other researchers have argued that semantic and affective associations are stored independently in different areas of the brain (e.g., Amodio & Devine, 2006), which stands in contrast to our assumption that affective responses are driven by the valence of associated semantic information.

Even though there does not seem to be a straightforward answer to the thorny question of the affect–cognition interface (and some have despaired that such conclusions will always prove to be elusive; see Halgren, 1992), it is worth noting that each theoretical account has its own limitations. On one hand, models that propose independent representations of affective and semantic associations are unable to explain evidence for US-revaluation effects in EC, which refers to the phenomenon where subsequent changes in the valence of a US after pairing with a neutral CS lead to corresponding changes in the valence of the CS. US-revaluation effects indicate that EC effects are driven by a semantic association between the CS and the US in memory (*stimulus–stimulus learning*) instead of the CS becoming directly associated with a particular affective response (*stimulus–response learning*).

For instance, in a study by Walther, Gawronski, Blank, and Langer (2009) neutral faces (CS) were repeatedly paired with either positive or negative faces (US). After the presentation of CS–US pairings, the initial valence of the US faces was reversed by means of an impression formation task, such that initially positive faces were presented with negative information and initially negative faces were presented with positive information. In a control condition, the US faces were presented with neutral information. Results in the control condition showed that the CS faces elicited evaluative responses in line with the initial valence of the US faces, such that CS faces that were paired with positive US faces were evaluated positively, whereas CS faces that were paired with negative US faces were evaluated negatively. Importantly, CS faces in the revaluation condition reflected the newly acquired, rather than the initial, valence of the US faces. That is, CS faces that were paired with initially positive, now negative, US faces were evaluated negatively, whereas CS faces that were paired with initially negative, now positive, US faces were evaluated positively (see also Langer et al., 2009). These results are difficult to explain by accounts that assume independent representations of semantic and affective associations (e.g., Amodio & Devine, 2006). However, they are consistent with our assumption that affective responses are driven by the overall valence of associated concepts that are activated in response to a given object.

That being said, our account has difficulties explaining the origin of affective responses as long as it does not go beyond semantic associations between concepts. For instance, in the above example of an association between the concepts *African American* and *hostile*, one would have to explain the origin of the negative affective response to the concept *hostile*. Stating that the affective response comes from a semantic association to another concept is insufficient, as it would imply an infinite chain of links between semantic concepts (Fazio, 2007). To avoid this problem one would have to propose a unique type of affective association that is qualitatively distinct from semantic associations. Such an extended account, however, would have to explain how semantic and affective representations interact with each other, which comes back to the notoriously difficult question of the affect–cognition interface.

A potential answer might be implied by recent research on embodied cognition (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). According to this approach, affectively significant stimuli can evoke relevant sensorimotor representations in modality-specific areas of the association cortex. The activation of these representations involves, essentially, some degree of reenactment or simulation of the physical pleasure, arousal, or other bodily experiences associated with the stimulus. For example, seeing a bowl of mangos may activate gustatory representations related to sensory experiences of their flavor and sweetness. Importantly, however, these concrete, sensory representations are assumed to be

embedded in a hierarchical representational structure that becomes increasingly more abstract (for a discussion, see Winkielman, Niedenthal, & Oberman, 2009). Representations at the abstract level of this hierarchy may reflect relatively "disembodied" or "cold" associations (i.e., semantic knowledge that a particular object is good or bad), whereas the representations at the more concrete level of the hierarchy may reflect relatively "embodied" or "hot" associations (i.e., bodily sensations and arousal levels that are associated with particular concepts). For instance, in the above example, the concept *African American* may be associated with the negative semantic concept *hostile*. To the extent that the activation of such evaluative concepts does not involve any embodied simulations (e.g., increase in arousal), these associations may be regarded as "cold" evaluative knowledge. If, however, the activation of evaluative concepts involves more concretely embodied responses, they may be experienced as "hot" affective responses. Even though such a conceptualization requires a number of revisions in the theoretical core of the APE model, future research investigating the unique roles of valence and arousal associations may be helpful to provide a deeper understanding of the mechanisms underlying implicit evaluations.

6.2. Evaluative conditioning

Another question that has been the subject of ongoing debates concerns the mechanisms underlying EC effects. In the original presentation of the APE model, we argued that EC represents the prototypical case of direct effects on associative processes, such that repeated pairings of a CS and a US create a mental association between the CS and the US in memory without involvement of higher-order propositional processes. Thus, CS–US pairings should have a direct effect on implicit evaluations, which may or may not produce a corresponding indirect effect on explicit evaluations. This prediction has received considerable empirical support, the strongest evidence coming from studies that manipulated the proposed indirect effect on explicit evaluations (e.g., Gawronski & LeBel, 2008; Grumm et al., 2009). However, as outlined by De Houwer (2007), equating a particular effect with a specific mechanism is problematic for conceptual reasons. Specifically, De Houwer argued that the term *EC* has been used for at least three different things: (a) the procedure of pairing a CS with a positive or negative US; (b) the observable effect of that procedure, namely a change in the valence of the CS; and (c) the psychological mechanism that underlies the observable effect. To avoid conceptual confusion, De Houwer suggested that EC should be defined as an effect, namely "a change in the valence of a stimulus that results from pairing the stimulus with another stimulus" (p. 230). Defined in this manner, it would be ill-founded to conceptually *equate* associative learning with EC. Rather, associative learning should be interpreted as a psychological mechanism that *explains* EC effects. Importantly, the

conceptual distinction between EC as an effect and associative learning as a psychological mechanism implies that EC effects could also be driven by other processes that involve propositional reasoning. Thus, whether EC effects are driven by associative or propositional processes is not a matter of definition, but an empirical question that has to be investigated as such.

Indeed, recent reviews of the EC literature suggest that there is no single mechanism that fully accounts for the available evidence (De Houwer, 2007; Hofmann, De Houwer, et al., 2010; Jones, Olson, & Fazio, 2010). Instead, EC effects seem to be driven by associative learning under some conditions, but by propositional learning under other conditions. The challenging task for EC researchers is to identify the particular conditions under which EC effects are driven by one or the other process. In this regard, the procedural details of the CS–US pairings seem important. For instance, whereas some researchers make every effort to eliminate participants' awareness of CS–US contingencies (e.g., Jones, Fazio, & Olson, 2009; Knight, Nguyen, & Bandettini, 2003; Olson & Fazio, 2001; Walther, 2002; Walther & Nagengast, 2006), other researchers use rather blatant pairings without any attempt to conceal CS–US contingencies (e.g., Corneille, Yzerbyt, Pleyers, & Mussweiler, 2009; Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Stahl & Unkelbach, 2009; Stahl, Unkelbach, & Corneille, 2009). Interestingly, whereas the former studies suggest that contingency awareness is not required for EC effects to occur (e.g., Jones et al., 2009; Knight et al., 2003; Olson & Fazio, 2001; Walther, 2002; Walther & Nagengast, 2006), awareness of CS–US contingencies has been shown to be a necessary precondition for EC effects in the latter studies (e.g., Pleyers et al., 2007; Stahl & Unkelbach, 2009; Stahl et al., 2009). Thus, instead of treating EC effects as a homogeneous phenomenon that is always produced by the same mechanism, it is important to consider the procedural details of the employed paradigms when studying the boundary conditions of EC effects (Jones et al., 2010). Such investigations may provide deeper insights into the conditions under which EC effects are driven by either associative or propositional processes.[8]

6.3. Limited effects of recategorization

A central assumption of the APE model is that categorizing an attitude object in terms of a particular category will activate evaluative attributes that are associated with that category. Thus, recategorizing the object in terms of another applicable category may activate different associations, thereby

[8] Note that these considerations also apply to the above distinction between stimulus–stimulus learning and stimulus–response learning. After all, stimulus–stimulus learning seems more likely when a CS is repeatedly paired with the same US, but it seems less likely when a CS is paired with multiple different US of the same valence (e.g., Sweldens, Van Osselaer, & Janiszewski, 2010).

producing a propositionally induced change in implicit evaluations. Even though this assumption is shared by various other models (e.g., Fazio, 2007) and even though it received considerable empirical support in a number of independent studies (e.g., Fujita & Han, 2009; Hofmann, Deutsch, et al., 2010; Mitchell et al., 2003; Olson & Fazio, 2003; Wheeler & Fiske, 2005), recent evidence suggests that categorization processes may not be as powerful as it is commonly assumed in preventing the activation of associations related to alternative categories. Instead, it seems that category-related features of an object are capable of influencing implicit evaluations even when the object is categorized in terms of an alternative category.

Evidence for the power of unattended category cues comes from a study by Gawronski, Cunningham, et al. (2010). In their study, participants completed one of two conceptually equivalent measures of implicit evaluation: Fazio et al.'s (1995) evaluative priming task or Payne et al.'s (2005) affect misattribution procedure.[9] The prime stimuli in the two tasks included faces of Black and White men of either young or old age. Adopting a manipulation by Olson and Fazio (2003), half of the participants were asked to keep a mental tally of how many Black and White faces were presented throughout the task; the remaining half were asked to keep a mental tally of how many young and old faces were presented throughout the task. Consistent with the assumption that categorization determines which associations are activated in response to a given stimulus, Fazio et al.'s (1995) evaluative priming task showed reliable priming effects only for the category dimension to which participants paid attention. That is, participants showed reliable priming effects of the race dimension when they paid attention to race, but not when they paid attention to age. Conversely, participants showed reliable priming effects of the age dimension when they paid attention to age, but not when they paid attention to race. These effects were reflected in (a) reduced effect sizes, (b) reduced internal consistencies, and (c) reduced correlations to corresponding self-reports when attention was directed toward the respective alternative dimension (see Table 2.1). Interestingly, the pattern of results was quite different for Payne et al.'s (2005) affect misattribution procedure, which showed reliable priming effects of both category dimensions regardless of whether participants paid attention to race or age. Using the same criteria

[9] In Fazio et al. (1995) evaluative priming task, participants are primed with an attitude object, which is followed by a positive or negative target word. Participants' task is to evaluate the target words as quickly as possible. The typical finding is that participants show faster responses to positive target words when they were primed with a positive stimulus than when they were primed with a negative stimulus, and faster responses to negative target words when they were primed with a negative stimulus than when they were primed with a positive stimulus. In Payne et al.'s (2005) affect misattribution participants are primed with an attitude object, which is followed by a neutral Chinese ideograph. Participants' task is to indicate whether they find the Chinese ideograph visually pleasant or visually unpleasant. The typical finding is that the Chinese ideographs are evaluated more favorably when participants were primed with a positive stimulus than when they were primed with a negative stimulus.

Table 2.1 Effect sizes (Cohen's d), split-half correlations, and correlations to explicit preference scores of priming scores reflecting implicit preferences for Whites over Blacks (implicit racism) and for young over old (implicit ageism) in Fazio et al.'s (1995) evaluative priming task and Payne et al.'s (2005) affect misattribution procedure as a function of attention to race versus age of face primes

		Evaluative priming task		Affect misattribution procedure	
		Implicit racism	Implicit ageism	Implicit racism	Implicit ageism
Effect size	Attention to race	0.28	−0.02	0.14	0.24
	Attention to age	0.04	0.24	0.27	0.16
Split-half correlation	Attention to race	0.25	−0.29	0.58	0.82
	Attention to age	−0.13	0.27	0.57	0.79
Correlation to explicit measure	Attention to race	0.26	−0.10	0.30	0.31
	Attention to age	−0.05	0.39	0.29	0.29

Adapted from Gawronski, Cunningham, et al. (2010) reprinted with permission.
Note: Effect sizes with negative signs indicate priming effects in the opposite direction of the respective preference scores.

that were applied to assess the reliability of priming effects in Fazio et al.'s (1995) task, the obtained pattern of results included (a) equally strong effect sizes, (b) equally high internal consistencies, and (c) equally high correlations to corresponding self-report measures independent of attention instructions (see Table 2.1). These results challenge the generality of categorization effects on implicit evaluations, suggesting that the impact of unattended category cues depends on conditions inherent in specific tasks. Thus, under some conditions, category cues may activate category-related associations even if these cues are not used to categorize the relevant stimulus (see also Ito & Urland, 2003). At this point, however, it is not clear which particular characteristic of the affect misattribution procedure allowed unattended category cues to influence implicit evaluations. Future research that systematically manipulates procedural characteristics of the task may be helpful to gain a better understanding of the conditions under which categorization is capable of preventing the activation of associations related to alternative categories.

6.4. Measurement issues

The results by Gawronski, Cunningham, et al. (2010) indicate that different indirect measures of the same construct may not necessarily show identical effects of the same experimental manipulation. In fact, there is evidence that

the same experimental manipulation can even lead to *opposite* effects on different measures of the same construct (e.g., Deutsch & Gawronski, 2009; Gawronski & Bodenhausen, 2005). These results suggest that the mechanisms underlying different tasks can shape participants' responses in a nontrivial manner (Gawronski, Deutsch, LeBel, & Peters, 2008). Thus, to avoid misinterpretations of experimentally induced changes in measurement scores it seems important to replicate these changes with different measures that are based on different underlying mechanisms (e.g., Peters & Gawronski, 2011b; Prestwich et al., 2010; Rydell & Gawronski, 2009).

Another important issue in this context concerns the lack of process-purity of indirect measures. As we outlined in the original presentation of the APE model, performance on indirect measures involves a combination of multiple processes that tend to be confounded in standard data analytical procedures. Applications of mathematical modeling techniques (e.g., Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Klauer, Voss, Schmitz & Teige-Mocigemba, 2007; Payne & Bishara, 2009; Stahl & Degner, 2007) provide a significant advance in this regard, as these procedures can disentangle the contributions of multiple qualitatively distinct processes to participants' performance on indirect measurement procedures (for a review, see Sherman, Klauer, & Allen, 2010). These techniques are superior to standard data analytical procedures in multiple respects, one of the most important issues being that an experimentally produced effect on the obtained measurement scores may not necessarily reflect a change in the underlying evaluative response. Instead, changes in measurement scores may sometimes be due to effects on other processes, such as the controlled inhibition of automatically activated associations (Sherman et al., 2008). Thus, researchers should be careful in drawing potentially premature conclusions about implicit evaluations from experimentally produced differences in measurement scores (for a more detailed discussion, see Gawronski & Sritharan, 2010).

6.5. Problems with testing mediation patterns

One of the most significant aspects in testing the predictions of the APE model is the different patterns of direct and indirect effects on associative and propositional processes (see Fig. 2.3). In the original presentation of the APE model, we used the notion of direct and indirect effects to derive specific predictions about different mediation patterns, such that an external factor may have a direct effect on either implicit or explicit evaluations, which may or may not lead to a corresponding effect on the respective other evaluation. In social psychology, the most common way of testing such mediation patterns is Baron and Kenny's (1986) regression-based approach, which we also advocated for testing predictions of the APE model.

However, there are a few issues that make Baron and Kenny's (1986) method suboptimal for this particular purpose.

One issue is that implicit and explicit evaluations are the *outcomes* of associative and propositional processes; they are not psychological processes *per se*.[10] However, our assumptions about mutual interactions refer to outcomes only for "bottom-up" but not for "top-down" influences. Specifically, we assume that the evaluation implied by activated associations serves as input for processes of propositional reasoning, such that this evaluation may be accepted or rejected as a basis for an evaluative judgment on the basis of its consistency with other relevant beliefs. For such cases, it seems well justified to use Baron and Kenny's (1986) method to test the proposed "bottom-up" mediation patterns (see Fig. 2.3A and B), as implicit evaluations reflect the input for propositional validation processes. However, the situation is different for "top-down" mediation patterns (see Fig. 2.3C and D), as implicit evaluations are not necessarily influenced by the outcome of the propositional validation process. Even though this may be the case in some situations (e.g., Peters & Gawronski, 2011b), implicit evaluations are primarily influenced by the associations that become activated in the course of propositional reasoning. Thus, instead of using explicit evaluations *per se* as the mediator in Baron and Kenny's (1986) regression approach, it would seem more appropriate to use participants' responses on open-ended thought listings, which may capture the mental contents that are activated during processes of propositional reasoning. However, such a strategy also seems suboptimal, as it requires different types of measures for testing "bottom-up" and "top-down" mediation.

Another problem is that the two basic patterns involving indirect effects (see Fig. 2.3A and D) generally imply corresponding effects on both implicit and explicit evaluations. Thus, a stringent test of the proposed mediation patterns with Baron and Kenny's (1986) method always requires a test of the respective other pattern. For instance, when predicting a significant indirect effect on explicit evaluations that is mediated by implicit evaluations, the opposite mediation pattern should ideally be nonsignificant (and *vice versa*). In many cases, however, the opposite mediation is at least close to significance or simply somewhat weaker compared with the predicted mediation (e.g., Gawronski & Walther, 2008; Whitfield & Jordan, 2009), which is due to the correlational nature of the relation between the mediator and the distal outcome in Baron and Kenny's (1986) approach. Thus, the particular direction of the obtained mediation pattern often comes down to the question which of the two measures shows a stronger effect of the experimental manipulation, which can be strongly influenced by the internal consistency of the employed measure. Given that many (though not all)

[10] Note that this concern applies to any psychological measure, including all measures that are claimed to provide a proxy of mental processes or mental representations (De Houwer, Gawronski, & Barnes-Holmes, 2010).

indirect measures suffer from relatively low internal consistencies (Gawronski, Deutsch, et al., in press), testing the mediation patterns predicted by the APE model can be a delicate endeavor with Baron and Kenny's (1986) method.[11]

For these reasons, we consider experimental approaches to be a superior means to test our predictions about direct and indirect effects compared with Baron and Kenny's (1986) regression-based approach (see Green, Ha, & Bullock, 2010; Jacoby & Sassenberg, 2011; Spencer et al., 2005). The APE model includes specific assumptions about mutual influences between associative and propositional processes, and these assumptions can be used to derive specific predictions about the factors that should moderate experimentally induced changes on the distal, but not the proximal, outcome. For instance, in support of our assumptions about associative learning processes underlying EC effects, we found that EC effects on explicit evaluations emerged only when participants focused on their feelings toward the CS, but not when they focused on their knowledge about the CS. Implicit evaluations, in contrast, showed reliable EC effects regardless of participants' introspection focus (e.g., Gawronski & LeBel, 2008; Grumm et al., 2009). Similar considerations can be applied to indirect effects on associative processes that are mediated by propositional processes. Thus, our assumptions about mediation patterns can often be tested by means of the moderating factors that should reduce experimentally induced changes on the distal outcome, with changes in the proximal outcome being unaffected.

6.6. Context-dependent activation of associations

A final issue concerns the effects of contextual cues on the activation of associations. A central assumption of the APE model is that the same object may activate different associations as a function of momentarily available contextual cues. To the extent that these associations differ in terms of their valence, the same object may elicit different evaluative responses depending on the context in which it is encountered. However, our explanation of context effects may be regarded as circular as long as it does not specify the conditions under which different contexts activate either the same or different associations in response to the same stimulus. Without a specification of these conditions, contextually induced differences in implicit

[11] One example is Hofmann, De Houwer, et al.'s (2010) meta-analysis of EC effects, which revealed larger effect sizes for explicit evaluations than for implicit evaluations. This difference led the authors to conclude that EC effects are driven by propositional processes rather than by associative processes. However, Hofmann et al.'s meta-analysis did not control for the reliability of the employed measures, making conclusions about the operation of associative versus propositional processes on the basis of effect sizes premature. Even though we appreciate the general value of meta-analyses, we believe that experimental manipulations of the proposed link between proximal and distal outcomes (see Spencer et al., 2005) represent a stronger test than meta-analytic differences in effect sizes without controlling for differences in reliability (e.g., Gawronski & LeBel, 2008; Grumm et al., 2009).

evaluations would be explained by the activation of different associations, but the only evidence for differences in activated association are the differences in implicit evaluation that need to be explained in the first place.[12]

To fill this theoretical gap, Gawronski, Rydell, Vervliet, and De Houwer (2010) proposed a representational account of generalization versus contextualization effects in evaluative learning that specifies the conditions under which implicit evaluations reflect (a) initially acquired information about an object, (b) subsequently acquired, counterattitudinal information about the object, or (c) a mixture of both. Their account assumes that the encoding of evaluative information about an object produces a mental association that links the object to that information. When individuals are later exposed to information that is evaluatively incongruent with the initially acquired information, the resulting expectancy violation is assumed to trigger a search for contextual factors that may explain the observed discrepancy, thereby drawing attention to momentarily available contextual cues (Roese & Sherman, 2007). As a result, these cues are integrated into a contextualized representation of the object which includes the newly acquired, counterattitudinal information and the particular context in which it was acquired. Thus, the mental representation of the object acquires a "dual" nature involving (a) a *context-free representation* that includes the object and the initially acquired evaluative information, and (b) a *contextualized representation* that includes the object, the subsequently acquired, counterattitudinal information, and the context in which this information was acquired. Hence, encountering the object in the initial learning context should activate the context-free representation, thereby producing implicit evaluations that reflect the valence of the initially acquired information. However, encountering the object in the second learning context should activate the contextualized representation, thereby producing implicit evaluations that reflect the valence of the subsequently acquired counterattitudinal information. Finally, encountering the object in a novel context should activate the initially formed, context-free representation, thereby producing implicit evaluations that reflect the valence of the initially acquired information.

Evidence for these assumptions can be found in a study by Rydell and Gawronski (2009). In their research, participants were first presented with either positive or negative statements about a target person against a meaningless, colored background (e.g., a yellow screen). In a second learning block, participants were presented with information 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, implicit evaluations of the target

[12] Note that the criticism of circularity also applies to other accounts of context effects on implicit evaluations, such as online constructions of implicit evaluations (e.g., Schwarz, 2007) or changes in the object of evaluation rather than the evaluation of the object (e.g., Fazio, 2007).

person were assessed with a sequential priming task (Payne et al., 2005) in which the target person was presented against: (a) the background of the first learning block, (b) the background of the second learning block, or (c) a novel background that was not part of the learning task. Results showed that after the first learning block implicit evaluations reflected the valence of the information acquired in that block regardless of the background against which the target person was presented. More importantly, implicit evaluations after the second learning block reflected the valence of the initially learned information when the target individual was presented against the background of the first learning block or a novel background that was not part of the learning task; the subsequently presented counterattitudinal information influenced implicit evaluations only when the target was presented against the background of the second learning block (see Fig. 2.8). All of these effects were limited to implicit evaluations of the target individual of the learning task and did not generalize to implicit evaluations of other individuals. This result suggests that the contexts as such did not become associated with a corresponding evaluative response; instead, context modulated the response that was elicited by the target. Expanding on these findings, Gawronski, Rydell, et al. (2010) provided further evidence for their representational account by showing that (a) the impact of initial experiences was reduced for implicit evaluations in novel contexts when context salience during the encoding of initial information was enhanced; (b) context effects were eliminated altogether when context salience during the encoding of counterattitudinal information was reduced; and (c) enhanced context salience during the encoding of counterattitudinal information produced context-dependent automatic evaluations even when there was no contingency between valence and contextual cues.

Gawronski, Rydell, et al.'s (2010) representational account fills the bemoaned theoretical gap by specifying the conditions under which different contexts activate either the same or different associations in response to the same stimulus. However, their account has not yet been integrated with the core assumptions of the APE model, which raises a number of important questions. First, the empirical tests that have been conducted thus far focused exclusively on implicit evaluations. Thus, it remains an open question if (or when) similar effects emerge on explicit evaluations. From the perspective of the APE model, one could argue that context effects on explicit evaluations should depend on whether the affective reactions resulting from activated associations are consistent with other momentarily considered information. However, given that context effects on implicit evaluations occur only when there is evaluatively inconsistent information about an object (Rydell & Gawronski, 2009), a derivation of specific predictions for explicit evaluations is not really straightforward. One possibility is that context cues influence which information comes to mind most rapidly upon encountering the target object. With increasing delays,

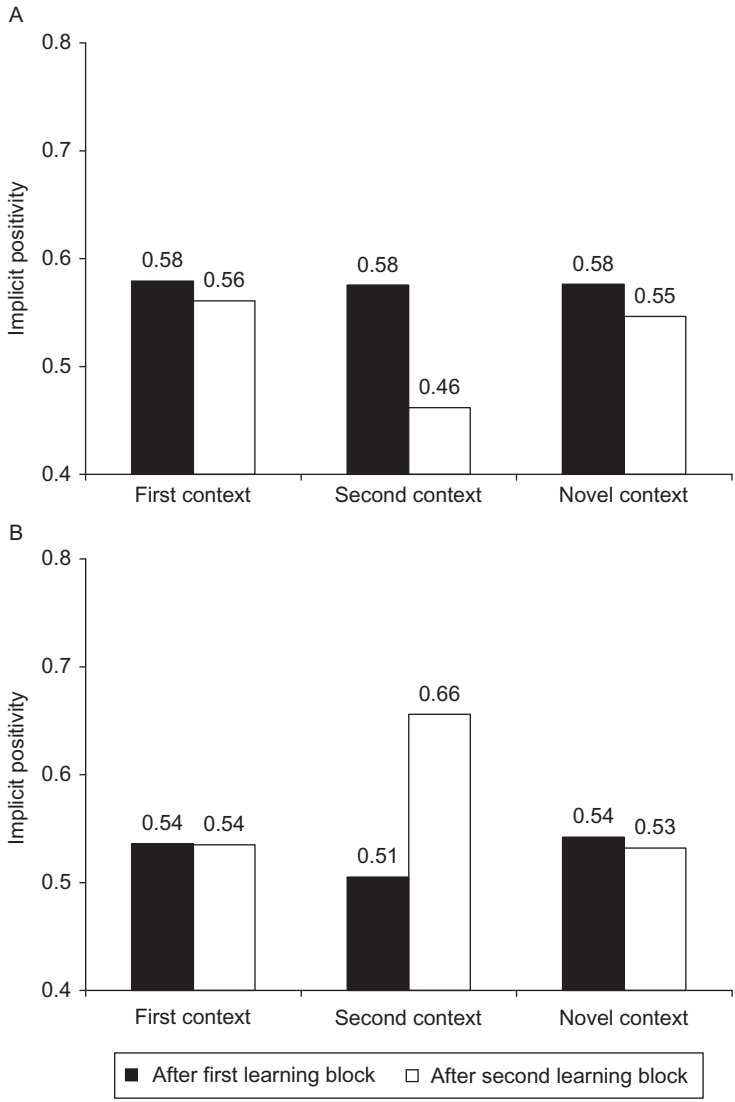

Figure 2.8 Implicit positivity toward target individual as a function of context (context of first learning block vs. context of second learning block vs. novel context) and time (after first learning block vs. after second learning block). (A) Evaluations when participants learned positive information in the first learning block and negative information in the second learning block (Rydell & Gawronski, 2009, Experiment 3); (B) Evaluations when participants learned negative information in the first learning block and positive information in the second learning block (Rydell & Gawronski, 2009, Experiment 4). Adapted from Rydell and Gawronski (2009), reprinted with permission.

however, perceivers may additionally retrieve other information from memory, including less accessible information that has been learned in other contexts. If these speculations are correct, the emergence of context effects on explicit evaluations may depend on the weight that is given to the two kinds of information in the course of making an evaluative judgment. To the extent that perceivers give more weight to information that comes to mind easily, the obtained context effects may well generalize to explicit evaluations. If, however, less accessible information is given equal weight in an integrated judgment that combines all available information regardless of how rapidly it comes to mind, the context effects obtained for implicit evaluations may not necessarily generalize to explicit evaluations. Drawing on research on the ease-of-retrieval heuristic (Schwarz et al., 1991), a potential moderator of the two outcomes could be the personal relevance of the attitude object, such that information that comes to mind rapidly may have a stronger impact when personal relevance is low, but not when it is high (see Rothman & Schwarz, 1998). Future research investigating the effects of contextual cues on explicit evaluations under conditions of high versus low relevance may provide deeper insights in this regard.

Another question is whether similar effects can be obtained in situations that involve processes of associative rather than propositional learning. A central assumption of Gawronski, Rydell, et al.'s (2010) account is that expectancy violations trigger a search for contextual factors that may explain the observed discrepancy, thereby drawing attention to momentarily available contextual cues (see Roese & Sherman, 2007). As a result, these cues are integrated into a contextualized representation of the object which includes the newly acquired, counterattitudinal information and the particular context in which it was acquired. However, as processes of associative learning may not necessarily involve the formation of conscious expectations that could be confirmed or disconfirmed (see Perruchet, Cleeremans, & Destrebcqz, 2006), it is not clear whether similar effects can be obtained for lower-level processes of associative learning. Future research combining Rydell and Gawronski's (2009) contextualized learning paradigm with tasks that involve the direct formation of associative links on the basis of mere co-occurrences (e.g., Olson & Fazio, 2001) may help to provide deeper insights into the learning mechanisms underlying context effects on implicit evaluations.

7. Conclusions

Inspired by evidence showing that implicit and explicit evaluations provide unique contributions to the prediction of behavior (Friese, Hofmann, & Schmitt, 2008; Perugini, Richetin, & Zogmaister, 2010), social psychologists have become increasingly interested in the causal antecedents of the two kinds of responses. In addition to providing a theoretical

integration of the heterogeneous findings on implicit and explicit evaluation, the APE model includes specific predictions about the conditions under which a given factor should produce (a) changes in implicit but not explicit evaluations, (b) changes in explicit but not implicit evaluations, or (c) corresponding changes in both implicit and explicit evaluations. The available evidence provides strong support for the predictions of the APE model and this evidence goes far beyond its original application to attitude change. Although there are several intriguing issues for future research to address, we believe that the APE model provides a comprehensive framework for understanding the psychological processes underlying implicit and explicit evaluations.

ACKNOWLEDGMENT

Preparation of this chapter was supported by grants from the Canada Research Chairs Program and the Social Sciences and Humanities Research Council of Canada awarded to B. G.

REFERENCES

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*, 652–661.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Aronson, E., & Cope, V. (1968). My enemy's enemy is my friend. *Journal of Personality and Social Psychology*, *8*, 8–12.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie*, *48*, 145–160.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5–22.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic activation with a pronunciation task. *Journal of Experimental Social Psychology*, *32*, 104–128.

Bargh, J. A., Litt, J., Pratto, F., & Spielman, L. A. (1989). On the preconscious evaluation of social stimuli. In A. F. Bennett & K. M. McConkey (Eds.), *Cognition in individual and social contexts: Proceedings of the XXV International Congress of Psychology* (pp. 357–370). Amsterdam: Elsevier.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82–93.

Blair, I. V., Ma, J., & Lenton, A. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*, 828–841.

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, *79*, 631–643.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, *8*, 539–546.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology*, *52*, 384–389.

Brochu, P. M., Gawronski, B., & Esses, V. M. (in press). The integrative prejudice framework and different forms of weight prejudice: An analysis and expansion. *Group Processes and Intergroup Relations*.

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.

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological Bulletin*, *125*, 47–63.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

Corneille, O., Yzerbyt, V., Pleyers, G., & Mussweiler, T. (2009). Beyond awareness and resources: Evaluative conditioning may be sensitive to processing goals. *Journal of Experimental Social Psychology*, *45*, 279–282.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Cunningham, W. A., Raye, C. I., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience*, *16*, 1717–1729.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. Wiers & A. Stacy (Eds.), *Handbook of implicit cognition and addiction* (pp. 11–28). New York: Sage.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, *10*, 230–241.

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2010). *A functional-cognitive framework for attitude research*. Manuscript submitted for publication.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*, 347–368.

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.

Degner, J., & Wentura, D. (2008). The extrinsic affective Simon task as an instrument for indirect assessment of prejudice. *European Journal of Social Psychology*, *38*, 1033–1043.

Deutsch, R., & Gawronski, B. (2009). When the method makes a difference: Antagonistic effects on "automatic evaluations" as a function of task characteristics of the measure. *Journal of Experimental Social Psychology*, *45*, 101–114.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385–405.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology*, *86*, 345–355.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568–584.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). San Diego, CA: Academic Press.

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.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row Peterson.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–210.

Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter? The use of automatic associations towards ethnic groups in person judgments. *Journal of Experimental Social Psychology*, *37*, 518–524.

Franco, F. M., & Maass, A. (1999). Intentional control over prejudice: When the choice of the measure matters. *European Journal of Social Psychology*, *29*, 469–477.

Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, *19*, 285–338.

Fujita, K., & Han, H. A. (2009). Moving beyond deliberative control of impulses: The effect of construal levels on evaluative associations in self-control conflicts. *Psychological Science*, *20*, 799–804.

Galdi, S., Gawronski, B., Arcuri, L., & Friese, M. (2010). *Selective exposure in decided and undecided individuals: Differential relations to automatic associations and conscious beliefs*. Manuscript submitted for publication.

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology*, *50*, 141–150.

Gawronski, B., & Bodenhausen, G. V. (2005). Accessibility effects on implicit social cognition: The role of knowledge activation versus retrieval experiences. *Journal of Personality and Social Psychology*, *89*, 672–685.

Gawronski, B., & Bodenhausen, G. V. (2006a). 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. (2006b). Associative and propositional processes in evaluation: Conceptual, empirical, and meta-theoretical issues. Reply to Albarracín, Hart, and McCulloch (2006), Kruglanski and Dechesne (2006), and Petty and Briñol (2006). *Psychological Bulletin*, *132*, 745–750.

Gawronski, B., & Bodenhausen, G. V. (2007a). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition*, *25*, 687–717.

Gawronski, B., & Bodenhausen, G. V. (2007b). What do we know about implicit attitude measures and what do we have to learn? In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 265–286). New York: Guilford Press.

Gawronski, B., & Bodenhausen, G. V. (2009). Operating principles versus operating conditions in the distinction between associative and propositional processes. *Behavioral and Brain Sciences*, *32*, 207–208.

Gawronski, B., Bodenhausen, G. V., & Becker, A. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*, 221–232.

Gawronski, B., Brochu, P. M., Sritharan, R., & Strack, F. (in press). Cognitive consistency in prejudice-related belief systems: Integrating old-fashioned, modern, aversive and implicit forms of prejudice. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press.

Gawronski, B., & Creighton, L. A. (in press). Dual-process theories. In D. E. Carlston (Ed.), *The Oxford handbook of social cognition*. New York: Oxford University Press.

Gawronski, B., Cunningham, W. A., LeBel, E. P., & Deutsch, R. (2010). Attentional influences on affective priming: Does categorization influence spontaneous evaluations of multiply categorizable objects? *Cognition and Emotion*, *24*, 1008–1025.

Gawronski, B., Deutsch, R., & Banse, R. (in press). Response interference tasks as indirect measures of automatic associations. In K. C. Klauer, C. Stahl, & A. Voss (Eds.), *Cognitive methods in social psychology*. New York: Guilford Press.

Gawronski, B., Deutsch, R., LeBel, E. P., & Peters, K. R. (2008). Response interference as a mechanism underlying implicit measures: Some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, *24*, 218–225.

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic stereotype activation. *Journal of Experimental Social Psychology*, *44*, 370–377.

Gawronski, B., Geschke, D., & Banse, R. (2003). Implicit bias in impression formation: Associations influence the construal of individuating information. *European Journal of Social Psychology*, *33*, 573–589.

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition*, *15*, 485–499.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*, 1355–1361.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science*, *2*, 181–193.

Gawronski, B., Peters, K. R., Brochu, P. M., & Strack, F. (2008). Understanding the relations between different forms of racial prejudice: A cognitive consistency perspective. *Personality and Social Psychology Bulletin*, *34*, 648–665.

Gawronski, B., Peters, K. R., & Strack, F. (2008). Cross-cultural differences vs. universality in cognitive dissonance: A conceptual reanalysis. In R. M. Sorrentino & S. Yamaguchi (Eds.), *Handbook of motivation and cognition across cultures* (pp. 297–314). New York: Elsevier.

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

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

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology*, *40*, 535–542.

Gawronski, B., Strack, F., & Bodenhausen, G. V. (2009). Attitudes and cognitive consistency: The role of associative and propositional processes. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 85–117). New York: Psychology Press.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin*, *34*, 1276–1289.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology*, *41*, 618–626.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research*, *35*, 178–188.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, *46*, 107–119.

Green, D. P., Ha, S. E., & Bullock, J. G. (2010). Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Sciences*, *628*, 200–208.

Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 147–170). New York: Academic Press.

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist*, *47*, 766–779.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Gregg, A. P., Seibt, B., & Banaji, M. R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, *90*, 1–20.

Gross, J. J. (1998). Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology*, *74*, 224–237.

Grumm, M., Nestler, S., & von Collani, G. (2009). Changing explicit and implicit attitudes: The case of self-esteem. *Journal of Experimental Social Psychology*, *45*, 327–335.

Gschwendner, T., Hofmann, W., & Schmitt, M. (2008). Differential stability: The effects of acute and chronic construct accessibility on the temporal stability of the Implicit Association Test. *Journal of Individual Differences*, *29*, 70–79.

Halgren, E. (1992). Emotional neurophysiology of the amygdala within the context of human cognition. In J. Aggleton (Ed.), *The amygdala: Neurobiological aspects of emotion, memory, and mental dysfunction* (pp. 191–228). New York: Wiley.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heine, S. J., & Lehman, D. R. (1997). Culture, dissonance, and self-affirmation. *Personality and Social Psychology Bulletin*, *23*, 389–400.

Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology*, *52*, 180–186.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming

as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, *40*, 217–234.
Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*, 390–421.
Hofmann, W., Deutsch, R., Lancaster, K., & Banaji, M. R. (2010). Cooling the heat of temptation: Mental self-control and the automatic evaluation of tempting stimuli. *European Journal of Social Psychology*, *40*, 17–25.
Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.
Hoshino-Browne, E., Zanna, A. S., Spencer, S. J., Zanna, M. P., Kitayama, S., & Lackenbauer, S. (2005). On the cultural guises of cognitive dissonance: The case of Easterners and Westerners. *Journal of Personality and Social Psychology*, *89*, 294–310.
Houben, K., Havermans, R. C., & Wiers, R. W. (2010). Learning to dislike alcohol: Conditioning negative implicit attitudes toward alcohol and its effect on drinking behavior. *Psychopharmacology*, *211*, 79–86.
Insko, C. A. (1984). Balance theory, the Jordan paradigm, and the Wiest tetrahedron. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 89–140). San Diego, CA: Academic Press.
Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*, 616–626.
Jacoby, J., & Sassenberg, K. (2011). Interactions do not only tell us when, but can also tell us how: Testing process hypotheses by interaction. *European Journal of Social Psychology*, *41*, 180–190.
Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology*, *96*, 933–948.
Jones, E. E., & Gerard, H. B. (1967). *Foundations of social psychology*. New York: Wiley.
Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Evaluative conditioning: The "how" question. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 43, pp. 205–255). New York: Academic Press.
Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.
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.
Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any "free" choice: Self and dissonance in two cultures. *Psychological Science*, *15*, 527–533.
Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process-components of the Implicit Association Test: A diffusion model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.
Knight, D. C., Nguyen, H. T., & Bandettini, P. A. (2003). Expression of conditional fear with and without awareness. *Proceedings of the National Academy of Sciences*, *100*, 15280–15283.
Koole, S. L., Dijksterhuis, A., & Van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*, 669–685.
Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.
Langer, T., Walther, E., Gawronski, B., & Blank, H. (2009). When linking is stronger than thinking: Associative transfer of valence disrupts the emergence of cognitive balance. *Journal of Experimental Social Psychology*, *45*, 1232–1237.

LeBel, E. P. (2010). Attitude accessibility as a moderator of implicit and explicit self-esteem correspondence. *Self and Identity*, *9*, 195–208.

Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology*, *48*, 563–574.

Lieberman, M. D. (2003). Reflective and reflexive judgment processes: A social cognitive neuroscience approach. In J. P. Forgas, K. R. Williams, & W. von Hippel (Eds.), *Social judgments: Implicit and explicit processes* (pp. 44–67). New York: Cambridge University Press.

Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science*, *12*, 135–140.

Logan, G. D. (1988). Toward and instance theory of automatization. *Psychological Review*, *95*, 492–527.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology*, *69*, 397–407.

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*, 224–253.

Martinie, M.-A., Olive, T., & Milland, L. (2010). Cognitive dissonance induced by writing a counterattitudinal essay facilitates performance on simple tasks but not on complex tasks that involve working memory. *Journal of Experimental Social Psychology*, *46*, 587–594.

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs. "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*, 433–449.

McConahay, J. B. (1986). Modern racism, ambivalence, and the modern racism scale. In J. D. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91–125). Orlando, FL: Academic Press.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology*, *46*, 961–978.

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

Mitchell, D. G. V., Nakic, M., Pine, D. S., & Blair, R. J. R. (2007). The impact of processing load on emotion. *Neuroimage*, *34*, 1299–1309.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455–469.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, *67*, 949–971.

Morsella, E., Zarolia, P., & Gazzaley, A. (in press). Cognitive conflict and consciousness. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*, 184–211.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413–447.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636–639.

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

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*, 272–314.

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

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

Perruchet, P., Cleeremans, A., & Destrebcqz, A. (2006). Dissociating the effects of automatic activation and explicit expectancy on reaction times in a simple associative learning task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *32*, 955–965.

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 255–277). New York: Guilford Press.

Peters, K. R., & Gawronski, B. (2011a). Mutual influences between the implicit and explicit self-concepts: The role of memory activation and motivated reasoning. *Journal of Experimental Social Psychology*, *47*, 436–442.

Peters, K. R., & Gawronski, B. (2011b). Are we puppets on a string? Comparing the effects of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, *37*, 557–569.

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., Cacioppo, J. T., & Schumann, D. W. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. *Journal of Consumer Research*, *10*, 135–146.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 130–144.

Prestwich, A., Perugini, M., Hurling, R., & Richetin, J. (2010). Using the self to change implicit attitudes. *European Journal of Social Psychology*, *40*, 61–71.

Quinn, K. A., Hugenberg, K., & Bodenhausen, G. V. (2004). Functional modularity in stereotype representation. *Journal of Experimental Social Psychology*, *40*, 519–527.

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

Richeson, J. A., & Nussbaum, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417–423.

Richter, T., Schroeder, S., & Wöhrmann, B. (2009). You don't have to believe everything you read: Background knowledge permits fast and efficient validation of information. *Journal of Personality and Social Psychology*, *96*, 538–558.

Roefs, A., Quaedacjers, L., Werrij, M. Q., Wolters, G., Havermans, R., Nederkoorn, C., et al. (2006). The environment influences whether high-fat foods are associated with palatable or with unhealthy. *Behaviour Research and Therapy*, *44*, 715–736.

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.

Rothman, A. J., & Schwarz, N. (1998). Construal perception of vulnerability: Personal relevance and the use of experiential information in health judgments. *Personality and Social Psychology Bulletin*, *24*, 1053–1064.

Rozin, P., Markwith, M., & Ross, B. (1990). The sympathetic magical law of similarity, nominal realism, and neglect of negatives in response to negative labels. *Psychological Science*, *1*, 383–384.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*, 79–82.

Rudman, L. A., & Lee, M. R. (2002). Implicit and explicit consequences of exposure to violent and misogynous rap music. *Group Processes and Intergroup Relations*, *5*, 133–150.

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., Mackie, D. M., & Strain, L. M. (2006). Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes. *Psychological Science*, *17*, 954–958.

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., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. A., & Groom, C. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–335.

Sherman, J. W., Klauer, K. C., & Allen, T. J. (2010). Mathematical modeling of implicit social cognition: The machine in the ghost. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 156–175). New York: Guilford Press.

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

Smith, C. T., & Nosek, B. A. (in press). Affective focus increases the concordance between implicit and explicit attitudes. *Social Psychology*.

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, *89*, 845–851.

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.

Stahl, C., & Degner, J. (2007). Assessing automatic activation of valence: A multinomial model of EAST performance. *Experimental Psychology*, *54*, 99–112.

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology. Animal Behavior Processes*, *35*, 286–291.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology*, *97*, 404–420.

Stone, J. (in press). Consistency as a basis for behavioral interventions: Using hypocrisy and cognitive dissonance to motivate behavior change. In B. Gawronski, & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition*. New York: Guilford Press.

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

Sweldens, S., Van Osselaer, S., & Janiszewski, C. (2010). Evaluative conditioning procedures and the resilience of conditioned brand attitudes. *Journal of Consumer Research*, *37*, 473–489.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778–793.

Trope, Y., & Liberman, A. (1996). Social hypothesis testing: Cognitive and motivational mechanisms. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 239–270). New York: Guilford Press.

Verosky, S. C., & Todorov, A. (2010). Generalization of affective learning about faces to perceptually similar faces. *Psychological Science*, *21*, 779–785.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology*, *82*, 919–934.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the backdoor: The US-revaluation effect. *Cognition and Emotion*, *23*, 889–917.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four picture recognition test. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 454–459.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52.

Wellens, A. R., & Thistlewaite, D. L. (1971). An analysis of two quantitative theories of cognitive balance. *Psychological Review*, *78*, 141–150.

Wheeler, M. E., & Fiske, S. T. (2005). Controlling racial prejudice: Social-cognitive goals affect amygdala and stereotype activation. *Psychological Science*, *16*, 56–63.

Whitfield, M. (2009). *Intuition, rational analysis and the relation between implicit and explicit attitudes*. Unpublished dissertation. Wilfrid Laurier University, Canada.

Whitfield, M., & Jordan, C. H. (2009). Mutual influences of explicit and implicit attitudes. *Journal of Experimental Social Psychology*, *45*, 748–759.

Wiest, N. M. (1965). A quantitative extension of Heider's theory of cognitive balance applied to interpersonal perception and self-esteem. *Psychological Monographs: General & Applied*, *79*, 1–20.

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

Winkielman, P., Niedenthal, P. M., & Oberman, L. M. (2009). Embodied perspective on emotion-cognition interactions. In J. A. Pineda (Ed.), *Mirror neuron systems* (pp. 235–257). New York: Humana Press.

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationships with questionnaire measures. *Journal of Personality and Social Psychology*, *72*, 262–274.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.

Zhang, H., & Chan, D. K.-S. (2009). Self-esteem as a source of evaluative conditioning. *European Journal of Social Psychology*, *39*, 1065–1074.