Ψ Psychology Press
Taylor & Francis Group

What moderates implicit–explicit consistency?

Wilhelm Hofmann, Tobias Gschwendner
University of Koblenz-Landau, Germany

Brian A. Nosek
University of Virginia, USA

Manfred Schmitt
University of Koblenz-Landau, Germany

Implicit and explicit indicators of attitudes or personality traits are positively, and variably, related. This review places the question of implicit–explicit consistency into the tradition of attitude/trait–behaviour consistency (e.g., Wicker, 1969). Drawing on dual-process models, such as the recent distinction between associative and propositional representations (Strack & Deutsch, 2004), we identify a working model of implicit–explicit consistency that organises the empirical evidence on implicit–explicit moderation into five factors: *translation* between implicit and explicit representations (e.g., representational strength, awareness), additional *information integration* for explicit representations (e.g., need for cognition), properties of *explicit assessment* (e.g., social desirability concerns), properties of *implicit assessment* (e.g., situational malleability), and research *design factors* (e.g., sampling bias, measurement correspondence).

A significant proportion of psychological research over the last three decades concerns the automatic nature of information processing (Bargh, 1997; Khilstrom, 1999; Wegner & Bargh, 1998). Theory and empirical data have broadened notions of core psychological concepts like attitudes, stereotypes, self-concept, goals, personality, and self-esteem to include not

just the explicit assessments that are products of introspection, but also implicit components of these constructs that may occur outside of conscious awareness or control (Asendorpf, Banse, & Mücke, 2002; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). Whereas implicit constructs are assumed to operate automatically and may be inaccessible to conscious experience, their explicit counterparts are conceptualised as reflective (conscious) and capacity-consuming mental representations that influence action through deliberation (e.g., Bargh, 1994; Greenwald & Banaji, 1995).

This distinction plays a role in a variety of dual-process theories that distinguish two modes of information processing such as implicit vs explicit (Greenwald & Banaji, 1995; Wilson et al., 2000), automatic vs controlled (Bargh, 1994), impulsive vs reflective (Strack & Deutsch, 2004), and associative vs rule-based (Sloman, 1996; Smith & DeCoster, 2000). These theories share an assumption that the two modes of processing will be differentially influential on perception, judgement, and action depending on the availability of cognitive resources or other situational constraints. However, the theories differ widely in their presumptions about how the two modes of processing interact, if at all (Gilbert, 1999).

The theoretical dual-process distinctions are paralleled by a recent surge in measurement innovation with tools designed to assess associations indirectly. There is great variability in the new methods, but all share a difference from traditional self-report in that their measurement features capture one or more of Bargh's (1994) hallmarks of automaticity: lack of intention, lack of awareness, lack of control, or efficiency of processing. Some of the techniques include: evaluative priming (e.g., Fazio, Jackson, Dunton, & Williams, 1995; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998), the Go/No-go Association Task (GNAT, Nosek & Banaji, 2001), the lexical-decision task (e.g., Wittenbrink, Judd, & Park, 1997), the Affective Simon task (De Houwer, 2003; De Houwer & Eelen, 1998), the Evaluative Movement Assessment (Brendl, Markman, & Messner, 2005), the affect misattribution paradigm (Payne, Cheng, Govorun, & Stewart, 2005), the word fragment completion task (e.g., Dovidio, Kawakami, Johnson, & Johnson, 1997; Hetts, Sakuma, & Pelham, 1999), the name-letter task (e.g., Koole, Dijksterhuis, & van Knippenberg, 2001; Nuttin, 1985), and the analysis of linguistic biases (e.g., Franco & Maass, 1999; von Hippel, Sekaquaptewa, & Vargas, 1997).

Probably the most widely used implicit measurement technique is the Implicit Association Test (IAT) developed by Greenwald et al. (1998). We describe the IAT procedure in more detail because of the large body of research it has stimulated across fields of psychology (see Nosek, Greenwald, & Banaji, in press, for a recent review) making it the most cited measure in this review (for descriptions of other implicit measures see

Blair, 2002; Brauer, Wasel, & Niedenthal, 2000). The IAT measures associations by comparing average response times for categorising stimulus exemplars into their superordinate categories in two distinct response conditions. Items belonging to four categories (e.g., Male, Female, Self, Other) are categorised as quickly as possible using two response keys. In one condition, items representing two of the categories (e.g., Male and Self) are categorised with one response key, while items representing the other two categories (e.g., Female and Other) are categorised with the alternate key. In the second condition, the response mapping is switched. For this example, Female and Self items would be categorised with one key and Male and Other items would be categorised with the alternate key. The core logic of the task is that it should be easier (faster) to make the same response for two categories when those categories are associated compared to when they are not. So, most women categorise items more quickly when Female and Self items share a response key, whereas most men categorise items more quickly when Male and Self share a response key. The difference in average response latency between the two conditions provides a comparative assessment of association strengths between the two response conditions.

The IAT and other implicit methods sometimes reveal mental representations that are starkly different from those indicated by the more common explicit methods. For example, Fazio et al. (1995) found an implicit racial attitude measure based on evaluative priming to be unrelated to the Modern Racism Scale (McConahay, 1986) even though both measures showed predictive validity of race-related judgements and behaviour. Findings like these lend credence to the suggestion that the measures assess distinct processes that have different consequences for behaviour. At the same time, implicit and explicit measures are sometimes strongly associated. For instance, from data collected about the 2000 US presidential election, Greenwald, Nosek, and Banaji (2003) reported a correlation of .86 between self-report and an Implicit Association Test (IAT) measuring attitudes towards George Bush relative to Al Gore.

While dual-process theories are well specified in distinguishing modes of processing and how they influence action, few identify *when* the dual processes will elicit similar or distinct responses. The purpose of this chapter is to review the existing evidence of variables that moderate implicit–explicit consistency, i.e., the relationship between implicit and explicit indicators of attitudes and traits. First, we note the conceptual similarity of identifying moderators of implicit–explicit correspondence to the trait–behaviour consistency controversies of the last century (Epstein & O'Brien, 1985). Second, we introduce a working model of implicit–explicit consistency that identifies five superordinate factors that influence when implicit and explicit measurement will be associated or dissociated. Third, we review the empirical evidence on moderators of implicit–explicit consistency using

the five factors as an organisational scheme. This working model should provide a framework for pursuing continued theoretical development towards a comprehensive model of the relationship between implicit and explicit constructs.

While the implicit–explicit distinction is relevant to a wide variety of constructs, domains, and measurement approaches, this review is constrained by the existing literature. The literature is not a random sampling of domains and methods. The most common applications concern investigations of the attitude construct, racial stereotypes, and prejudice as a content domain, and the Implicit Association Test (IAT) as an implicit measurement method. Findings with other constructs including stereotypes, self-concept, self-esteem, and personality are included, as are findings from other measurement methods, most notably evaluative priming and lexical decision tasks, but the skew in the literature is obvious. If nothing else, this review identifies the importance of broadening the investigation of implicit–explicit consistency to a wider variety of constructs and measurement methods to determine whether the present conclusions are particular to a subset of methodologies or content domains.

A BRIEF HISTORY OF ATTITUDE/TRAIT–BEHAVIOUR CONSISTENCY CONTROVERSIES

Most attitude and personality trait models assume that there exist stable and generalisable differences in attitudes or personality characteristics between persons (e.g., Thurstone & Chave, 1929). If persons differ on these dimensions, and if these differences are relatively stable over time, then the indicators of attitudes or traits should effectively describe individuals and predict behaviour. In the history of psychology, this claim has been the focus of considerable debate. The controversy first appeared in the 1930s with the seminal studies on attitude–behaviour consistency by LaPiere (1934) and on trans-situational consistency of moral behaviour by Hartshorne and May (1928). LaPiere (1934) found that American hosts' self-reported willingness to accommodate Chinese guests was almost unrelated to their actual behaviour. Similarly, in the Hartshorne and May studies, different behavioural indicators of honesty correlated weakly, .19 on average.

A second surge in this debate was sparked off by Wicker's (1969) review on attitude–behaviour consistency and Mischel's (1968) book on the predictive validity of personality and achievement tests. Wicker (1969) argued that (explicitly assessed) attitudes and behaviour were correlated .30 or less, and that attitudes, therefore, accounted for a maximum of 10% of the variance in observed behaviour. He concluded that there was "little evidence to support the postulated existence of stable, underlying attitudes

within the individual which influence both his verbal expressions and his actions'' (p. 75). In a similar vein, Mischel (1968) stated that self-report trait measures and behaviour were correlated .30 at best. Both Wicker (1969) and Mischel (1968) concluded that behaviour was predominantly a function of the situation, challenging the long-standing assumption that behaviour could be predicted by relatively stable attitudes or traits.

Both controversies and the subsequent rejoinders to the criticisms were characterised by a typical sequence of argumentation and empirical investigations that can be described as generations of research investigations (Zanna & Fazio, 1982). First, the debate focused on the absolute degree of consistency—*whether* a relationship exists, and if so, *how much* of a relationship. Second, attention shifted to methodological explanations which attributed inconsistencies to poor measurement techniques (e.g., Ajzen & Fishbein, 1977) and a search for psychological *moderator variables* that identify *when* attitudes or traits reliably predict behaviour (Ajzen, 1987; Bem & Allen, 1974; Block, 1977; Schmitt, 1990; Snyder & Ickes, 1985). Third, the identification and cataloguing of moderators provided an opportunity to develop theoretical models illustrating a deeper under-standing of *why* traits and behaviours are highly related in some cases and not others (e.g., Kenrick & Funder, 1988). This sequence bears notable similarities (and differences) to the present investigations of implicit – explicit consistency.

SIMILARITIES AND DIFFERENCES BETWEEN ATTITUDE/TRAIT – BEHAVIOUR AND IMPLICIT – EXPLICIT CONSISTENCY

Research on implicit – explicit consistency is following a course that closely resembles the earlier trait – behaviour debates. Researchers first wrestled with the same question of *whether* and *how much* implicit – explicit con-sistency exists. However, while the implications of strong, positive relations were considered important to establish convergent validity for attitude – behaviour relations, the opposite goal was evident for initial research on implicit – explicit consistency. Weak trait – behaviour relations were a threat to the trait theorists' suggestion that stable inter-individual differences exist and are relevant for predicting behaviour, and similar weak implicit – explicit relations were considered initial evidence for the notion of distinct implicit and explicit constructs (e.g., Greenwald & Banaji, 1995).

Interest in implicit measures was fostered both by dissatisfaction with measures that required a willingness to self-disclose, and by theoretical innovations postulating implicit constructs that are inaccessible to introspection. If implicit and explicit measures are redundant (i.e., highly related), then the notion that they serve as indicators of distinct constructs is

undermined (Nosek & Smyth, 2005). As such, early theory and research on the relationship between implicit and explicit measures emphasised their dissociation (Banaji, 2001).

Meanwhile, however, a robust literature has accumulated on the implicit – explicit relationship allowing for a quantitative summary of consistency across a wide range of domains. Hofmann, Gawronski, Gschwendner, Le, & Schmitt (2005a), for example, meta-analysed a total of 126 studies from 53 different content domains that had both an IAT and an explicit measure of the same construct, and obtained a mean corrected implicit – explicit correlation of .24. Likewise, a study that assessed attitudes with the IAT and explicit warmth ratings across 57 content domains reported an average correlation of .36 (Nosek, 2005).[1] Thus, the accumulated evidence suggests that implicit and explicit measures are positively related, refuting the notion of complete dissociation among implicit and explicit constructs.

Also, like attitude/trait – behaviour consistency (e.g., Kraus, 1995), implicit – explicit consistency appears to be highly variable (e.g., Blair, 2001; Brauer et al., 2000; Dovidio, Kawakami, & Beach, 2001; Hofmann et al., 2005a). In the Nosek (2005) investigation, for example, implicit – explicit correlations ranged from a low of – .05 (male – female attitudes) to a high of .70 (pro choice – pro life attitudes). Thus, average coefficients about the *how much* of implicit – explicit consistency do not capture the complexity of findings. Rather, the high variability in correlations suggests that moderator variables exist that determine *when* correlations between implicitly and explicitly assessed responses are high or low. As with the earlier trait – behaviour consistency research, the field is moving to a second generation of research identifying these moderators. This growing body of research has not been summarised, providing the impetus for this review.

Even though implicit – explicit consistency research could be characterised as being in the second generation of theoretical development, there are a few theoretical models that identify *why* implicit – explicit consistency should occur. One is the MODE model (Fazio, 1990) that was originally applied to the attitude – behaviour relationship but could be similarly relevant to the question of implicit – explicit consistency (Fazio & Olson, 2003). The MODE model distinguishes between spontaneous and deliberate processes by which an attitude may influence judgements and behaviour. Implicit indicators are assumed to reflect the spontaneous process, and explicit indicators will assess

[1]Somewhat lower correlation in the Hofmann et al. (2005a) study may be due to the inclusion of a more heterogeneous set of explicit measures, some of which corresponded less well with the implicit measures in terms of content and specificity than the measures employed by Nosek (2005). Also, sampling of content domains for the meta-analysed studies was driven by research interests of different laboratories and may ''oversample'' from domains in which a dissociation is expected, whereas the Nosek (2005) study sampled across a wide range of domains.

the same unless the report is altered by effortful, deliberate processing. For deliberate processing to occur, both opportunity (time and cognitive resources) and motivation are required. Thus, high implicit – explicit consistency is expected when individuals have neither the opportunity nor the motivation to adjust their explicit response. As opportunity and motivation to alter the automatic response increase, the consistency between implicit and explicit indicators should decline. In other words, the MODE model assumes a *single mental representation* that can either be measured immediately upon automatic activation (implicit), or after some deliberative processing has been invoked that could alter the automatic response (explicit). The MODE model is especially suited to accounting for moderator effects that reflect response biases in self-report such as self-presentation (e.g., Nosek, 2005), motivation to control prejudiced reactions (e.g., Dunton & Fazio, 1997) or spontaneity of judgement (e.g., Koole et al., 2001).

In contrast to the MODE model, most other dual-process theories propose that implicit and explicit assessments reflect distinct mental representations or operations (Devine, 1989; Smith & DeCoster, 2000; Strack & Deutsch, 2004; Wilson et al., 2000). These theories make few predictions about when implicit and explicit indicators will correspond, but compared to a single representation framework they have the advantage of being able to incorporate moderators that influence the similarity of mental representations, beyond response biases that influence self-report. Under-standing implicit – explicit consistency may be heightened by taking advantage of this notion of dual representation to spell out hypotheses about distinct features of implicit and explicit representations, to specify the processes by which the constructs may become transformed into each other, and to identify factors influencing how these hypothetical explicit and implicit constructs manifest themselves regarding their respective indicators. As reviewed below, there is evidence for a number of moderators that fit into a dual-representation framework, but would be difficult to integrate into a single-representation model.

In the following section, we suggest a working model of implicit – explicit consistency that draws on the distinction between associative and proposi-tional representations (Strack & Deutsch, 2004). This model integrates five factors by which the correspondence between implicit and explicit indicators is affected. The model then serves as an "advanced organiser" for an empirical review of moderator variables of implicit – explicit consistency.

A WORKING MODEL OF IMPLICIT – EXPLICIT CONSISTENCY

In accordance with dual-process frameworks (Smith & DeCoster, 2000; Strack & Deutsch, 2004), we assume that implicit and explicit representations

are two structurally separable components of an attitude or trait (see Figure 1). The implicit representation constitutes the associative component of an attitude or trait (e.g., Gawronski & Bodenhausen, 2004; Gawronski & Strack, 2004; Strack & Deutsch, 2004). Associations can be activated automatically, without awareness, without intention, and with minimal use of cognitive resources (Bargh, 1994). Furthermore, associative representations are independent of the assignment of truth values (e.g., Gawronski & Strack, 2004; Strack & Deutsch, 2004). In other words, these associations can be activated whether people approve or disapprove of them (see also Devine, 1989). For example, negative attitudes towards spiders can be conceived of as the association between the concept of spiders and negative evaluative nodes such as fear and disgust (Teachman, Gregg, & Woody, 2001). This association may become activated unintentionally and effortlessly whenever a spider is encountered—independent of whether a person consciously endorses the negative evaluation.

In contrast, the explicit component of an attitude or trait is mentally represented in a propositional format. Propositional representations (e.g., ''spiders are disgusting'') are the result of higher-order mental processes in the so-called reflective system which is designed to generate or transform declarative knowledge (Strack & Deutsch, 2004). Whereas associative representations can be activated and influence subsequent processing no matter whether a person considers them to be true or false, propositional representations generally influence explicit responses only when regarded as true (Gawronski & Strack, 2004).

Most important, we assume that implicit and explicit representations may feed into each other as indicated by the bi-directional causal paths between implicit and explicit latent representations shown in Figure 1. As argued in more detail below, implicit representations may form the elementary basis of higher-order propositional (explicit) representations (Strack & Deutsch, 2004). Likewise, propositional representations may have a retroactive influence on associative ones (Gawronski & Bodenhausen, 2004).

Like all hypothetical constructs, the implicit and explicit latent representations are identified as causal influences on their implicit and explicit indicators (measures), reflecting the fundamental connection between measurement and theory. There are many factors that influence the degree to which a measure appropriately reflects the construct of interest. And, formally speaking, implicit–explicit consistency is observed between indicators, not between the representations themselves. Translating an understanding of correlations between implicit and explicit measures into an understanding of the structure and function of the underlying representations requires consideration of both the factors that affect the correlation among the representations themselves and the factors that affect the indicators' measurement properties.
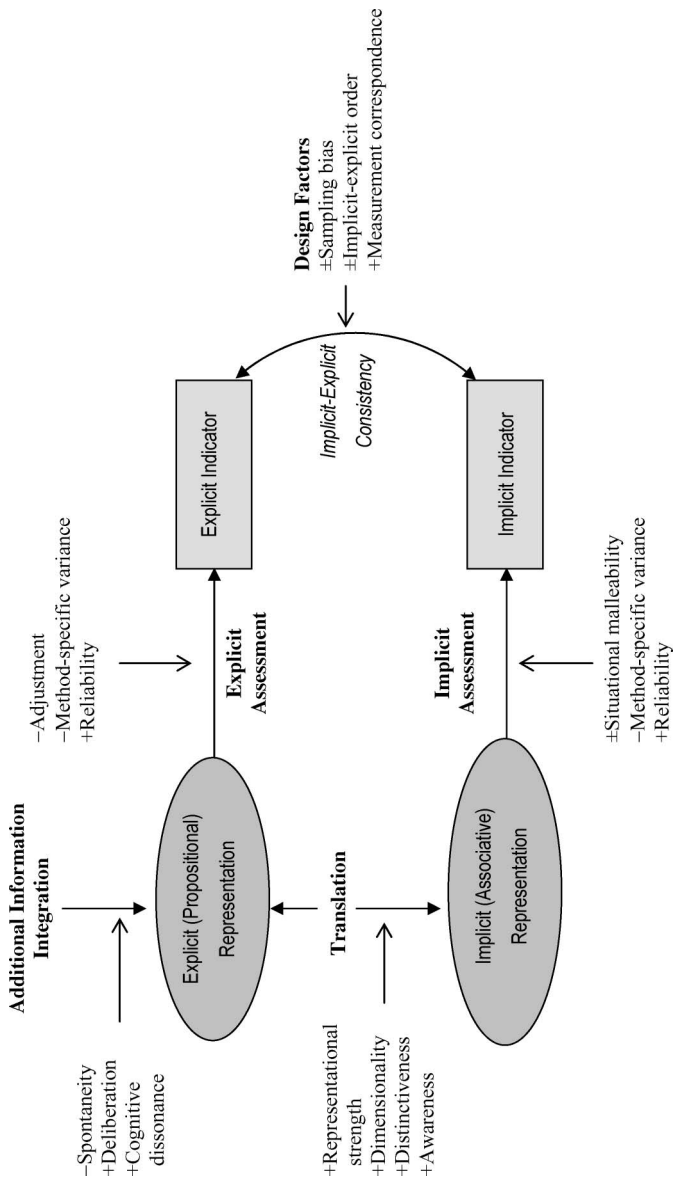
Figure 1. A working model of implicit−explicit consistency. Plus and minus signs indicate the hypothesised direction of moderator effects on the respective pathway or correlation. A ± sign indicates that both directions are possible (see text).

343

The proposed model reflects potential moderation of implicit–explicit consistency as occurring between the hypothesised representations and as a function of the measurement of those representations: The degree of consistency between implicit and explicit indicators is a function of how strongly associative and propositional representations feed into each other and how well the implicit and explicit indicators reflect their latent representations respectively. In our view, most of the empirical evidence on moderators of implicit–explicit consistency can be meaningfully organised into five factors that correspond to this model: *translation* between implicit and explicit representations, additional *information integration* for explicit representations, properties of *explicit assessment*, properties of *implicit assessment*, and research *design factors* such as sampling selectivity and measurement correspondence (see Figure 1). These factors provide a parsimonious, superordinate structure for organising a variety of moderator variables that are supported by empirical evidence. Also, this structure may be useful for incorporating additional moderators that have yet to be identified. Next, we provide an overview of the five factors and review the empirical evidence for their relevance for understanding implicit–explicit consistency (see the Appendix for a summary table of the reviewed studies).

TRANSLATION BETWEEN IMPLICIT AND EXPLICIT REPRESENTATIONS

It is plausible to assume that associative and propositional representations, while structurally distinct, have meaningful interrelations. Some theories suggest that momentarily activated associations form the elementary basis, or the "building blocks", of our propositional thinking (e.g., Gawronski & Bodenhausen, 2004; Strack & Deutsch, 2004). Likewise, propositional thinking may in turn shape the structure of the underlying associative representations. The quality of the *translation process* should hinge on a number of moderating factors that determine the extent to which associative processing influences propositional representations, and the extent to which reflective processing shapes associative representations.

Importantly, translation processes are hypothesised to be bi-directional. Associative structures may influence propositional representations such as the self-awareness moderators described below in which associative information provides an experience or produces a behaviour that is incorporated into one's propositional thinking. Likewise, propositional thinking may influence associative structures, such as Logan's (1988) instance theory of automatisation suggesting that the practice or repetition of a response increases the likelihood that it will be automatically activated in the future (see also Smith & DeCoster, 2000). For example, in one study

on spider phobia (Teachman & Woody, 2003), cognitive therapy involving countering maladaptive beliefs and gradual exposure led to a reduction of implicit spider – fear associations over the course of treatment. Likewise, multicultural training seminars (Rudman, Ashmore, & Gary, 2001) or stereotype negation programmes (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000) can reduce automatic race bias, lending further credence to the notion that propositional influences feed back on associative structures. The moderators reviewed next vary in their presumed directional causal implications. First, we discuss representational strength-related variables that may foster the flow of information among implicit and explicit representations. Then we mention the dimensionality (unipolar vs bipolar) and social distinctiveness of representations as additional moderators. Finally, we consider potential mechanisms and moderator variables that determine the extent to which people may become propositionally aware of their associative representations.

Representational strength

One translational moderator of implicit – explicit consistency may be the strength of the representations in memory. Strong attitudes are stable over time, resistant to persuasion, guide information processing, and strongly predict behaviour (Petty & Krosnick, 1995; Visser, Krosnick, & Simmons, 2003). This notion of attitude strength can be generalised as *representational strength* to be inclusive of other constructs such as stereotypes, personality characteristics, and self-concepts. Strong representations are more practised and, as a consequence, may show greater consistency between implicit and explicit representations (Logan, 1988; Smith & DeCoster, 2000).

Strength, especially in the domain of attitudes, is conceived of as a heterogeneous construct including importance, thought frequency, familiarity, cognitive elaboration, knowledge, stability, prior experience, and certainty (see Petty & Krosnick, 1995, for a review). Nosek (2005) tested a combined strength factor of attitude importance, thought frequency, and familiarity as a moderator of implicit – explicit consistency across 57 content domains in a large sample. A multilevel analysis revealed that stronger attitudes were associated with greater consistency between the IAT and self-reported evaluations than weaker attitudes. This approach showed the potency of strength as a moderator of implicit – explicit consistency irrespective of the content domain.

Similarly, Nosek and Banaji (2002) had raters estimate the average degree of *cognitive elaboration*—raters' judgements of the extent to which people in general would spend thinking about a topic—for 15 different content domains. Those ratings predicted the observed implicit – explicit consistency

across 15 large datasets in which the IAT and self-reported attitudes were measured at various websites. Domains rated as eliciting high cognitive elaboration also showed relatively strong implicit – explicit consistency. Unlike the multilevel analysis described above (Nosek, 2005), this study was limited to inferences across content domains because individual differences in cognitive elaboration were not assessed.

Comparable effects have been observed for inter-individual strength moderators in a variety of studies examining specific content domains. For example, differences in *importance* positively moderated implicit – explicit correspondence in a study on stereotyping (Hofmann, Gschwendner, & Schmitt, 2005b) and in two studies on voting attitudes and consumer attitudes (Karpinski, Steinman, & Hilton, 2005).

Further support was observed in a known-groups comparison of men and women concerning the importance of body weight. Women consistently report greater weight and body image concerns than men (e.g., Wadden, Brown, Foster, & Linowitz, 1991). And, consistent with the other importance moderators, research shows significant consistency between implicit and explicit weight identity for women but not for men (Grover, Keel, & Mitchell, 2003).

Another indicator of representation strength is the extent of *prior experience* with an attitude object (e.g., Fazio & Zanna, 1978). In two studies on condom use (Marsh, Johnson, & Scott-Sheldon, 2001), the relationship between implicitly and explicitly assessed attitudes towards condoms was stronger for people who reported having more prior experience with condoms.

A final strength-related variable, *certainty*, has been investigated by Lambert, Payne, Ramsey, and Shaffer (2005) in the stereotype domain. Specifically, stereotype certainty was operationalised as the degree of perceived outgroup homogeneity. Prior research has found that people who perceive a given outgroup as homogeneous on various traits report their attitudes towards this group more quickly and place greater subjective certainty in their attitudes towards the group and group members (Lambert, Barton, Lickel, & Wells, 1998) than people who perceive the group to be heterogeneous. In two studies, implicit attitudes towards Blacks as assessed via a lexical decision task were more strongly related to an explicit impression formation task of a Black target person for those participants who perceived the outgroup as homogeneous compared to those who perceived heterogeneity.

Dimensionality

Nosek (2005) also proposed *dimensionality* as a moderator of implicit – explicit attitude consistency. Judd and Kulik (1980) showed that some

evaluations may be conceived of as a bipolar continuum—being "in favour" of one position (e.g., pro-choice) implies being "against" a different position (i.e., pro-life)—and hypothesised that bipolar structure facilitates encoding, retention, and retrieval of the representation. This suggests that bipolarity may simplify representational structure so that evaluations may be activated more consistently across time and contexts. Conversely, the lack of simple structure for unipolar representations may result in a less stable representation. This bears some relationship with Hsee and colleagues' notion of evaluability, in which the difficulty of evaluating attributes of different concepts varies as a function of whether they are being judged comparatively or separately (Hsee, Loewenstein, Blount, & Bazerman, 1999). Consistent with this hypothesis, Nosek (2005) observed that attitudes fitting a bipolar structure yielded stronger implicit–explicit consistency than attitudes with a more unipolar structure across a wide variety of domains. This study used the IAT as its implicit measure which, because of its comparative format, may be especially suited for observing such effects of dimensionality.

Distinctiveness

Individual differences in implicit–explicit consistency may also be a function of *interpersonal comparison*, highlighting the role of social comparison processes that often influence people's assessments of their thoughts and feelings (Festinger, 1954). Focused on attitudes, Nosek (2005) defined distinctiveness as the magnitude of the discrepancy between one's own attitudes and the perceived norm. Unlike representational strength, attitudes can be distinctive because they are stronger or weaker than perceived norms.

Nosek (2005) suggested that distinctive attitudes should be more self-definitional, even if they are weakly held, so self-reports of relative standing compared to other people may be enhanced when one's own attitude contrasts with others. Supporting this hypothesis, greater distinctiveness was associated with higher implicit–explicit consistency, and this relationship held even after partialling out variation due to representational strength. Distinctiveness as a moderator of implicit–explicit consistency suggests that representations can be identified in relation to a social context, not just on the basis of internal experience.

Awareness

The translation process in our model may also depend on the degree to which people are able to form an accurate propositional representation of their underlying associative representation (e.g., Hofmann et al., 2005b). Most theories of conscious awareness do not suggest that awareness

involves direct access or oversight of brain function. Without delving deeper into debates on consciousness, it is sufficient to note that most theories indicate that awareness reflects mental experience but does not provide a view of mental operations or representational structures themselves (e.g., Nisbett & Wilson, 1977). As such, it does not make much sense to wonder if people are *directly* aware of their associative representations. As Anthony Greenwald noted about locating and identifying his associations, "I would not know where to look" (A. G. Greenwald, personal communication, June 2004). Rather, activation of associative representations may produce mental experiences or observable actions that assist in drawing inferences about the nature of the underlying representations.

There are two senses of awareness that might be relevant for understanding implicit–explicit consistency. For one, individuals may have experiential awareness of the feelings or outcomes of implicit process, even if they have no knowledge of how the results were generated (Monteith, Devine, & Zuwerink, 1993; Nisbett & Wilson, 1977; Strack & Deutsch, 2004; e.g., "When I see a Black person, I feel uneasy"). Alternatively, individuals could generate self-inferences about their implicit representations via observation of their own behaviour (Bem, 1972), even in the absence of mental experiences stemming from the associative representations. For example, during an interaction with a black partner, a white person may observe her own bodily avoidance reactions caused by a negative implicit attitude, and integrate this self-observation into her propositional attitude representation.

In the first sense of awareness, the individual could have privileged access to mental experiences that are influenced by associative representations. In the second, the individual could gain knowledge about internal mental representations in the same way that he or she learns about other people, by watching behaviour. This is an underexamined area of research, but there are a few investigations, especially in individual differences research that have some relevance.

Private self-consciousness. Private self-consciousness is an individual difference assessment of the extent to which individuals pay attention to their bodily states and experiences (Fenigstein, Scheier, & Buss, 1975). In the attitude–behaviour consistency literature, persons high in private self-consciousness tend to show greater consistency between their attitude reports and observed behaviour (Gibbons, 1983; Pryor, Gibbons, Wicklund, Fazio, & Hood, 1977). Individuals high in private self-consciousness might therefore demonstrate stronger implicit–explicit consistency because their increased attention to internal experiences would increase the likelihood of generating explicit self-inferences that correspond with implicit representations (via a self-perception explanation), or their explicit responses might be based more on their internal experiences than on subsequent propositional

thinking (via attention to mental experiences that are shaped by associative representations). While plausible, assessments of private self-consciousness moderated implicit – explicit consistency in just one of three studies in which it was examined (Gschwendner, Hofmann, & Schmitt, 2004), whereas effects were close to zero in the other two studies (Hofmann et al., 2005b).

Mindfulness. Mindfulness is conceptually related to private self-consciousness and can be defined as "an enhanced attention to and awareness of current experience or present reality...which may be reflected in a more regular or sustained consciousness of ongoing events and experiences" (Brown & Ryan, 2003, p. 822). Consistent with the self-consciousness hypothesis, participants higher in mindfulness showed stronger implicit – explicit consistency in implicitly and explicitly assessed affective states than participants lower in mindfulness.

In their meta-analysis of 53 different content domains, Hofmann et al. (2005a) collected raters' judgements of how strongly people on average are likely to introspect about what their attitude or personality characteristic is like for each domain. However, the expected positive relationship between introspection and implicit – explicit consistency across domains did not emerge. Rather, the nonsignificant regression weight pointed in the opposite direction, suggesting that, if anything, higher degrees of introspection tend to weaken consistency at the level of content domains. It is possible that raters did not generate accurate theories about average differences in self-awareness efforts across content domains, making *awareness* moderation with individual difference and experimental manipulations an attractive avenue for future research.

Summary

The moderator variables reviewed in this section are assumed to directly influence the interrelation or "communication" among implicit and explicit representations. The available research on representational strength-related variables points in the same direction: Attitudes and other personality characteristics that are strong, important, certain, bipolar, and often experienced yield higher consistency than attitudes that are weak, unimportant, uncertain, unipolar, and rarely experienced. Thus, the pattern of findings supports the notion that consistency is a function of structural features of the representation, such as the strength of its embodiment in memory. Furthermore, dimensionality and perceived distinctness from the norm have been shown to moderate consistency over and above representational strength (Nosek, 2005), pointing to the importance of structural features of representations and of interpersonal comparison processes as a source of consistency, respectively.

Less work is available to make strong inferences about whether intro-spection increases (Brown & Ryan, 2003; Gschwendner et al., 2004) or decreases consistency (Hofmann et al., 2005a). One speculative explanation is that introspective efforts may trigger a multitude of processes and that context factors may determine which process gains more weight in which case (cf. Hixon & Swann, 1993). On the one hand, thinking about *what* exactly their attitudes or personality characteristics are like may lead people to form a more accurate propositional representation of their underlying association-based attitudes. On the other hand, introspective efforts may give rise to a host of additional thoughts about *why* one holds a particular attitude, some of which may be supportive and some of which may question the initial viewpoint (Wilson, Dunn, Kraft, & Lisle, 1989), thus leading to reduced consistency. Future research will have to disentangle the various mechanisms instigated by peoples' active efforts to get to know themselves. Also, the significance of self-perception processes on people's inferences about underlying associative representations should be investigated more directly.

ADDITIONAL INFORMATION INTEGRATION

An important property of the reflective system is the ability to transform, generate, and gather information for reasoning. Associative representations about the target concept may be just one source of information on which propositions are based. Other sources include autobiographical memory and general knowledge structures about the world and the self (Gawronski & Bodenhausen, 2004; Koole et al., 2001) or propositions generated on the spot (e.g., Converse, 1970; e.g., Wilson & Hodges, 1992). For example, a person's explicit representation of a Cadillac may be based on more than just the implicit evaluation of the car, incorporating also judgements derived from information about price, fuel consumption, and maintenance expenses, the more the person deliberates about the car. Thus, reflective integration of additional information should decrease the strength of the relationship between the associative structure and its corresponding propositional representation.

Spontaneity versus deliberation

If individuals are unable or unmotivated to search for and integrate additional information when making an explicit judgement, then the explicit representation should be more consistent with the implicit one than after substantial deliberation (Fazio, 1990). Two common ways of limiting additional information integration are to introduce time pressure requiring a rapid response, and to introduce cognitive load that disrupts the normal course of information processing.

Ranganath, Smith, and Nosek (2005) measured self-reported "gut feelings" and "actual feelings" toward gay and straight people. Two-factor structural equation models with "gut" and implicitly measured evaluations (IAT; GNAT; SPF-4, Nosek & Vianello, 2005) as a single factor and "actual feelings" as a second factor fitted the data better than models loading the two self-report measures on a single factor and the implicit measures on a second factor. In a second study, the authors measured self-reported attitudes in two conditions—under time pressure (fast) and with unlimited time to respond (slow). The better-fitting model had fast responses and the IAT loading on one factor and slow responses on a second factor. Thus, despite the fact that "fast" and "slow" responses were both direct measures of attitudes, the responses under time pressure corresponded more with indirectly assessed evaluations via the IAT than with the slower direct assessments.

Investigating implicit and explicit self-esteem, Koole et al. (2001) found similar evidence for explicit judgements made spontaneously or under cognitive load. In two studies, implicit self-esteem as assessed with the name-letter effect (Nuttin, 1985) and explicit self-esteem as assessed via the Rosenberg (1965) self-esteem scale correlated more strongly for faster than slower respondents (Study 3) and for participants under high versus minimal cognitive load when completing the explicit measure (Study 4). Lack of deliberation or available cognitive resources minimised self-reflection. So, the authors argued, when cognitively taxed or speedy, explicit self-esteem is based on chronically accessible implicit self-evaluations whereas with increasing deliberation, self-reports are influenced by additional sources of information from autobiographic memory.[2]

Likewise, in their meta-analysis, Hofmann and colleagues (2005a) showed that domains in which explicit responses were judged as more likely to be based on gut-level reactions yielded stronger implicit–explicit correspondence than domains judged as likely to elicit deliberation prior to making a response. These demonstrations show that explicit self-reports reflect underlying associations to a greater extent when people do not retrieve and deliberate about additional information from memory in the course of making an explicit judgement.

The flipside of spontaneous judgements are those made with deliberation. A popular individual difference variable indicating tendencies to deliberate is the need for cognition (Cacioppo & Petty, 1982). If need for

---

[2]An alternative interpretation in terms of motivational influences is that spontaneity hindered participants from socially adjusting their responses on the explicit measure, leading to higher implicit–explicit consistency. However, this assumption is inconsistent with the finding that fast responders endorsed *higher* explicit levels of self-esteem than slow responders. If social adjustment were the primary source of influence, then the reversed pattern of means would have been expected.

cognition elicits additional information integration in explicit judgement, then it should be associated with weaker implicit–explicit consistency. Consistent with this prediction, Florack, Scarabis, and Bless (2001b) found that participants high in need for cognition showed less consistency between an IAT measuring attitudes towards Turks and a corresponding explicit attitude measure as well as a Turkish target impression formation task.

Further moderators may be derived from considerations of cognitive styles and mood states. One prominent conception, regulatory-focus theory (Higgins, 1998), distinguishes between a *promotion focus*, in which people focus on accomplishments and advancements, using approach strategies in order to attain a desired goal or state, and a *prevention focus*, in which people focus on protection, safety, and responsibility, using avoidance strategies in order to avoid potential negative outcomes. Recent evidence on consumer attitudes suggests that people in a promotion focus tend to base their explicit preferences towards consumer products to a stronger extent on implicit evaluations than people in a prevention focus (Florack, Scarabis, & Gosejohann, 2005). The authors suggested that people in a promotion focus "trust" their immediate responses, whereas those in a prevention focus do not consider their automatically activated attitudes a valid basis for explicit judgements, and therefore rely more strongly on other information than just their automatic preferences. Similar effects may be derived from theories of mood and information processing (e.g., Clore, Schwarz, & Conway, 1994). People in a positive mood may be more prone to base their explicit judgements on their spontaneously activated associations, whereas people in a negative mood may engage in more systematic information processing and integration of information from various sources other than their associative representations.

Finally, a known-groups comparison of gender differences in reliance on spontaneous feelings provides some support for spontaneity as a moderator of implicit–explicit consistency. Pelham and colleagues (Pelham, Koole, Hardin, Hetts, Seah, & DeHart, 2005) argued that women are socialised to base their judgements more on intuitions and gut-level reactions than men do (cf. Pacini & Epstein, 1999). Across six independent samples from several cultures, the authors found that implicit self-esteem as measured via the name-letter task (Nuttin, 1985) and a word-fragment completion paradigm (Implicit self-evaluation scale; Hetts et al., 1999) correlated more strongly with explicit self-esteem (Rosenberg, 1965) for women than for men.

Cognitive dissonance

According to dissonance theory (Festinger, 1957), people are motivated to reduce an aversive state resulting from logically inconsistent propositions.

As consistency between two propositions depends on the truth values assigned to them, dissonance can only be resolved by changing the truth value of one proposition, or by finding an additional proposition that resolves the inconsistency (e.g., Kruglanski, 1989). Thus, cognitive dissonance should operate only in the reflective system on the level of propositions that have a truth value assigned to them (see above), but not in the impulsive system on the level of associative representations that are independent of truth values (Gawronski, Strack, & Bodenhausen, in press). For example, when forced to behave in a counter-attitudinal way, people may be motivated to search for additional propositions or to change the truth value of one proposition, thereby altering the content of their propositional (explicit) attitude representation, while their associative (implicit) attitude stays the same. Consequently, implicit – explicit correspondence should be reduced.

These assumptions were tested in two studies by Gawronski and Strack (2004) in which participants had to write a counter-attitudinal essay closely following the original Festinger and Carlsmith (1959) paradigm. In the "forced position" condition, participants were ostensibly randomly assigned to generate arguments against their own attitude. In the "induced compliance" condition, however, they were led to believe that counter-attitudinal essays were scarce and that the researchers would be pleased about such essays, but that everybody was free to choose whatever position they deemed best. As expected, subsequently assessed implicit and explicit attitudes were correlated more strongly when a strong justification (i.e., forced assignment) for writing the essay was provided. However, when no strong justification was given, the relationship between implicit and explicit measures was weaker. This is consistent with the argument that dissonance was resolved by changing the propositionally represented explicit attitude, reducing reliance on the activated associations and leading to weaker implicit – explicit correspondence.

Summary

The evidence on additional information integration consistently supports the notion that factors that foster the spontaneity with which an explicit response is given heighten consistency whereas deliberation reduces it. The more an explicit representation includes additional propositions stemming from sources beyond the translation of an associative representation, the more likely it is that implicit and explicit representations will differ. Further support is observed in research on cognitive dissonance. Implicit – explicit consistency is weaker when dissonance is resolved by questioning the validity of one's explicit attitude (Gawronski & Strack, 2004).

EXPLICIT ASSESSMENT

The interdependence of theory and measurement is illustrated in the scientific truism that hypothetical constructs cannot be investigated without measurement and even valid indicators are never perfectly isomorphic with the intended construct. Explicit indicators are influenced by a variety of factors besides the explicit representation. If not identified, the unintended influences on explicit assessment could lead to misspecification of implicit – explicit consistency. As a consequence, implicit – explicit consistency will be understood properly only if the factors influencing assessment are also included in the model.

Adjustment

As stated in the introduction to this article, explicit measures—in contrast to implicit measures—are assumed to be more susceptible to *adjustment processes* by which the explicitly given response deviates from the underlying explicit representation. We define adjustment as a motivated self-regulatory process by which the content of an explicit representation is edited before being reported in order to fit an internalised standard or external norm. As a consequence of this editing process, adjustment should reduce the consistency between the explicit and implicit indicator. Adjustment should only affect the explicit response as it hinges on the availability of control resources. This conjecture is consistent with research showing (a) that implicit measures are more difficult to fake than explicit measures (e.g., Asendorpf et al., 2002; Egloff & Schmukle, 2002; Kim, 2003) and (b) that greater correspondence results among implicit indicators and spontaneous behaviour on the one hand and explicit indicators and controllable behaviour on the other hand (e.g., Asendorpf et al., 2002; Dovidio, Kawakami, & Gaertner, 2002; Dovidio et al., 1997; Fazio et al., 1995; Perugini, 2005).

One pervasive motivational source of adjustment is self-presentation, social desirability, or impression management (Edwards, 1967; Tedeschi & Riess, 1981). Self-presentation has been examined in comparison of content domains that are more or less likely to elicit socially desirable responses, in the comparison of individuals who are more or less susceptible to socially desirable responding in general, and as a situational factor in which some situations elicit more adjustment than others.

Comparisons of content domains.   Swanson, Rudman, and Greenwald (2001) reported higher implicit – explicit consistency between two IAT measures and two explicit indicators (feeling thermometer, semantic differential) for a non-stigmatised domain (vegetarianism) as compared with a stigmatised one (smoking). In a similar vein, Franco and Maass

(1999) showed that implicit and explicit measures of prejudice correlated positively only for a target outgroup (Islamic Fundamentalists) that was not normatively protected against discrimination (as indicated by a pretest), but were not positively associated for a normatively protected outgroup (Jews). Also, Nosek and Banaji (2002) showed that raters' judgements of the social desirability demands of 15 different content domains predicted the strength of implicit – explicit consistency across those domains. Higher social desirability demands were associated with weaker implicit – explicit consistency. Finally, in a meta-analysis that should have sealed the case for social desirability as a moderator of implicit – explicit consistency, Hofmann et al. (2005a) could not replicate the moderation effect, even though on a descriptive level higher implicit – explicit correlations were obtained for consumer attitudes and correlations were lower for the more sensitive stereotype and self-esteem domains. This last finding introduces some caution about the near-consensus conclusion that differences in social desirability concerns across content domains are a determinant of implicit – explicit consistency.

Interindividual differences within a specific content domain. Instead of relying on the average level of social desirability for a given domain, Nosek (2005) assessed individually rated self-presentation concerns for 57 content domains. Whether the domain was social, political, or consumer attitudes, individuals who felt stronger self-presentation concerns about a given topic showed weaker implicit – explicit consistency than individuals who felt weaker self-presentation concerns. The magnitude of this effect was surprisingly small. Together with the meta-analysis by Hofmann et al. (2005a), this suggests that the field's emphasis on self-presentation as a (the) moderator of implicit – explicit consistency may be too strong.

Even so, in the domain of prejudice, numerous investigations have found evidence for implicit – explicit consistency moderation using Dunton and Fazio's (1997) motivation to control prejudice scale as a domain-specific adjustment construct (Banse & Gawronski, 2003; Banse, Seise, & Zerbes, 2001; Dunton & Fazio, 1997; Fazio et al., 1995; Gawronski, Geschke, & Banse, 2003; Hofmann et al., 2005b; Olson & Fazio, 2004b; Payne et al., 2005; but see Fazio & Hilden, 2001).

Plant and Devine (1998) further distinguish two types of stereotype control motivation: *internal and external motivation to control prejudice.* People high in internal motivation are assumed to control for prejudice due to strongly internalised personal standards while those high in external motivation are externally constrained to comply with society's egalitarian norms.

Hofmann et al. (2005b) found that only internal motivation negatively moderated implicit – explicit relations concerning evaluations of Turkish

people in a German sample. Likewise, Payne et al. (2005) reported that only internal motivation moderated the consistency between an implicit affect misattribution measure and self-reported feelings towards Blacks. These findings suggest that adjustment to internal standards may be a stronger moderator than adjustment to external norms. Finally, Ziegert and Hanges (2005) employed only the external subscale and found it moderated the relationship between a race IAT and the attitudes towards blacks scale (Brigham, 1993), but not the more subtle modern racism scale (McConahay, 1986). This latter finding suggests that external motivation may be a significant moderator only for blatant explicit measures that clearly elicit social desirability concerns and not for explicit measures including more subtle but socially acceptable items.

General interindividual differences.   Social desirability also varies as a relatively stable inter-individual difference trait regardless of content domain. Of three studies examining social desirability trait effects on implicit – explicit consistency, only one showed a reliable influence (Banse & Fischer, 2002), while two others found no effect (Egloff & Schmukle, 2003; Hofmann et al., 2005b).

Other personality traits may be related to the extent to which people adjust their explicit responses to external norms, such as *public self-consciousness* (Buss, 1980; Fenigstein et al., 1975) or *self-monitoring* (Snyder, 1979). Public self-consciousness indicates the tendency to direct attention to those aspects of one's behaviour that are observable by others. People high in public self-consciousness should tune their behaviour more strongly to social expectations (Turner, 1978). Hofmann et al. (2005b) found that public self-consciousness moderated implicit – explicit consistency in one study on stereotyping, but not in an attempted replication.

Self-monitoring has often been investigated as a moderator in past research on attitude – behaviour consistency (Snyder & Ickes, 1985). According to Snyder (1979), some people tend to closely monitor social cues from the environment in order to adjust their behaviour (high self-monitors) while others tend to rely mainly on internal standards, attitudes, value beliefs, and feelings (low self-monitors). Consequently, high rather than low self-monitors should show a stronger sensitivity to situational and interpersonal cues to social appropriateness when making a verbal self-report (e.g., Snyder & Cantor, 1980), resulting in reduced implicit – explicit consistency.[3] However, this pattern was not borne out in the one known

---

[3]Note that self-monitoring is a construct that bears both on adjustment (low self-monitors are less prone to adjustment than high self-monitors) and awareness of internal states (low self-monitors are more strongly aware of inner states). Hence, one could even expect a joint effect of both factors such that consistency is particularly strong for low self-monitors.

application to implicit–explicit consistency (Hofmann et al., 2005b). Taken together, support for trait predictors of social adjustment such as social desirability, public self-consciousness, and self-monitoring varies from weakly supportive to inconclusive.

Situational manipulations.   A recent study (Nier, 2005) adopted the bogus pipeline paradigm (Jones & Sigall, 1971) and suggested to participants in one condition that the IAT was a valid, non-fakeable procedure for the assessment of prejudice. Implicit–explicit consistency was significantly positive for this condition, and was not in a "no information" control condition and in a condition that introduced the IAT as a poor measure. This suggests that participants in the IAT-valid condition reduced deliberate adjustment of self-report responses for fear of being caught "lying". In another study, Nosek (2002) had participants report attitudes in public (interview) and private (alone, on a computer). Attitudes reported in private mediated relations between IAT effects and attitudes reported in public.

Finally, Ziegert and Hanges (2005) manipulated the legitimacy of explicit racial bias against black job applicants in an experimental hiring context. Some participants received a memo from the fictitious company's president stating that it was essential to put a white person in the position for reasons related to the company's team spirit; other participants did not receive a memo. As expected, implicit racial attitudes, as measured by an IAT, were predictive of explicit racial bias in a hiring decision when expressing bias was legitimised by the company president, but not in the control condition.

Outgroup threat

Feeling threatened by an outgroup may disinhibit or bypass the social adjustment process by creating an emotionally uneasy state in which the self-regulation of one's overt response fails (cf. Baumeister & Heatherton, 1996). In line with these assumptions, Florack, Scarabis, and Bless (2001a) found that implicit attitudes towards Turks as assessed with a race-IAT correlated more strongly with a judgement of guilt towards a Turkish target person for German participants who were induced to feel threatened by the outgroup of Turks. However, implicit attitudes and guilt judgements were unrelated for controls who were not threatened. Although it was not interpreted this way by the authors, comparable results may be suggested in a study by Rudman and Lee (2002, Study 2) in which outgroup threat may have been created experimentally by presenting white participants with violent rap-music prior to explicit report. Participants' race IAT scores in the "rap-music" condition correlated significantly with explicit hostility and sexism ratings of a black target person, but not in a control group exposed to popular music. Taken together, these results suggests that outgroup

threat may increase the likelihood that negative implicit concerns about the respective outgroup are explicitly expressed.

Method-specific variance

In addition to social adjustment effects, a variety of other *method-specific factors* may affect the validity of the explicit indicator, such as acquiescence, extreme response sets, or question context effects (cf. Schwarz & Sudman, 1992; Sudman, Bradburn, & Schwarz, 1996). Although there is little research directly examining these effects in the context of implicit–explicit relations, the control or elimination of such method-specific sources should result in increased consistency.

Summary

A considerable amount of research has addressed the question of whether implicit–explicit consistency varies as a function of the social adjustment motive. Meta-analytical evidence for differences in consistency across topics varying in social desirability is surprisingly weak (Hofmann et al., 2005a). Even so, replicable effects have been observed showing that adjustment operates in individual differences (Fazio et al., 1995) and situational factors (Nier, 2005).

Adjustment as a personal moderator can be conceptualised as a rather general trait such as the tendency to give socially desirable answers, or as a domain-specific trait such as the motivation to control prejudiced reactions. While evidence is mixed for general moderators such as social desirability and public self-consciousness, clear support has been obtained for the more specific motivation to control prejudiced reactions in the stereotypes and prejudice domain.

A second useful distinction can be drawn between internal and external sources of the motivation to adjust (Plant & Devine, 1998), with internal motivation generally yielding stronger negative moderator effects on consistency. This distinction raises the interesting theoretical problem of whether adjustment to internal and adjustment to external norms both operate at the same stage or at different stages in the model. Asked another way, can internally motivated or "honest" forms of adjustment be parsed from externally motivated or "dishonest" adjustment? One possible answer to this question is that internal motivation to adjust may even affect the contents of the explicit representation themselves because internal standards such as egalitarian values may constitute an "additional" source of information that shapes the explicit representation (representation-relevant adjustment), while external motivation may lead to an adjustment of the explicit indicator away from the underlying explicit representation (representation-irrelevant adjustment). For the sake of parsimony, we

grouped all adjustment moderators into the same section in this review but future theoretical and empirical work may define conditions to better clarify this distinction between representation-relevant and -irrelevant forms of adjustment.

Making the matter of exactly where adjustment operates even more complicated, one may argue that adjustment need not necessarily be a controlled process that exclusively operates on active explicit representations, but may even become automatised (Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). Such automatic forms of adjustment may inhibit the translation process prior to an implicit representation interacting with a propositional representation in the first place (cf. Wilson et al., 2000). As with the above discussion about distinguishing representation-relevant and -irrelevant adjustment, automatic adjustment processes introduce interesting complications that future models will have to address to clarify the nature of adjustment processes as moderators of consistency.

IMPLICIT ASSESSMENT

Implicit measures are assumed to tap associative representations by employing techniques that decrease deliberation, introspection, awareness, or controllability of responses. These techniques have a variety of properties that impact the validity and context-sensitivity of the measures. As a consequence, the degree of implicit – explicit consistency also varies.

Situational malleability

The implicit pathway may be influenced by the personal and situational preactivation of associations (Blair, 2002). Accordingly, implicit representations reflect a combination of state and trait variance (Schmukle & Egloff, 2004). State variance of implicit measures emerges whenever different measurement contexts pre-activate implicit associations differentially. Furthermore, implicit – explicit consistency depends on the degree to which implicit and explicit measures are sensitive to the same (increased consistency) versus different (decreased consistency) context factors.

Initial evidence for the influence of the situational context on implicit – explicit consistency stems from a series of studies by Wittenbrink, Judd, and Park (2001). In one study, participants saw either a video clip of Black Americans in a gang-related setting or a video clip of Black Americans at an outdoor barbecue, before completing an implicit subliminal priming measure of stereotyping and several explicit measures. Positive implicit – explicit correlations emerged only for participants in the gang-related video condition. However, because explicit measures were also administered after the video manipulation, this study does not allow us to rule out an influence

of the video manipulation on the explicit pathway by rendering the expression of prejudice more acceptable. This alternative explanation was ruled out in a second experiment in which the manipulation of the context was integrated into a sequential priming paradigm such that the presentation of a Black or White face prime immediately followed the presentation of a context prime (a dilapidated street corner vs the inside of a Baptist church) before an evaluative response had to be given. Again, significant implicit – explicit correlations were obtained only for the negative context. One possible interpretation of these findings is that the negative context facilitated the assessment of *negative* stereotypes on the implicit measure which corresponded more closely to the explicit measure designed to assess primarily the negative component of the Black stereotype, while the positive context led to less correspondence between both measures.[4]

Method-specific variability

Implicit measures, like all psychological measures, are not perfect indicators of their intended constructs. Rather, their validity is attenuated by a variety of method factors. If these method factors are specific, i.e., have no effect on explicit measures, the relationship between indicators will underestimate the implicit – explicit consistency between representations. Contrarily, if implicit and explicit measures have method factors in common, consistency between implicit and explicit indicators could overestimate the relationship between the representations. Given the profoundly different assessment procedures between most implicit and explicit measures, the former is a more likely scenario than the latter.

Among implicit measures, the IAT (Greenwald et al., 1998) has undergone the most rigorous testing in the identification of method-specific variability. This extensive attention has benefited the IAT because subsequent procedural or analytic innovations have reduced the impact of some of the construct-irrelevant contaminating influences (Greenwald et al., 2003). Other implicit measures will mature and improve as they receive similar scrutiny. Because of the extensive data available, this section will emphasise the IAT.

Method-specific influences on the IAT. Nosek et al. (in press) reviewed methodological influences on the IAT that are conceptually unrelated to the assessment of association strengths. The presentation order of the critical

---

[4]As an anonymous reviewer noted, another possibility to account for the facilitative effect of negative context could be that thinking of gang members made the expression of prejudice more acceptable. Thus, the results by Wittenbrink et al. (2001) may be interpreted as also affecting the explicit assessment of prejudice via the disinhibition of social desirability concerns through changing the context.

response tasks are an extraneous influence on IAT effects (Greenwald & Nosek, 2001) that may alter the distribution of individual scores when presentation order is counterbalanced across participants (e.g., Gawronski, 2002). Nosek, Greenwald, and Banaji (2005) suggested a procedural correction that minimises the impact of presentation order.

Cognitive fluency, or average response latency, is another extraneous influence of the IAT that disrupts inter-individual comparisons (Greenwald et al., 2003; McFarland & Crouch, 2002) as it varies with age and fluid intelligence (Kray & Lindenberger, 2000). The influence of cognitive fluency (Cai, Sriram, Greenwald, & McFarland, 2004) and a related factor called task-switching ability (Back, Schmukle, & Egloff, 2005; Mierke & Klauer, 2003)[5] is reduced with the application of improved scoring procedures suggested by Greenwald and colleagues (2003).

Another extraneous influence is prior experience with the IAT. Effect magnitudes with the task tend to decline with experience (Nosek et al., in press). Hence, implicit – explicit consistency may be stronger for samples in which practice level is homogeneous among participants as opposed to samples with largely varying levels of practice. Also, the impact of this extraneous factor can be reduced when using scoring procedures recommended by Greenwald et al. (2003).

Some authors have suggested that the IAT (and possibly also other implicit measures) are confounded by *extra-personal associations*—information that exists in memory but is not construct-relevant in that it is not an active contributor to thought and action related to the target construct (Karpinski & Hilton, 2001; Olson & Fazio, 2004a). The most cited potential extra-personal influence is cultural knowledge on IATs designed to measure attitudes (see also Arkes & Tetlock, 2004). For example, an IAT measuring racial attitudes may be confounded by people's cultural knowledge that Blacks are devalued in general. In an attempt to "personalise" the IAT, Olson and Fazio (2004a, Studies 3 and 4) obtained higher implicit – explicit correlations in comparison to a traditional IAT when changing the labels of the attribute categories from "positive" vs "negative" to "I like" vs "I don't like", respectively.

In contrast, further empirical data do not support the conclusion that cultural knowledge influences the IAT significantly: In seven studies with large samples across more than 50 topical domains, Nosek and Hansen (2005) observed reliable relations between self-reported attitudes and IAT effects, but no independent relation between cultural knowledge measures and the IAT, even among the 11 domains that elicited more

---

[5]According to these authors, task-switching ability can be defined as the speed with which people are able to switch between the two different task sets for classifying targets and attributes in the IAT, respectively.

inter-individual variability in cultural knowledge than self-reported attitudes. Further, this pattern of results was identical for the traditional version of the IAT and a "personalised" version as used by Olson and Fazio (2004a). As such, whether extra-personal associations influence the IAT and, if they do, whether they should be considered a confounding influence, is ambiguous.

Faking. The ease of faking responses on explicit measures played a role in fostering development of implicit measures that are more resistant to intentional alteration. An attractive quality of implicit measures is that they assess the target construct indirectly by limiting awareness of the topic of assessment or the controllability of the response. Despite this, implicit measures are not impervious to alteration and depend on the willing participation of the respondent for reliable assessment. Reviewing the data on fakeability of the IAT, Nosek and colleagues (in press, p. 14) concluded that "Investigations of IAT fakeability across multiple content domains including shyness, extraversion, moral identity, attitudes toward flowers versus insects, attitudes toward sexual orientation, and attitudes toward racial groups collectively suggest that (a) the IAT is much less fakeable than self-report, (b) the IAT is not very fakeable when subjects are given only abstract instructions to do so (e.g., 'try not to appear shy') and, (c) two factors, experience with the IAT and explicit instructions about how to control IAT scores, increase fakeability" (e.g., Asendorpf et al., 2002; Banse et al., 2001; Egloff & Schmukle, 2002; Fiedler & Bluemke, in press; Kim, 2003; Steffens, 2004). Other implicit measures such as subliminal priming paradigms may be even less susceptible to faking because participants are often unaware of what is being measured and how.

Thus, when faking influences *both* implicit and explicit responses, implicit–explicit consistency may be overestimated. Kim (2003), for example, showed that implicitly and explicitly assessed attitudes towards Blacks corresponded more strongly when participants were instructed on how to fake the IAT compared to controls.

Reliability

One pervasive but often neglected psychometric influence is error of measurement. Both implicit and explicit indicators will inevitably contain measurement error, thus having less than perfect loadings on their underlying representations. However, this issue is more dramatic for implicit assessment because many of the measures suffer from weak reliability (e.g., Bosson, Swann, & Pennebaker, 2000; De Houwer, 2003). Sequential priming, one of the more popular implicit measures, often shows

near zero internal consistency (e.g., Olson & Fazio, 2003). These minimal reliabilities have serious consequences for the assessment of consistency between implicit and explicit representations because they could mislead one to interpret a strong relationship as a weak one. Not accounting for reliability is a likely factor for many null implicit – explicit relations when the measures are unreliable (Bosson et al., 2000). Additionally, differences in reliability from study to study may introduce unsystematic variation between research findings that may be misinterpreted as indicative of moderator effects (Hunter & Schmidt, 1990). Moreover, the consequences of unreliability for moderator testing should not be underestimated. For example, in moderated regression analysis, the power of the interaction term is strongly reduced by unreliable predictors (Aiken & West, 1991).

One of the attractive qualities of the IAT is that it shows substantially better internal consistency than other implicit measures, usually ranging between .7 and .9 (Hofmann et al., 2005a; Nosek et al., in press). The stronger internal consistency leads to greater power for observing relations with other variables. Also, researchers have shown that disattenuating unreliability through structural equation modelling reveals more robust implicit – explicit consistency than is observed with raw score comparisons (Cunningham, Preacher, & Banaji, 2001; Nosek & Smyth, 2005). For example, Nosek and Smyth (2005) reanalysed a large dataset of implicit – explicit correlations (Nosek, 2005). The median correlation of .37 across 57 domains was .48 after reanalysis with structural equation modelling—a 68% increase in shared variance.

Summary

The studies reviewed on the implicit assessment process provide clear evidence for the notion that implicitly assessed representations are subject to a variety of situational and extraneous influences (Blair, 2002). These variables include the situational activation of the relevant context which renders the implicit association more accessible during measurement (Wittenbrink et al., 2001) and a variety of method-specific influences such as cognitive skill confounds and faking. With the continuing improvement of implicit (and explicit) measures, the distracting influence of extraneous factors will be reduced and will improve identification of the principles by which implicit and explicit representations influence each other.

DESIGN FACTORS

Design factors are not psychological influences on implicit – explicit consistency. Rather this category encompasses components of the research

process that can influence the presence or absence of implicit–explicit relations that are unrelated to the underlying constructs of interest. These include biases in sampling and measurement correspondence.

Sampling biases

Selective sampling can often lead to reduced variability in the measured indicators that results in an underestimation of the "true" relationship between variables (e.g., Schmitt, 1990). In the present context, variance restriction is especially likely for explicit measurement of attitudes or traits in socially desirable domains (Greenwald & Nosek, 2001). For example, assessments of explicit racial prejudice in university samples tend to elicit reduced variability as most university undergraduates are unwilling to endorse non-egalitarian values (Flammer, 2001; Schuman, Steeh, Bobo, & Krysan, 1997). When more varied samples are examined, stronger correlations with regard to implicit measures can emerge. For example, Greenwald and colleagues (2003) reported a correlation of .36 between implicit and explicit racial attitudes in a diverse sample ($N = 10,475$) of drop-in visitors to a demonstration website.

Further, because the psychological qualities of content domains differ widely on factors that may contribute to implicit–explicit consistency (Nosek, 2005), it is risky to generalise from a selective set of domains. For example, across three content domains, Karpinski and Hilton (2001) observed little implicit–explicit consistency and concluded that the IAT was unrelated to self-reported attitudes. But, in broader sampling of content domains, substantial correlations between the IAT and self-report are often observed (Nosek, 2005).

Order of implicit–explicit measurement

Does the order in which implicit and explicit measures are administered affect the magnitude of the correlation between the two? On the one hand, one may argue that demanding an explicit judgement from participants first may provide participants with a relevant context and thus selectively activate associations that are more compatible with explicit responses than with the reversed order, leading to increased consistency between measures. Equally plausible, however, one may hold that activating associative representations first may determine explicit propositions to a greater extent (hence increasing consistency) than when the propositional representation is assessed first, because propositional influences on associations may take more time in order to shape associations (e.g., Smith & DeCoster, 2000). In a first investigation addressing potential order effects, Bosson et al. (2000)

reported higher average consistency between a number of implicit and explicit measures of self-esteem when explicit measures were administered first. In contrast to this finding, however, Nosek et al. (2005) and Hofmann et al. (2005a) found no reliable differences in implicit – explicit correlations as a function of whether the IAT or the explicit measure was administered first, suggesting that order effects may be negligible for consistency.

Measurement correspondence

Perhaps the most influential methodological explanation of attitude – behaviour (in)consistency was offered by Ajzen and Fishbein (1977) who highlighted the *principle of correspondence*. According to their reasoning, consistency is higher when attitude and behaviour measures are comparable in content and specificity. Ajzen and Fishbein (1977) reviewed more than a hundred studies on the attitude – behaviour relationship and concluded that most studies in which attitude and behaviour corresponded to each other in content and specificity yielded correlations above .40, whereas most studies with low correspondence between measures yielded non-significant correlations.

Likewise, implicit – explicit consistency should increase with increasing correspondence among implicit and explicit indicators in their content and specificity. Modern explicit attitude measures, for example, often infer attitudes by aggregating levels of agreement to a variety of propositional statements. For example, items on the Modern Racism Scale (Brigham, 1993) concern beliefs about race-related issues such as support for affirmative action. The richness of these explicit measures may actually be a hindrance to observing strong implicit – explicit consistency because implicit measures are presumed to reflect relatively simple associative relations such as the association of racial groups with positive or negative evaluation, rather than the rich nuance of support or disapproval of race-relevant political issues. So, explicit measures that directly assess evaluations of the target concepts may show greater consistency with implicit measures because of correspondence in specificity. Consistent with this assumption, Hofmann et al. (2005a) reported significantly higher consistency of the IAT with regard to simple adjective ratings, feeling thermometers, and semantic differentials, as compared with more complex questionnaire scales.

Another important component of comparability of measures is whether the measures are comparative or absolute in nature (Nosek & Banaji, 2001). The IAT, for example, is a relative measure contrasting associations between pairs of concepts—e.g., Black – White, male – female, rich – poor (Greenwald et al., 1998; Nosek et al., in press). As such, implicit – explicit

consistency for the IAT should be maximised when the explicit measure also provides a comparative assessment rather than an absolute assessment of just one of the content domains. In support, the meta-analysis by Hofmann et al. (2005a) revealed that correlations were higher for relative as compared with absolute self-report measures.

Moreover, some researchers have characterised implicit measures such as the IAT as measuring affective rather than cognitive evaluations (Marsh et al., 2001; Spence, in press). In line with this assumption, the IAT showed higher consistency with regard to affective rather than cognitive explicit measures (Hofmann et al., 2005a).

Correspondence in content is illustrated by multitrait–multimethod investigations that reveal convergent and discriminant validity for implicit and explicit attitude measures. For example, in a study by Gawronski (2002), implicit preferences for Germans relative to Turks correlated better with explicit preferences for Germans relative to Turks than with explicit preferences for Germans relative to Asians. Likewise, implicit preferences for Germans relative to Asians correlated better with explicit German–Asian preferences than German–Turk preferences. Taking advantage of powerful methods of multitrait–multimethod testing in structural equation modelling, Nosek and Smyth (2005) further showed convergent and discriminant validity for the IAT and explicit attitude measures across seven domains; attitude measures with matched content domains yielded stronger correspondence than measures with mismatched domains.

Summary

Sampling biases in participants and content domains is an empirical limitation that cannot be eliminated just by acknowledging the limitation. Most of the early work in implicit–explicit consistency emphasised their dissociation because the work focused on a single content domain (racial attitudes) and a sample that showed little variability in explicit responses in that domain (university students). More recent work investigating a wider range of people and domains shows that the conclusions of dissociation between implicit and explicit processes were premature.

Just as for attitude–behaviour consistency (e.g., Ajzen & Fishbein, 1977), correspondence of measures is a major research design factor: Consistency appears to be a direct function of implicit and explicit measures' comparability in content and specificity. Still, determining what exactly renders both measures comparable may require not only considerations of more technical measurement issues, but also some consideration for the properties of implicit and explicit *representations*. This becomes evident

from the above findings on affective vs cognitive contents (Hofmann et al., 2005a)—pointing again to the interdependence of theory and measurement that characterises our model.

DISCUSSION

The goal of the present review was to organise the empirical findings on the moderation of implicit – explicit consistency into a working model. Just like attitude – behaviour consistency (e.g., Zanna & Fazio, 1982), implicit – explicit consistency varies as a function of multiple factors. Drawing on the recent distinction between associative and propositional representations (Strack & Deutsch, 2004), we assigned moderator variables to five primary factors (see Figure 1): *translation* between implicit and explicit representations, additional *information integration* for explicit representations, factors influencing *explicit assessment*, factors influencing *implicit assessment*, and features of research *design and sampling*. In our review we have discussed both the dominant issues and lingering questions for future investigations. We end by discussing potential limitations of the reviewed literature and suggesting general strategies for future research.

Limitations

This review is limited by at least three features of the existing literature. First, most research on implicit – explicit consistency exists in research on attitudes, so other areas such as stereotyping and personality are under-represented. Second, despite the growing diversity of implicit measurement methods, most of the available empirical evidence used the IAT. The generality of these findings to other methods will emerge in the next generation of implicit cognition research as more work is done with other methods, and the relations among implicit measures are better understood. Finally, reviews are biased by the domains investigated, and potential biases in what type of work is more likely to be published. The first issue is surely a factor as most of the available empirical work is in domains with high social sensitivity and on topics where little implicit – explicit consistency is expected. The extent of publication biases is unknown. In terms of implicit – explicit consistency, it seems unlikely to be a major factor, as many of the reviewed studies were not primarily concerned with implicit – explicit consistency moderation. More importantly perhaps, findings of association and dissociation between implicit and explicit measures are valued for different purposes, making it less likely that publication biases lead to an emphasis of one over the other.

Research strategies for future investigations

Next to the investigation of theoretically important but yet largely unexplored moderator candidates such as introspection, self-observation, mood states, automatic adjustment, and others, a number of more general research strategies may be transferred from the past work on attitude– behaviour consistency. One first promising approach is to investigate how *functionally equivalent* personal and situational moderators interact with each other. Such an approach is rooted in the concept of modern interactionism, according to which any behaviour is a function of a particular person in a particular situation (Magnusson & Endler, 1977). Compared with a purely personal or situational approach, this strategy allows for conceptual replication and mutual validation while methodological heterogeneity is at its maximum. In addition, a simultaneous test allows the researcher to investigate *whether* personal and situational moderators interact in an additive (i.e., no interaction on the moderator level), synergistic (i.e., person and situation intensify each other as moderators), or compensatory (i.e., differences among persons are compensated by a ''strong'' situation) manner. For example, a recent investigation by Gschwendner et al. (2004) provided the first evidence that inter-individual differences in awareness interact synergistically with a situational manipulation of awareness.

Second, it should be considered how *different* moderators interact with each other. The simultaneous test of different moderators has important benefits and can lead to refinements in theory (e.g., Zuckerman, Koestner, DeBoy, Garcia, Maresca, & Sartoris, 1988), even though it introduces risks of increased complexity and challenges in replicating interactive effects (Cronbach & Snow, 1977). Hofmann et al. (2005b), for example, found that the moderating effect of motivation to control prejudiced reactions was more pronounced for persons high in private self-consciousness and persons holding important attitudes, as opposed to low private self-consciousness or unimportant attitudes respectively, suggesting that adjustment may hinge on successful translation of one's implicit representations. At the same time, Nosek (2005) found no interactive effects among four moderators—self-presentation, attitude strength, dimensionality, and distinctiveness.

Third, during the attitude/trait–behaviour consistency debate, several authors argued for a definition of *consistency as a trait*, with some persons showing generally higher consistency than others (e.g., Baumeister & Tice, 1988; Schmitt, 1990). For implicit–explicit consistency, the degree of *intra-individual* (in)consistency between implicit and explicit representations may have implications for information processing and behaviour. Recently, Jordan and colleagues (Jordan, Spencer, & Zanna, 2005; Jordan, Spencer,

Zanna, Hoshino-Browne, & Correll, 2003) showed that the combination of high explicit self-esteem and low implicit self-esteem is a strong predictor of defensive behaviour such as dissonance reduction, in-group bias, and discrimination. In a similar vein, Briñol, Petty, and Wheeler (2002) demonstrated that people with high discrepancies between implicit and explicit self-conceptions (Studies 1–3) or self-esteem (Study 4) exhibit a more thorough elaboration of discrepancy-related information than people with small discrepancies, suggesting that highly discrepant people may be motivated to reduce their potentially aversive self-inconsistencies by additional information integration. This approach treats implicit–explicit consistency as an independent rather than a dependent variable, and may broaden our understanding of the consequences of implicit–explicit consistency for affect, well-being, and behaviour.

Coda

Despite early emphasis on dissociation between implicit and explicit constructs, there is now substantial evidence that they are related, and that the magnitude of that relationship varies across content domains and individuals. This chapter reviewed empirical evidence for a wide variety of moderators of implicit–explicit consistency, helping to reveal *when* and *why* implicit–explicit consistency is observed. Continued empirical investigations of potential moderators will clarify the value of distinguishing implicit and explicit constructs, and help to refine the theoretical models that specify their interrelations.

REFERENCES

Aiken, L. S., & West, S. G. (Eds.). (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behaviour in personality and social psychology. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 1–63). New York: Academic Press.

Ajzen, I. T., & Fishbein, M. (1977). Attitude–behaviour relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*, 888–918.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry*, *15*, 257–278.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behaviour. *Journal of Personality and Social Psychology*, *83*, 380–393.

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring task-switching ability in the implicit association test. *Experimental Psychology*, *52*, 167–179.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder*. Washington, DC: American Psychological Association.

Banse, R., & Fischer, I. (2002). *Implicit and explicit aggressiveness and the prediction of aggressive behaviour*. Poster presented at the 11th conference on Personality of the European Association of Personality Psychology, Jena, 21–25 July.

Banse, R., & Gawronski, B. (2003). Die Skala Motivation zu vorurteilsfreiem Verhalten: Skaleneigenschaften und Validierung [The scale Motivation to Act without Prejudice: Psychometric properties and validity]. *Diagnostika, 49*, 4–13.

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.

Bargh, J. (1997). The automaticity of everyday life. In J. R. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 1–61). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In J. R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 2, pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry, 7*, 1–15.

Baumeister, R. F., & Tice, D. M. (1988). Metatraits. *Journal of Personality, 56*, 571–598.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.

Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistency in behaviour. *Psychological Review, 81*, 506–520.

Blair, I. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242–261.

Block, J. (1977). Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads* (pp. 37–63). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Brauer, M., Wasel, W., & Niedenthal, P. (2000). Implicit and explicit components of prejudice. *Review of General Psychology, 4*, 79–101.

Brendl, C. M., Markman, A. B., & Messner, C. (2005). Indirectly measuring evaluations of several attitude objects in relation to a neutral reference point. *Journal of Experimental Social Psychology, 41*, 346–368.

Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology, 23*, 1933–1967.

Briñol, P., Horcajo, J., Becerra, A., Falces, C., & Sierra, B. (2002). Cambio de actitudes implicitas [Implicit attitude change]. *Psicothema, 14*, 771–775.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*, 822–848.

Buss, A. H. (Ed.). (1980). *Self-consciousness and social anxiety*. San Francisco, CA: Freeman.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131.

Cai, H., Sriram, N., Greenwald, A. G., & McFarland, S. G. (2004). The Implicit Association Test's *d* measure can minimize a cognitive skill confound: Comment on McFarland and Crouch (2002). *Social Cognition, 22*, 673–684.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. Wyer & T. Srull (Eds.), *Handbook of social cognition: Vol. 1. Basic processes* (pp. 323–417). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Converse, P. E. (1970). Attitudes and non-attitudes: Continuation of a dialogue. In E. E. Tufte (Ed.), *The quantitative analysis of social problems* (pp. 168 – 189). Reading, MA: Addison-Wesley.

Cronbach, L. J., & Snow, R. E. (Eds.). (1977). *Aptitudes and instructional methods. A handbook for research on interactions.* New York: Irvington.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measurement: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163 – 170.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50,* 77 – 85.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12,* 45 – 61.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56,* 5 – 18.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. Gaertner (Eds.), *Blackwell handbook of social psychology: Vol. 4. Intergroup processes* (pp. 175 – 197). Oxford, UK: Blackwell.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82,* 62 – 68.

Dovidio, J. F., Kawakami, K., Johnson, C., & Johnson, B. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology, 33,* 510 – 540.

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.

Edwards, A. L. (1967). The social desirability variable: A review of the evidence. In I. A. Berg (Ed.), *Response set in personality assessment* (pp. 48 – 70). Chicago, IL: Aldine.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for assessing anxiety. *Journal of Personality and Social Psychology, 83,* 1441 – 1455.

Egloff, B., & Schmukle, S. C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences, 35,* 1697 – 1706.

Epstein, S., & O'Brien, E. (1985). The person – situation debate in historical and current perspective. *Psychological Bulletin, 98,* 513 – 537.

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

Fazio, R. H., & Hilden, L. E. (2001). Emotional reactions to a seemingly prejudiced response: The role of automatically activated racial attitudes and motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin, 27,* 538 – 549.

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.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology, 54,* 297 – 327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50,* 229 – 238.

Fazio, R. H., & Zanna, M. P. (1978). On the predictive validity of attitudes: The role of direct experience and confidence. *Journal of Personality, 46,* 228 – 243.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 42,* 522 – 527.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–140.

Festinger, L. (Ed.). (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.

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

Fiedler, K., & Bluemke, M. (in press). Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*.

Flammer, L. J. (2001). *The nature of prejudice: Dimensions and patterns of racism, sexism, classism, and heterosexism among social groups*. Dissertation, Temple University, PA, USA.

Florack, A., Scarabis, M., & Bless, H. (2001a). Der Einfluss wahrgenommener Bedrohung auf die Nutzung automatischer Assoziationen bei der Personenbeurteilung [The impact of perceived threat on the use of automatic associations in person judgements]. *Zeitschrift für Sozialpsychologie*, *32*, 249–259.

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

Florack, A., Scarabis, M., & Gosejohann, S. (2005). Regulatory focus and cue relevance in product choice and persuasion. *Unpublished manuscript*.

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.

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice related IATs. *Experimental Psychology*, *49*, 171–180.

Gawronski, B., & Bodenhausen, G. V. (2004). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Unpublished manuscript*.

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., & 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. (in press). 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 wave of implicit measures* Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Gibbons, F. X. (1983). Self-attention and self-report: The ''Veridicality'' hypothesis. *Journal of Personality*, *51*, 517–542.

Gilbert, D. (1999). What the mind's not. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 3–11). New York: Guilford Press.

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

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

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie*, *48*, 85–93.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test I: An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Grover, V. P., Keel, P. K., & Mitchell, J. P. (2003). Gender differences in implicit weight identity. *International Journal of Eating Disorders*, *34*, 125 – 135.

Gschwendner, T., Hofmann, W., & Schmitt, M. (2004). On implicit – explicit consistency: Do functionally equivalent situational and personal variables of awareness and adjustment serve as moderators? *Unpublished manuscript*.

Hartshorne, H., & May, M. A. (Eds.). (1928). *Studies in the nature of character* [Vol. 1: *Studies in deceit*]. New York: Macmillan.

Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology*, *35*, 512 – 559.

Higgins, E. T. (1998). Promotion and prevention: regulatory focus as a motivational principle. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 30, pp. 1 – 46). New York: Academic Press.

Hixon, J. G., & Swann, W. B. (1993). When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. *Journal of Personality and Social Psychology*, *64*, 35 – 43.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005a). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369 – 1385.

Hofmann, W., Gschwendner, T., & Schmitt, M. (2005b). On implicit – explicit consistency: The moderating role of individual differences in awareness and adjustment. *European Journal of Personality*, *19*, 25 – 49.

Hsee, C. K., Loewenstein, G., Blount, S., & Bazerman, M. (1999). Preference reversals between joint and separate evaluations of options: a theoretical analysis. *Psychological Bulletin*, *125*, 576 – 590.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*, 349 – 364.

Jordan, C. H., Spencer, S. J., & Zanna, M. P. (2005). Types of high self-esteem and prejudice: How implicit self-esteem relates to ethnic discrimination among high explicit self-esteem individuals. *Personality and Social Psychology Bulletin*, *31*, 693 – 702.

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., & Correll, J. (2003). Secure and defensive high self-esteem. *Journal of Personality and Social Psychology*, *85*, 969 – 978.

Judd, C. M., & Kulik, J. A. (1980). Schematic effects of social attitudes on information processing and recall. *Journal of Personality and Social Psychology*, *38*, 569 – 578.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774 – 788.

Karpinski, A., Steinman, R. B., & Hilton, J. L. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. *Personality and Social Psychology Bulletin*, *31*, 949 – 962.

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

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person – situation debate. *American Psychologist*, *43*, 23 – 34.

Khilstrom, J. F. (1999). The psychological unconscious. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality*. *Theory and research* (pp. 424 – 442). New York: Guilford Press.

Kim, D. Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly*, *66*, 83 – 96.

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.

Kraus, S. J. (1995). Attitudes and the prediction of behaviour: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin, 21*, 58–75.

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging, 15*, 126–147.

Kruglanski, A. W. (Ed.). (1989). Lay epistemics and human knowledge: Cognitive and motivational bases. New York: Plenum Press.

Lambert, A. J., Barton, L. R., Lickel, B., & Wells, J. (1998). The influence of group variability and processing goals on the ease of making judgements about social categories. *Personality and Social Psychology Bulletin, 24*, 807–820.

Lambert, A. J., Payne, G. C., Ramsey, S., & Shaffer, L. M. (2005). On the predictive validity of implicit attitude measures: The moderating effect of perceived group variability. *Journal of Experimental Social Psychology, 41*, 114–128.

LaPiere, R. T. (1934). Attitudes versus action. *Social Forces, 13*, 230–237.

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

Magnusson, D., & Endler, N. S. (1977). Interactional psychology: Present status and future prospects. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads* (pp. 3–31). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behaviour. *Zeitschrift für Experimentelle Psychologie, 48*, 161–175.

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

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20*, 483–510.

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*, 1180–1192.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Monteith, M. J., Devine, P. G., & Zuwerink, J. R. (1993). Self-directed vs. other-directed affect as a consequence of prejudice-related discrepancies. *Journal of Personality and Social Psychology, 64*, 198–210.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167–184.

Nier, J. A. (2005). How dissociated are implicit and explicit attitudes? A bogus pipeline approach. *Group Processes and Intergroup Relations, 8*, 39–52.

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

Nosek, B. A. (2002). *Moderators of the relationship between implicit and explicit attitudes* Doctoral Dissertation, Yale University, USA.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General, 134*, 565–584.

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*, 625–666.

Nosek, B. A., & Banaji, M. R. (2002). (At least) two factors moderate the relationship between implicit and explicit attitudes. In R. K. Ohme & M. Jarymowicz (Eds.), *Natura Automatyzmow* (pp. 49–56). Warszawa: WIP PAN & SWPS.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding the Implicit Association Test II: Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166–180.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behaviour*. New York: Psychology Press.

Nosek, B. A., & Hansen, J. J. (2005). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit cognition. *Unpublished manuscript*.

Nosek, B. A., & Smyth, F. L. (2005). A multitrait – multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Unpublished manuscript*.

Nosek, B. A., & Vianello, M. (2005). The sorting paired features task. *Unpublished manuscript*.

Nuttin, J. M. (1985). Narcissism beyond gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, *15*, 353 – 361.

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. (2004a). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653 – 667.

Olson, M. A., & Fazio, R. H. (2004b). Trait inferences as a function of automatically activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology*, Mar, *26*, 1 – 11.

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*, 972 – 987.

Payne, B. K., Cheng, C. 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.

Pelham, B. W., Koole, S. L., Hardin, C. D., Hetts, J. J., Seah, E., & DeHart, T. (2005). Gender moderates the relation between implicit and explicit self-esteem. *Journal of Experimental Social Psychology*, *41*, 84 – 89.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, *44*, 29 – 45.

Petty, R., & Krosnick, J. A. (Eds.). (1995). Attitude strength. Antecedents and consequences. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811 – 832.

Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R. B., & Hood, R. (1977). Self-focused attention and self-report validity. *Journal of Personality*, *45*, 513 – 527.

Ranganath, K., Smith, C., & Nosek, B. (2005). Separating automatic and controlled processes in the expression of attitudes. *Unpublished manuscript*.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, *81*, 856 – 868.

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.

Schmitt, M. (1990). *Konsistenz als Persönlichkeitseigenschaft? Moderatorvariablen in der Persönlichkeits- und Einstellungsforschung [Consistency as a trait? Moderator variables in personality and attitude research]*. Berlin: Springer.

Schmukle, S. C., & Egloff, B. (2004). Does the Implicit Association Test for assessing anxiety measure trait and state variance? *European Journal of Personality*, *18*, 483 – 494.

Schuman, H., Steeh, C., Bobo, L., & Krysan, M. (1997). *Racial attitudes in America: Trends and interpretations*. Cambridge, MA: Harvard University Press.

Schwarz, N., & Sudman, S. (1992). *Context effects in social and psychological research*. New York: Springer.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Smith, E. C., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, *4*, 108–131.

Snyder, M. (1979). Self-monitoring processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12, pp. 86–128). New York: Academic Press.

Snyder, M., & Cantor, N. (1980). Thinking about ourselves and others: Self-monitoring and social knowledge. *Journal of Personality and Social Psychology*, *39,* 222–234.

Snyder, M., & Ickes, W. (1985). Personality and social behaviour. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 883–947). New York: Random House.

Spence, A. (in press). Using implicit tasks in attitude research: A review and a guide. *Social Psychology Review*.

Steffens, M. C. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology*, *51*, 165–179.

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

Sudman, S., Bradburn, M. N., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude–behaviour consistency for stigmatised behaviour. *Cognition and Emotion*, *15*, 207–230.

Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology*, *110*, 226–235.

Teachman, B. A., & Woody, S. R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology*, *112*, 100–109.

Tedeschi, J. T., & Riess, M. (1981). Verbal strategies in impression management. In C. Antaki (Ed.), *The psychology of ordinary explanations of social behaviour* (pp. 271–309). London: Academic Press.

Thurstone, L. L., & Chave, E. J. (Eds.). (1929). The measurement of attitude. Chicago, IL: University of Chicago Press.

Turner, R. G. (1978). Consistency, self-consciousness, and the predictive validity of typical and maximal personality measures. *Journal of Research in Personality*, *12*, 117–132.

Visser, P. S., Krosnick, J. A., & Simmons, P. J. (2003). Distinguishing the cognitive and behavioural consequences of attitude importance and certainty: A new approach to testing the common-factor hypothesis. *Journal of Experimental Social Psychology*, *39*, 118–141.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1997). The linguistic intergroup bias as an implicit indicator of prejudice. *Journal of Experimental Social Psychology*, *33*, 490–509.

Wadden, T. A., Brown, G., Foster, G. D., & Linowitz, J. R. (1991). Salience of weight-related worries in adolescent males and females. *International Journal of Eating Disorders*, *10*, 407–414.

Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 446–496). New York: McGraw-Hill.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioural responses to attitude objects. *Journal of Social Issues*, *25*, 41 – 78.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude – behaviour consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 287 – 343). New York: Academic Press.

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37 – 65). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship 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.

Zanna, M. P., & Fazio, R. H. (1982). The attitude – behaviour relation: Moving toward a third generation of research. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behaviour: The Ontario symposium* (Vol. 2, pp. 283 – 301). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Ziegert, J. C., & Hanges, P. J. (2005). Employment discrimination: The role of implicit attitudes, motivation, and climate for racial bias. *Journal of Applied Psychology*, *90*, 553 – 562.

Zuckerman, M., Koestner, R., DeBoy, T., Garcia, T., Maresca, B. C., & Sartoris, J. M. (1988). To predict some of the people some of the time: A reexamination of the moderator variable approach in personality theory. *Journal of Personality and Social Psychology*, *54*, 1006 – 1019.

APPENDIX

The following table provides a catalogue of the studies in each section in the order in which they are reviewed. For each study, the left-hand column indicates the moderator investigated. Column one also provides information on *how* a given moderator candidate has been investigated (see brackets): Following a taxonomy of moderators by Snyder and Ickes (1985) and Schmitt (1990), we distinguish between *qualifying attributes* (QA) of attitude or trait domains (e.g., differences in the degree of social desirability elicited by different attitude domains), *personal moderators* (P) (e.g., individual differences in socially desirable responding), and *situational moderators* (S) (e.g., differences in the degree to which various situations elicit social desirability concerns).

TABLE 1
Overview of empirical findings on the moderation of implicit–explicit consistency

Moderator	Study	Topic	Implicit measure(s)	Central findings
1. Translation				
Attitude strength (QA, P)	Nosek, 2005	57 attitude objects	IAT	Higher consistency for strong attitudes (composite of importance, thought frequency, and familiarity) in a multilevel-analysis.
Cognitive elaboration (QA)	Nosek & Banaji, 2002	15 content domains	IAT	Higher consistency for topics associated with high cognitive elaboration (meta-analysis, $\beta_{P*M} = .61*$).
Attitude importance (P)	Hofmann et al., 2005b	Attitudes towards Turks	IAT	Higher consistency for persons high in attitude importance ($\beta_{P*M} = .21*$).
	Karpinski et al., 2005	Political preference/ consumer attitudes	IAT	Attitude importance positively moderates I-E consistency for political preference (Study 1: sig. interaction, $t(169) = 2.41*$) and consumer products (Study 2: $r = .41*$ vs. $r = .03$).
Gender diff. in attitude importance (P)	Grover et al., 2003	Weight identity	IAT	Women show higher consistency for implicitly and explicitly assessed weight identity than men ($r = .36*$ vs. $r = .14$).
Prior experience (QA)	Marsh et al., 2001	Condom use	IAT / EP	Higher consistency for persons with prior experience (IAT: $r = .46*$ vs. $-.16$; Priming: $r = .30*$ vs. $-.14$).

(*continued*)

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
Stereotype certainty (P)	Lambert et al., 2005	Attitudes towards Blacks	LDT	Persons high on stereotype certainty (i.e., high perceived outgroup variability) show stronger correspondence among implicit attitudes towards Blacks and explicit rating of a Black target person ($\beta_{P*M} = .36*$).
Dimensionality (QA, P)	Nosek, 2005	57 content domains	IAT	Higher consistency for attitudes with a bipolar structure than for attitudes with a unipolar structure (multilevel-analysis).
Distinctiveness (QA, P)	Nosek, 2005	57 attitude objects	IAT	Higher consistency when own attitudes are perceived as discrepant from the general norm (multilevel-analysis).
Private self-consciousness (P)	Gschwendner et al., 2004	Attitudes toward Turks	IAT	Higher consistency for persons high in private self-consciousness ($\beta_{P*M} = .35*$).
	Hofmann et al., 2005b	Attitudes towards East Germans / Turks	IAT	No significant moderator effect for private self-consciousness ($\beta_{P*M} = .03$ in Study 1, $\beta_{P*M} = .01$ in Study 2).
Mindfulness (P)	Brown & Ryan, 2003	Affective state	IAT	Higher consistency between implicitly and explicitly assessed affective state for persons high in mindfulness ($\beta_{P*M} = .24*$).

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
Introspection (QA)	Hofmann et al., 2005a	53 content domains	IAT	Higher consistency for topics associated with low cognitive introspection (meta-analysis, $\beta_{P*M} = -.13*$).
2. Additional information integration				
Gut reactions (S)	Ranganath et al., 2005	Homosexuality	IAT, GNAT, SPF-4	A structural equation model with self-reported "gut feelings" and implicitly measured evaluations as a single factor and self-reported "actual feelings" as a second factor fits the data significantly. better than a model loading implicit vs. self-report measures on separate factors.
Time pressure (S)	Ranganath et al., 2005	Homosexuality / jazz vs. teen pop	IAT	In both content domains, a model with a speeded self-report measure and an IAT as a single factor and a standard self-report measure as a second factor fits the data significantly better than a model loading implicit vs. self-report measures on separate factors.
Response time of explicit self-report (P)	Koole et al., 2001,	Self-esteem	NLT	Significant correlation between implicitly (Name Letter Task) and explicitly assessed self-esteem (RSE) only for fast responders but not for slow responders to the explicit measure ($r = .51*$ vs. $r = -.06$).

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
Cognitive load (S)	Koole et al., 2001,	Self-esteem	NLT	Significant implicit–explicit correlation only for participants put under high cognitive load ($r = .48$* vs. $r = -.15$).
Spontaneity (QA)	Hofmann et al., 2005a	53 content domains	IAT	Higher consistency for topics associated with a high degree of spontaneous responding (meta-analysis, $\beta_{P*M} = .20$*).
Need for cognition (P)	Florack et al., 2001b	Attitude towards Turks	IAT	Higher consistency for persons low in need for cognition between IAT and explicit attitude towards Turks ($\beta_{P*M} = -3.23$*) and rating of Turkish target person ($\beta_{P*M} = -3.50$*).
Regulatory focus (S)	Florack et al., 2005	Fruits vs. chocolate	IAT	Participants primed with a promotion focus show greater correspondence between implicitly assessed consumer attitudes and explicitly assessed preferences than participants primed with a prevention focus ($r = .40$* vs. $r = -.20$).
Gender as proxy for spontaneity (P)	Pelham et al., 2005	Self-esteem	ISES / NLT	Women show higher consistency between implicitly and explicitly assessed self-esteem than men in six samples (average r: $.27$* for women vs. $-.04$ for men).

(*continued*)

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
Cognitive dissonance (S)	Gawronski & Strack, 2004	Attitudes towards alcohol prohibition / scholarship policy towards Blacks	IAT	Higher consistency between implicitly and explicitly assessed attitudes when writing a counter-attitudinal essay was justified (forced assignment) than when it was not justified (free choice) (Study 1: $r = .46*$ justification vs. $r = -.11$ no justification; Study 2: $r = .35$ justification vs. $r = -.25$ no justification).
3. Explicit assessment				
Stigmatisation (QA)	Swanson et al., 2001	Attitudes towards vegetarian food / smoking	IAT	Significant correlations between implicitly and explicitly assessed attitudes toward vegetarian food ($r = .47*$) but not for (stigmatised) smoking ($r = -.06$).
Normative protection against discrimination (QA)	Franco & Maass, 1999	Attitudes towards Islamic Fundamentalists / Jews	LIB	Consistency among implicit prejudice measure and explicit liking/reward allocation for Islamic Fundamentalist target group without normative protection against prejudice ($r = .36*$ / $r = .31*$), but not for normatively protected Jewish outgroup ($r = .03$ / $r = -.20*$).
Social desirability (QA)	Nosek & Banaji, 2002	15 content domains	IAT	Higher I-E correlations for socially *in* sensitive topics (meta-analysis, $\beta_{P*M} = -.84*$).

(*continued*)

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
	Hofmann et al., 2005a	53 content domains	IAT	No significant moderator effect of social desirability (meta-analysis, $\beta_{P*M} = -.02$).
Social desirability (P), topic-specific	Nosek, 2005	57 attitude objects	IAT	Higher I-E correlations for socially *insensitive* topics (multilevel-analysis).
Motivation to control prejudiced reactions (P)	Banse & Gawronski, 2003	Gender stereotypes	IAT	Higher consistency between implicitly and explicitly assessed attitudes for persons low in MCPR ($\beta_{P*M} = -.26^*$).
	Banse et al., 2001	Attitudes towards homosexuality	IAT	Persons with a negative implicit attitude towards homosexuality and low MCPR report most negative explicit attitudes (contrast test, $t(23) = 2.3^*$).
	Dunton & Fazio, 1997	Attitudes towards Blacks	EP	Higher consistency between implicitly (BFP) and explicitly assessed attitudes towards Blacks for persons low in MCPR (β_{P*M} = n.r., $t(42) = 2.03^*$).
	Fazio & Hilden, 2001	Attitudes towards Blacks	EP	No significant moderator effect of MCPR on the relationship between implicit attitude and three explicit affect scores.
	Fazio et al., 1995	Attitudes towards Blacks	EP	Higher consistency between implicitly and explicitly assessed attitudes for persons low in MCPR (β_{P*M} = n.r., $F_{1,106} = 5.38^*$).

(*continued*)

383

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
	Gawronski et al., 2003	Attitudes towards Turks	IAT	Higher consistency between implicitly and explicitly assessed attitudes for persons low in MCPR ($r = .37^*$ for low MCPR vs. $r = -02$ for high MCPR).
	Hofmann et al., 2005b	Attitudes towards Turks	IAT	Higher consistency between implicitly and explicitly assessed attitudes for persons low in MCPR ($\beta_{P*M} = -.17\dagger$/ $-.06$ in Study 1, $\beta_{P*M} = -.19^*$ in Study 2).
	Olson & Fazio, 2004b	Attitudes towards Blacks	EP	Higher consistency between implicit attitude towards Blacks and explicit trait inferences of Black target persons for participants low in MCPR ($\beta_{P*M} = $ n.r., $t(51) = 2.28^*$).
	Payne et al., 2005	Attitudes towards Blacks	AMP	Higher consistency between implicitly and explicitly assessed attitudes for persons low in Concern ($\beta_{P*M} = $ n.r.*) and Restraint Subscale ($\beta_{P*M} = $ n.r.*).
Internal and external motivation (P)	Hofmann et al., 2005b	Attitudes towards Turks	IAT	Internal motivation negatively moderates I-E consistency ($\beta_{P*M} = -.20^*$), but not external Motivation ($\beta_{P*M} = -.04$).

(*continued*)

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
	Payne et al., 2005	Attitudes towards Blacks	AMP	Higher consistency between implicitly assessed evaluation of Blacks and self-reported feelings for persons low in internal ($\beta_{P*M} = $ n.r.*) but not external motivation ($\beta_{P*M} = $ n.r.).
External motivation (P)	Ziegert & Hanges, 2005	Attitudes towards Blacks	IAT	External motivation negatively moderates the relationship between IAT and ATB scale ($F_{inc}(1,95) = 4.76$*; HLM analysis) but not between IAT and MRS scale ($F_{inc}(1,95) = 2.09$; HLM analysis).
Social desirability (P), general	Banse & Fischer, 2002	Aggression	IAT	Higher consistency for persons with low tendency for socially desirable responding ($\beta_{P*M} = $ n.r.†).
	Egloff & Schmukle, 2003	Anxiety	IAT	No significant moderator effect for social desirability (CM: $\beta_{P*M} = .05$ in Study 1, BIDR: $\beta_{P*M} = .10$ in Study 2.
	Hofmann et al., 2005b	Attitudes towards East Germans	IAT	No significant moderator effect for social desirability ($\beta_{P*M} = -.01/.14$).
Public self-consciousness (P)	Hofmann et al., 2005b	Attitudes towards East Germans / Turks	IAT	Higher I-E consistency for persons low in public self-consciousness ($\beta_{P*M} = -.27$*/$-.24$† in Study 1, but $\beta_{P*M} = -.08$ in Study 2).
Self-monitoring (P)	Hofmann et al., 2005b	Attitudes towards East Germans	IAT	No significant moderator effect of self-monitoring ($\beta_{P*M} = .12/.01$).

385

TABLE 1
(Continued)

Moderator	Study	Topic	Implicit measure(s)	Central findings
Social desirability (S)	Nier, 2005	Attitudes towards Blacks	IAT	Significant I-E correlation for persons led to believe that IAT was true measure of attitudes ($r = .51^*$ vs. $r = .18/.14$ CGs).
	Nosek, 2002	Attitudes towards science and homosexuality / political preferences	IAT	Privately assessed explicit attitudes mediate the relationship between IAT attitude measures and publicly assessed explicit attitudes in two (homosexuality, science) of three tests of mediation.
	Ziegert & Hanges, 2005	Attitudes towards Blacks	IAT	Corporate climate (climate for equality vs. climate for bias) moderates the relationship between implicit attitude towards Blacks and explicit bias against black job applicants ($\beta_{P*M} = .17^*$ in HLM analysis).
Outgroup threat (S)	Florack et al., 2001a	Attitudes towards Turks	IAT	Significant correlation between implicit measure of prejudice towards Turks (IAT) and attribution of guilt towards Turkish target person only for participants who feel threatened by the Turkish outgroup ($r = -.58^*$ vs. $r = .21$ CG); no moderator effect for trait rating of target.

(continued)

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
	Rudman & Lee, 2002	Attitudes towards Blacks	IAT	Significant correlation between implicit measure of prejudice towards Blacks (IAT) and explicit rating of Black target person only when participants are presented with violent rap music (rating of hostility: $r = .48^*$ vs. $r = .16$; rating of sexism: $r = .47^*$ vs. $r = .12$).
4. *Implicit assessment*				
Situational preactivation (S)	Wittenbrink et al., 2001	Attitudes towards Blacks	IAT	Significant correlations between implicitly and explicitly assessed prejudice towards Blacks only for persons presented with a movie clip of Black Americans in a negative stereotypic setting ($r = .34^*$ for negative setting vs. $r = -.19$ for positive setting).
	Wittenbrink et al., 2001	Attitudes towards Blacks	LDT	Significant correlations between sequential priming paradigm and explicit measures only for primes embedded in a negative (dilapidated street corner) setting as compared with a positive (church) setting ($r = .35^*$ vs. $r = .05$).

(*continued*)

TABLE 1
(*Continued*)

Moderator	Study	Topic	Implicit measure(s)	Central findings
Extra-personal associations (S)	Olson & Fazio, 2004	Attitudes towards apples vs. candy bars / Bush vs. Gore	IAT	Higher relationship between personalised IAT version and explicit self-report compared with a traditional IAT version (Study 3: $r = .69*$ vs. $r = .09$; Study 4: $r = .81*$ vs. $r = .45*$).
	Nosek & Hansen, 2005	58 content domains	IAT	Explicit attitudes predict traditional and personalised IAT scores equally well across 58 domains (average $\beta = .33*$ vs. average $\beta = .34*$). Further, cultural knowledge is insignificantly and not differentially associated with the traditional and the personalised IAT (average $\beta = .04$ vs. average $\beta = -.03$).
Faking (S)	Kim, 2003	Attitudes towards Blacks	IAT	Consistency among implicitly and explicitly assessed prejudice for participants who were asked to fake the IAT and given instructions on how to achieve this ($r = .46*$) but not for control group ($r = .12$).
Reliability	Cunningham et al., 2001	Attitudes towards Blacks	IAT, EP	Higher consistency when modelling a second-order implicit factor and controlling for measurement error in SEM.

(*continued*)

TABLE 1
(Continued)

Moderator	Study	Topic	Implicit measure(s)	Central findings
	Nosek & Smyth, 2005	57 content domains	IAT	Higher consistency with structural equation modelling (median of $r = .48^*$) compared with zero-order correlation (median of $r = .37^*$; reanalysis of data from Nosek, 2005).
5. Design factors				
Implicit–explicit order (S)	Bosson et al., 2000	Self-esteem	IAT, WFC, NLT, SPP, SBP, ISES, Stroop	Higher implicit–explicit correlations when explicit measures are administered first ($F(1,6) = 15.54^*$).
	Nosek et al., 2005	8 attitude and stereotype domains	IAT	No moderator effect of implicit–explicit order on consistency (average effect $q = .00$).
	Hofmann et al., 2005a	57 content domains	IAT	No moderator effect of implicit–explicit order on consistency ($r_{\text{expl-impl}} = .24^*$ vs. $r_{\text{impl-expl}} = .22^*$).
Measurement correspondence	Hofmann et al., 2005a	53 content domains	IAT	Significantly higher correlations of IAT with regard to trait ratings ($r = .29^*$), semantic differentials ($r = .28^*$), and feeling thermometers ($r = .24^*$) than with regard to more complex questionnaire scales ($r = .18$).

(*continued*)

389

TABLE 1
(Continued)

Study	Topic	Implicit measure(s)	Central findings
Hofmann et al., 2005a	53 content domains	IAT	Higher consistency between IAT and relative (vs. absolute) explicit measures (meta-analysis, $r = .27^*$ vs. $r = .19$).
Hofmann et al., 2005a	53 content domains	IAT	Higher consistency between IAT and affective (vs. cognitive) explicit measures (meta-analysis, $r = .28^*$ vs. $r = .18$).
Gawronski, 2002	Attitudes toward Turks and Asians	IAT	Significant I-E correlations for corresponding measures only (SEM).
Nosek & Smyth, 2005	7 attitude domains	IAT	Median implicit–explicit correlation $r = .37^*$ for corresponding attitude measures compared to $r = .00$ for non-corresponding ones.

Correlations are scored such that positive numbers indicate higher consistency. QA = qualifying attribute; P = personal moderator variable (including gender differences); S = situational moderator variable; β_{P*M} = standardised regression weight of interaction term computed from predictor and moderator; a positive weight indicates an increase in implicit–explicit consistency. AMP = affect misattribution paradigm; BIDR = balanced inventory of desirable responding; CG = control group; CM = Crowne-Marlow inventory for socially desirable responding; EP = evaluative priming; GNAT = Go/No-Go Association Task; HLM = hierarchical linear modelling; IAT = Implicit Association Test; ISES = implicit self-evaluation scale; LIB = linguistic intergroup bias; LDT = lexical decision task; NLT = name letter task; n.r. = not reported; MCPR = motivation to control prejudiced reactions; SEM = structural equation modelling; RSE = Rosenberg self-esteem scale; SPF−4 = four-category sorting paired features task; SPP = supraliminal priming; SBP = subliminal priming; WFC = word fragment completion task.
$^{†}p < .10$. $^{*}p < .05$. Statistical parameters without indices are non-significant.