Implicit? What Do You Mean? A Comprehensive Review of the Delusive Implicitness Construct in Attitude Research
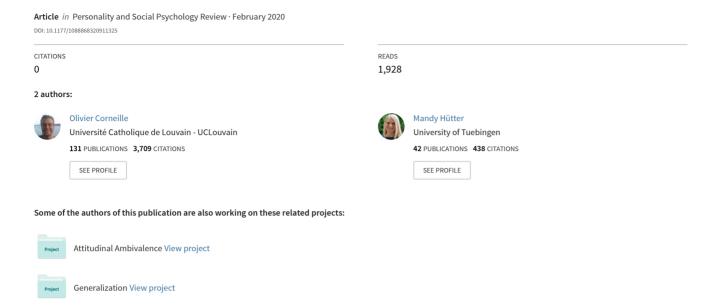
2 authors:

Olivier Corneille
Université Catholique de Louvain - UCLouvain

131 PUBLICATIONS   **3,709** CITATIONS

Mandy Hütter
University of Tuebingen

42 PUBLICATIONS   **438** CITATIONS

Some of the authors of this publication are also working on these related projects:

Attitudinal Ambivalence View project

Generalization View project

Implicit? What do you mean?

A Comprehensive Review of the Delusive Implicitness Construct in Attitude Research

Olivier Corneille, UCLouvain

Mandy Hütter, Universität Tübingen

To be published in *Personality and Social Psychology Review*

This is a preprint version

Abstract

This article provides a comprehensive review of divergent conceptualizations of the "implicit" construct that have emerged in attitude research over the past two decades. In doing so, our goal is to raise awareness of the harmful consequences of conceptual ambiguities associated with this terminology. We identify three main conceptualizations of the "implicitness" construct: The procedural conceptualization (implicit as indirect), the functional conceptualization (implicit as automatic), and the mental theory conceptualization (implicit as associative), as well as two hybrid conceptualizations (implicit as indirect and automatic, implicit as driven by affective gut reactions). We discuss critical limitations associated with each conceptualization and explain that confusion also arises from their co-existence. We recommend discontinuing the usage of the "implicit" terminology in attitude research and research inspired by it. We offer terminological alternatives aimed at increasing both the precision of theorization and the practical value of future research.

Implicit? What do you mean?

A Comprehensive Review of the Delusive Implicitness Construct in Attitude Research

"The slovenliness of our language makes it easier for us to have foolish thoughts"

George Orwell, *Politics and the English Language*

I.   Introduction

For the last two decades, a thriving body of influential research has been conducted on implicit attitudes, implicit evaluations, and implicit measures and, more generally, on implicit social cognition. While "evaluations" refer to evaluative responses, "attitudes" typically refer to the latent mental construct assumed to elicit these responses. These concepts are strongly connected, as one can only gain knowledge about a mental construct (i.e., the "implicit attitude") by measuring the observable responses it relates to (i.e., the "implicit evaluation"), typically on an "implicit measure." Research into the implicitness of attitudes, evaluations, and measures has garnered considerable attention across various psychological domains (e.g., health psychology, self and identity, personality, interpersonal and intergroup relations, behavioral and cognitive neuroscience) and beyond (e.g., moral philosophy, political science, law, consumer research, policymaking). It has contributed to the development of new measurement methods, research paradigms, and inspiring theories. The importance of these contributions should not be underestimated, nor should the importance of studying implicit processes in psychological research. Yet, this research relies on problematic and divergent definitions of the implicitness construct.

The present article provides a comprehensive review of three conceptualizations of this construct that have emerged in attitude research over the past two decades. In doing so, our goal is to raise awareness of the harmful consequences of conceptual ambiguities

associated with the "implicit" terminology. Most researchers making use of the "implicit attitudes and evaluations" constructs are probably aware that these constructs are associated with some limitations. However, an integrated discussion of these limitations is currently lacking. Throughout this article, we argue that the co-existence of diverging understandings of the "implicit" construct is problematic for theory and practice. We also point to constructive recommendations for enhanced conceptual and measurement clarity. Hence, whereas previous discussions of implicit attitudes have addressed their psychometric properties (e.g., Blanton, Jaccard, Strauts, Mitchell, & Tetlock, 2015; Schimmack, 2019), whether they reflect individual or extrapersonal representations (e.g., Arkes & Tetlock, 2004; Payne, Vuletich & Lundberg, 2017), and how they relate to explicit attitudes (e.g., Nosek, 2007) and behavior (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Kurdi, Seitchik, … Banaji, 2019; Meissner, Grigutsch, Koranyi, Müller, & Rothermund, 2019; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013), this article is essentially concerned with a *conceptual analysis* of the implicit attitude and evaluation constructs, linking this analysis to induction and measurement questions. Although conceptual analyses of these constructs have been proposed in the past and more recent literature (see e.g., Brownstein, Madva, & Gawronski, in press; De Houwer, 2006; De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; De Houwer & Moors, 2010; Gawronski, 2019; Gawronski & Hahn, 2019), the current review is distinct in its comprehensiveness and in its clear recommendation to get rid of the "implicitness" terminology in attitude research.

We first briefly discuss the theoretical background for research on implicit attitudes and evaluations (Section II). Next, in Section III, we review three major conceptual approaches to these constructs: (i) the procedural approach: implicit-as-indirect, (ii) the functional approach: implicit-as-automatic, and (iii) the mental theory approach: implicit-as-associative. In Section IV, we illustrate issues that arise from conceptual ambiguities in the

definition and operationalization of these constructs. We then discuss processing tree modeling as a way of enhancing accuracy in conceptualization and measurement (Section V). In the General Discussion (VI), we relate the main insights of this review to points of attention for future research.

II.   The background: Dual-process accounts of cognition and dual-learning models of attitudes

The implicit-explicit opposition in attitude research emerged in the broader context of dual-process accounts of cognition and behavior. In opening his widely cited review of dual-process models, Evans (2008) identified over twelve dual-process accounts of cognition, which he subsumed under an overarching System 1–System 2 opposition. Evans (2008) noted that specific dual-process accounts differ in fundamental assumptions (e.g., sequential or parallel operation of the two systems), but that they all oppose automatic (unconscious, rapid, and high capacity) to non-automatic (conscious, slow, and deliberative) mental processes.

Evans (2008) concluded that "(…) close inspection of the evidence suggests that generic dual system theory is currently oversimplified and misleading" (p. 270). Evans and Stanovich (2013) proposed to revert to a Type 1–Type 2 processing terminology in which categories of processes are opposed to each other without additionally assuming that they operate within exactly two (partially) distinct cognitive or neurological systems often associated with different phylogenetic origins (i.e., System 1 and System 2). A Type-2 process is seen as requiring central working memory capacities. It is, therefore, likely to depend on transient and chronic cognitive capacities.

The validity of the System 1–System 2 distinction was called into doubt from the very beginning by its most famous proponents. Kahneman (2011, p. 29), in his influential book *Thinking Fast and Slow*, warned the reader: *"I must make it absolutely clear that they are fictitious characters. Systems 1 and 2 are not systems in the standard sense of entities with*

interacting aspects or parts. And there is no one part of the brain that either of the systems would call home.*" It may seem perplexing that scientists promote frameworks they consider fictitious. The answer lies in the pragmatic value of this simplification. The System 1–System 2 framework offered a convenient way of simplifying communication on complex issues. It made it possible to reach out to the general public and to research domains that traditionally left little room for automatic processes. Hence, this conceptualization was useful.

Simplifications, however, while having a pragmatic value in specific contexts (e.g., raising awareness of the role of automatic processes among social psychologists, economists, or the general public) may be detrimental to others (i.e., advancing accurate knowledge). As Keren and Schul (2009, p. 534) nicely put it when discussing dual-process accounts: "(…) the proliferation of models, each using its own definition of the theoretical constructs or, even worse, using theoretical constructs that are not well defined, offers researchers and their readers a false sense of understanding. This also poses a problem for those who desire to consider the models critically because the entity under examination is not well defined."

The dual-process view of cognition has progressively found its way into social psychology. This field has been characterized by "the greatest proliferation of dual-processing labels and theories (…)" (Evans & Stanovich, 2013, p. 223). Several theories address the attitude acquisition question (i.e., dual-learning models of attitudes) by positing the existence of an association formation mode (Type 1, automatic), in addition to a propositional or rule-based learning mode (Type 2, non-automatic). The most influential of these models are the Associative-Propositional Evaluation (APE) model (Gawronski & Bodenhausen, 2006, 2011), the Systems of Evaluation Model (SEM; e.g., Rydell & McConnell, 2006; McConnell & Rydell, 2014), and the Meta-Cognitive Model (MCM; e.g., Petty & Briñol, 2006; Petty, Briñol, & DeMarree et al., 2007; for a detailed discussion of how these theories relate to associative assumptions, see Corneille & Stahl, 2019).

III.   Disparate views on the "implicit" construct in attitude research

Implicit attitude measures are classically viewed as measurement devices that reveal unconscious attitudes and assess evaluations elicited through automatic processes. These views are consistent with original definitions of the implicit construct in attitude and social cognition research:

"A template for definitions of specific categories of implicit cognition is: *An implicit C is the introspectively unidentified (or inaccurately identified) trace of past experience that mediates» R*. In this template, **C** is the label for that construct (such as attitude), and R names the category of responses (such as object-evaluative judgments) assumed to be influenced by that construct." (Greenwald & Banaji, 1995, p. 5)

"Because conceptions of attitudes as necessarily accessible to conscious awareness and control are difficult to shake off, the research summarized here will come to be viewed as historically important – as the first robust and reliable demonstrations that permitted a sufficient breakthrough to allow us to conceptualize attitudes as automatic evaluations." (Banaji, Lemm, & Carpenter, 2001, p. 147)

These quotations highlight major ambiguities in the "implicitness" terminology. First, whereas implicitness is defined both in terms of knowledge content (i.e., traces of past experiences) and in terms of a mediating factor in the first quotation, it refers to the evaluative responses themselves in the second quotation (i.e., attitudes as automatic evaluations). Second, whereas implicitness refers to the absence of accurate introspective access to knowledge content in the first quotation, it additionally refers to the automatic nature of evaluative responses in the second quotation. In sum, across and even within definitions, "implicit" refers to evaluative contents or to any feature of automaticity related to evaluative responses. Although not directly apparent in these quotations, "implicitness" can additionally refer to a mechanism, such as when individuals implicitly misattribute evaluative responses to

an object, construct, or person that was not causal in eliciting this response (e.g., Jones, Olson & Fazio, 2009).

In this section, we identify and discuss three conceptualizations of the "implicit" construct in attitude research. The first conceptualization relates this construct to the indirectness of the measurement (i.e., the procedural approach: "implicit-as-indirect"; Section III.1). The second relates it to the automaticity of evaluative responses (the functional approach: "implicit-as-automatic"; Section III.2). The third relates it to associative representations and associative learning processes (the mental theory approach: "implicit-as-associative"; Sections III.3.a and III.3.b). We additionally discuss two hybrid definitions of this construct ("implicit-as-indirect-and-automatic", and "implicit-as-driven-by-affective-gut-reactions"; Section III.4.a and III.4.b).

It should be evident throughout this section and later in this article that the existence of definitional disparities creates significant confusion. The goal of the present review is to provide a comprehensive and integrated understanding of these problems. In turn, this should help researchers and practitioners (i) build a shared understanding of the various usages of the "implicit' construct in implicit social cognition research, (ii) make informed terminological choices, and (iii) select measurement tasks and analytic strategies that best serve their particular interests.

III.1. The procedural approach: Implicit as indirect

In this first conceptualization, "implicit" refers to how attitudes are assessed, namely whether this assessment relies on self-reported evaluation of the attitude object or is inferred from overt behavior that excludes such self-report (e.g., De Houwer, 2006; De Houwer & Moors, 2010).

III.1.a. Conflation

The inventors of the IAT have recently published an article that clarifies this meaning of "implicit" by stating four equations (Greenwald & Banaji, 2017):

$$\text{implicit} = \text{indirect}; \text{explicit} = \text{direct}$$

$$\text{indirect} \neq \text{unconscious}; \text{direct} \neq \text{conscious}$$

The equations on the top indicate the procedural nature of the conceptualization: implicitness implies the use of indirect measures, whereas explicitness implies the use of direct measures. In the bottom equations, direct $\neq$ conscious indicates that the use of direct measures does not imply the study of conscious phenomena, and so does not exclude the study of unconscious ones. This is consistent with work by Jacoby (1991) and others, but also with the development of processing tree models in attitude research (e.g., Hütter, Sweldens, Stahl, Unkelbach, & Klauer, 2012; see Section V). To give one example, the mere exposure effect, often seen as the hallmark of implicit cognition (but see, e.g., Newell & Shanks, 2007; Wang & Chang, 2010) is typically evidenced on self-report measures. Finally, the indirect $\neq$ unconscious equation indicates that the use of indirect measures does not imply the study of unconscious phenomena, and so does not exclude the study of conscious ones.

The implicit-as-indirect conceptualization bears no relation to mental processes. In particular, direct and indirect measures can both inform conscious and unconscious phenomena. Greenwald and Banaji (2017) acknowledge that they "(…) find themselves occasionally lapsing to use implicit and explicit as if they had conceptual meaning" (Greenwald & Banaji, 2017, p. 862). This lapsing is unfortunate because it suggests that the choice of a direct versus indirect measure alone is informative about mental processes driving evaluative responses. Confusions may also apply when linking the "implicit-as-indirect" conceptualization to attitude acquisition and representation questions. For instance, Mann, Kurdi, and Banaji (2019) define implicit evaluations in terms of their indirect measurement:

"Whereas people seem quite able to set aside a first impression in their *explicit* (self-reported) judgments, first impressions can continue to linger in *implicit* (indirectly-measured) evaluations (attitudes)" (Mann et al., 2019, p. 1). The authors then explain why implicit evaluations may show lower sensitivity to updating: "From the perspective of dual-systems theories, implicit and explicit evaluation have been construed as the output of distinct mental systems that are subject to different learning mechanisms and contain different kinds of representations (Rydell & McConnell, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004)" (Mann et al., 2019, p. 2). However, if one is consistent with the procedural definition of "implicit", which is agnostic to mental processes, there is no reason why direct versus indirect measures should speak to "mental systems that are subject to different learning mechanisms and contain different kinds of representations."

As it appears, a major issue with the implicit-as-indirect view is that researchers who err into a process-loaded version of it are at risk of drawing unwarranted inferences about mental processes. Alternatively, those who subscribe to its process-free version endorse a conceptualization that has no evidential value for theorization. In the first case, the conceptualization is misleading. In the second, it is theoretically irrelevant. Of significant concern, one easily moves from one view to the other. For instance, Greenwald and Lai (2019) open the abstract of their *Annual Review of Psychology* article stating: "In the last 20 years, research on implicit social cognition has established that social judgments and behavior are guided by attitudes and stereotypes of which the actor may lack awareness" (p. 25.1), whereas they close their article stating: "Some interpret "implicit attitude" as meaning "unconscious attitude." This needlessly commits to a theoretical interpretation that is not established and seems unlikely to become established in the foreseeable future" (p. 25.21)." Hence, the abstract supports a process-loaded understanding of the implicit-as-indirect view, whereas the recommendation opposes it. Moreover, the phrasing in the abstract (i.e., "is

guided") suggests a *causal* influence of presumably unconscious attitudes and stereotypes on social judgments and behavior. This is unwarranted considering the lack of experimental evidence for this. We will come back to this critical point in the General Discussion.

III.1.b. Lack of Constraint

Finally, it is also not always clear how to set boundaries between indirect and direct measures. Evaluating a stimulus as positive or negative on a scale qualifies as a direct evaluative measure of this stimulus. Other cases, however, are more ambiguous, such as speeded evaluative measures, or a bogus pipeline measure, or a questionnaire assessing representations relevant to evaluations in a subtle way, such as the Modern Racism Scale (Morisson & Kriss, 2017). Therefore, researchers should specify what they mean by "indirect" measures of evaluations to facilitate precision in communication and theorizing.

III.1.c. Conclusions for the implicit-as-indirect conceptualization

Researchers who endorse the implicit-as-indirect conceptualization should keep in mind that this procedural definition has no theoretical ambition and should refrain from making process-related inferences based on the mere directness or indirectness of a task. In our view, this implies that this conceptualization should be abandoned. Strict adherence to the concept (i.e., the process-free version) cannot inform theories, whereas loose adherence (i.e., the process-loaded version) is misleading. Under this procedural conceptualization, referring to "direct" and "indirect" measures would greatly limit the risk of unwarranted inferences as well as the risk of confusions between concurrent conceptualizations of the implicitness construct. Here too, however, the opposition is theory uncommitted.

III.2. The functional approach: Implicit as automatic

To avoid unwarranted process-related inferences from the direct or indirect nature of a measurement, some researchers have proposed that the direct-indirect distinction should refer to the structure of a task (i.e., whether an evaluation is directly requested from the respondent

or not), and that the implicit-explicit distinction should refer to the conditions under which a response is produced on that task (e.g., De Houwer, 2006; De Houwer & Moors, 2010). According to this functional view, evaluations are implicit if "the impact of the to-be-measured psychological attribute on participants' responses is unintentional, resource-independent, unconscious, or uncontrollable (…)" (Gawronski & De Houwer, 2014, p. 284; see also De Houwer & Moors, 2007; De Houwer et al., 2009). Therefore, a task cannot be considered implicit. Against that recommendation, many researchers keep attaching the adjective "implicit" to the substantive "measure" (e.g., Gawronski & Hahn, 2019). The implicit-as-automatic view is also characterized by two issues that we now discuss: conflation and a lack of constraint.

III.2.a. Conflation

Various features of automaticity exist (e.g., awareness, efficiency, intentionality, controllability) that do not perfectly align with each other (e.g., Bargh, 1992, 1994; Fiedler & Hütter, 2014; Melnikoff & Bargh, 2018; Moors & De Houwer, 2006, 2007; for a recent analysis of automaticity, see Moors, 2016). Of particular concern here is that calling an evaluation "implicit" because it is elicited automatically puts an unwarranted weight on the "awareness" feature of automaticity. "Implicitness" is easily confused with unconscious processes, as it suggests that the evaluation remains "unexpressed" or "unrevealed". This semantic drift was discussed in the previous section and is also evident in research on implicit memory, sometimes called "unconscious memory" (e.g., Jacoby, Kelley, & Dywan, 1989). It is aggravated by the view that the IAT captures unconscious attitudes or unconscious biases. In other words, using "implicit" instead of "automatic" may suggest that a task speaks to unconscious processes or representations, whereas it involves processes that can be characterized by other features of automaticity (e.g., efficiency).

Under the implicit-as-automatic conceptualization, authors could warn their readers that "implicit" actually means "automatic." In this case, however, it seems preferable to use the "automatic" terminology. This choice would still mask differences between various features of automaticity. However, it would considerably reduce the risk of concluding that, just because an evaluation is automatic on one particular feature (e.g., unintentional or efficient), the evaluation or representation it captures is unconscious.

More generally, we encourage researchers to pay careful attention to how measurement tasks differently capture automatic processes, and which feature(s) of automaticity they allow assessing. Measurement tasks developed during the implicit cognition revolution help advance our understanding of evaluative responses under less optimal conditions (and in particular, conditions of reduced control, efficiency and intentionality). However, it is important to realize (i) that these tasks speak to various dimensions of automaticity, and (ii) that their automaticity should not be overstated. The most widely assumed automaticity feature of implicit attitudes is unawareness (Greenwald & Banaji, 1995; Greenwald, McGhee, & Schwartz, 1998). It is, however, debatable whether attitudes captured by measurement tasks meant to assess implicit attitudes and evaluations are unconscious. First, low correlations (e.g., Greenwald et al., 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005) between outcomes from these tasks and self-reported evaluations are not informative about the consciousness question (e.g., Blanton & Jacquard, 2015; Hofmann et al., 2005). Second, evidence suggests that people can introspect evaluations captured by these tasks (Gawronski, Hofmann, & Wilbur, 2006; Gawronski, LeBel, & Peters, 2007; Hahn, Judd, Hirsh & Blair, 2014) and may be aware of the mental contents and processes involved in the production of responses on these tasks (Cummins, Hussey, & Hughes, 2019). Likewise, these tasks - including the IAT and evaluative priming - are amenable to voluntary control (Fiedler & Bluemke, 2005; Klauer & Teige-Mocigemba, 2007; Steffens, 2004; Teige-

Mocigemba & Klauer, 2008; Teige-Mocigemba, Penzl, Becker, Henn, & Klauer, 2016). We will discuss in Section V how processing tree modeling allows quantifying the contribution of discrete features of automaticity to task outcomes.

III.2.b. Lack of constraint

A second issue has to do with the lack of constraint associated with this definition. Nosek et al. (2011) identified more than twenty social cognition measures, "to which the label 'implicit' is routinely applied" (p. 152) and acknowledged that "the inclusion/exclusion boundary for implicit measures in social cognition research is fuzzy" (p. 153). Under the implicit-as-automatic conceptualization, the number of tasks that can be characterized as "implicit" can be much higher than the one assembled by these authors. Theoretically, it is unconstrained because *any* task reveals partly automatic responses (e.g., Hütter & Klauer, 2016; Jacoby, 1991; Sherman, 2009).

If it has to be constrained, it is not clear what level of automaticity should be met for a task to be considered to measure implicit evaluations, and on what specific dimension of automaticity. Current research does not provide a satisfactory response to this important question. As a result, measurement procedures that strongly differ in their structural and functional properties (e.g., a seating distance measure, an evaluative priming measure, a linguistic bias measure, an affect misattribution procedure) are merged into an "implicit measures" category, as opposed to evaluative self-reports and dozens of measures that remain in a definitional limbo. As a consequence, measurement selection often seems guided by social convention rather than methodological considerations.

Some authors have started constraining the meaning of "implicit measures" by relating them to the use of computers: "The term 'implicit social cognition' is conventionally used to refer to research in social psychology that uses a particular class of computerized measurement instruments to infer thoughts and affective reactions without directly asking

participants to report on them" (Hahn & Gawronski, 2018, p. 395). It is true that the most widely used measures of implicit attitudes, such as the IAT or evaluative priming measures, typically use computers to record speed and accuracy data. If implicit evaluations are defined as automatic evaluations, however, the use of computers should not be critical to their measurement. Several authors have developed pencil-and-paper-based versions of originally computerized indirect measures (e.g., Lowery et al., 2001; for a discussion, see Vargas, Sekaquaptewa, & Von Hippel, 2007) and a wide range of indirect evaluative measures making no use of computers (e.g., word completion, seating distance) do also involve automatic processes. Finally, it is important to realize that even a *non-speeded self-report* may capture implicit attitudes and evaluations. For instance, people report higher levels of life satisfaction on a rainy day; their weather-driven mood unknowingly influences their direct evaluation (Schwarz & Clore, 1983; see also Section V).

III.2.c. Conclusions for the implicit-as-automatic conceptualization

We recommend that researchers who define implicit evaluations as automatic (i) call them "automatic" instead of "implicit", (ii) systematically refer to the feature of automaticity they are interested in, and (iii) select a task that is best suited for addressing this particular feature of automaticity. For instance, if interested in efficient evaluations, an evaluative priming measure seems more relevant than a seating distance measure. Also, (iv) researchers should be careful not to overgeneralize outcomes from one specific evaluative task to the superordinate "implicit evaluation", "implicit attitude", "implicit bias", or "unconscious bias" levels. This is because different tasks are characterized by the operation of different cognitive processes, showing different levels of automaticity on different dimensions of automaticity. It should be clear that these recommendations do not dispute the importance of studying evaluations under suboptimal processing conditions. Rather, they suggest discontinuing the confusing usage of the "implicit" terminology for its study.

III.2.d. Choice of wording for the remainder of this review

So far, we have argued that "implicit" refers to (i) the indirect measurement of attitudes and evaluations (i.e., implicit-as-indirect), and to (ii) responses indicative of evaluations produced under conditions of automaticity (i.e., implicit-as-automatic). The confusion and overgeneralizations this terminology elicits (see Section IV) are likely to apply, by logical necessity, to the present review as well, as it constrains us to use a construct that we say is confusing and whose usage we claim should be suspended. In the remainder of this review, we will use "IAEs" as a generic label when referring to measurement outcomes from tasks commonly said to capture implicit attitudes and evaluations. We realize this solution is not ideal, but the best we could think of. Otherwise, we would have to systematically refer to the lengthy "measurement outcomes from tasks commonly said to capture implicit attitudes and evaluations" or to the problematic "implicit measures" wording.

III.3. The mental theory approach: Implicit as associative

A third cluster relates the implicit construct to associative mental process theories. Consider, for instance, these quotations: "In contrast with explicit beliefs that are conscious and deliberate, implicit associations are automatically activated (…)" (Régner et al., 2019, p. 1171). "Implicit measures of attitudes that capture the associative system of evaluation (…) are strongly influenced (…) by cues that are subliminal in nature" (McConnell & Rydell, 2014, pp. 208-209). In this section, we explain that IAEs and associations have been theoretically related on the representation and acquisition levels. We discuss both views and limitations associated with them. The main limitation here relates to the lack of conclusive empirical support for associative attitude learning in general and the view that it is better indicated by IAEs in particular. In addition, we briefly discuss theoretical and measurement limitations related to the associative view.

III.3.a. Implicit as based on associative representations

Associations are typically seen as links between nodes that can be activated to varying degree and possibly inhibited. This simple associative representation does not easily accommodate the complexity of human cognition. For instance, activation spreading through the "Mary", "Paul" and "Love" nodes fails to represent whether Mary loves Paul, or Paul loves Mary, or whether they love each other. More complex associative representations exist (e.g., multilayer connectionist models), but all face this basic limitation to some degree (e.g., Hummel, 2010). Helpful discussions of the "association" construct have been proposed elsewhere (e.g., Gawronski & Sritharan, 2010). We encourage the reader to pay attention to these analyzes and also to consider propositional theories as a theoretical alternative to associative ones. Propositional theories of human cognition and behavior have proved increasingly influential in social psychological research (for a recent discussion, see De Houwer, Van Dessel, & Moran, 2019).

Despite rejecting the ambiguous "implicit attitude" construct, the MODE model (e.g., Fazio, 1990) was influential in relating the "implicit" terminology to the "association" construct (see, e.g., Gawronski & Sritharan, 2010). This theory posits that attitudes are represented as links between attitude objects and summary evaluation of these objects in memory. These links can be automatically activated upon encountering the attitude object when evaluative associations are strong enough (e.g., Fazio, 2007). In the MODE model, motivation and opportunity to engage in effortful processing moderate the impact of this automatic activation on judgments and behavior (including evaluative responses). Because tasks meant to capture IAEs typically rely on less intentional and controlled processes, these tasks are thought to reflect associative knowledge ideally. In line with this view, when participants lack motivation or opportunity to engage in effortful processing (e.g., when they

are asked for a speeded judgment), self-reported evaluations correspond more strongly with IAEs (e.g., Ranganath, Smith, & Nosek, 2008).

As it appears, whereas automaticity is considered on a functional level under the implicit-as-automatic conceptualization, it is meant to inform associative theories under the implicit-as-associative approach. However, neither do IAEs (exclusively) capture associations (for instance, they also capture executive functions; Ito et al., 2015, or causal relations; Hughes, Ye, Van Dessel, & De Houwer, 2018; Kurdi, Morris, Cushman, 2020; see also Fiedler, Messner, & Bluemke, 2006), nor do they all assess associations equally well. The IAT is often considered the ideal-case procedure for assessing associations. For instance, Charlesworth, Kurdi, and Banaji (2019) recently noted that their findings are "particularly impressive considering that implicit attitudes were measured on an arguably "associative" test (the Implicit *Association* Test)." (p. 7). We consider this an overstatement and argue that what is needed is a way to separate "associative" versus "non-associative" contributions to IAEs. We discuss in Section V how to achieve this.

III.3.b. Implicit as based on association formation

Whereas the above discussion essentially linked the "implicit" and "association" constructs on a representational level (i.e., IAEs reveal the automatic activation of stored associative knowledge), these constructs have also been linked on the acquisition level (i.e., IAEs are the outcome of association formation). The learning mode thought to underlie the formation of IAEs is termed "associative" and is described by Hebbian principles of joint co-activation (e.g., Gawronski & Bodenhausen, 2018). Contrary to propositional or rule-based learning, associative learning, it is claimed, automatically registers mere co-occurrences between events, independently of their meaning. Dual-learning theories also point to the slow-paced accumulation of information (e.g., McConnell & Rydell, 2014). In sum, automaticity, slow-pace, and unqualified links between concepts are considered core features of association

formation processes. No dual-learning theory states that IAEs can only capture associative learning processes. However, influential dual-learning theories such as the APE and the SEM assume double dissociations between learning mode and measurement modes under typical circumstances. As Gawronski and Bodenhausen (2011) put it: "Whereas implicit evaluations are the behavioral outcome of associative processes, explicit evaluations represent the behavioral outcome of propositional processes" (p. 61). Of note, we consider this theorizing coherent, and so it does not threaten the validity of the "implicit" construct. Rather, it is important to note that this theorizing has not received conclusive empirical support.

Attitude research primarily relied on the evaluative conditioning paradigm for testing assumptions from associative learning theories. This paradigm has a straightforward "associative" structure and so has been considered ideally suited for the study of associative attitude learning (e.g., Olson & Fazio, 2001). As Hahn and Gawronski (2018) recently noted: "The idea of associative learning is most prominently reflected in research on evaluative conditioning" (p. 407). In an evaluative conditioning procedure, a formerly neutral stimulus (i.e., a conditioned stimulus, or CS) typically acquires the valence of the valent stimulus after having been paired with it. For instance, a neutral pictogram becomes more positive after having been paired with a positive picture.

Corneille and Stahl (2019) reviewed evidence from the evaluative conditioning and related paradigms. They concluded that empirical evidence for associative attitude learning is weak at best, both when defining it in terms of operating conditions (i.e., automaticity) and operating principles (e.g., the establishment of unqualified associations). In the next section, we argue that, even assuming evidence exists for associative attitude learning theories, this evidence (i) is not distinctly found on IAEs and (ii) can often be explained based on alternative theoretical accounts. We discuss here three implicit-as-associative views: implicit-

as-automatically-acquired-associations, implicit-as-slowly-acquired-associations, and implicit-as-unqualified-associations.

III.3.b.1. Implicit-as-automatically-acquired-associations

It is often assumed that IAEs are established through association formation thought to occur automatically (e.g., Gawronski & Bodenhausen, 2014). There is, however, little evidence for the automatic acquisition of attitudes, and when there is, this evidence is most convincingly obtained on self-reported evaluations, not IAEs. For instance, results from the surveillance paradigm developed by Olson and Fazio (2001) suggest that attitudes may be formed under incidental learning conditions; that is, unintentionally and under conditions of low subjective awareness. However, this study relied on self-reports, not IAEs.

Generally, the evidence is weak that attitudes can be formed under suboptimal conditions (for a review, see Corneille & Stahl, 2019). Regarding the unconsciousness feature of automaticity, recent studies that relied on strong designs, measures, and analyses have systematically failed to observe subliminal evaluative learning effects, using procedures as diverse as brief CS presentations (e.g., Stahl Haaf, & Corneille 2016), parafoveal CS presentations (Dedonder, Corneille, Bertinchamps, & Yzerbyt, 2014), or continuous flash suppression of the CS (Hödgen, Hütter, & Unkelbach, 2018). This null effect was not just observed in the evaluative conditioning paradigm, but also the approach-avoidance paradigm (Van Dessel, De Houwer, Roets, & Gast, 2016). It was also observed on measures of subjective awareness (Stahl & Bading, 2019) and on IAEs such as the Affect Misattribution Procedure (AMP; Hödgen et al., 2018) and the IAT (Heycke, Aust, & Stahl, 2017; Van Dessel et al., 2016).

This is not to say that unconscious processes play no role in evaluations. Rather, this is to stress that the most relevant paradigm for studying associative *learning* (simple, and associative in its structure) has so far failed to provide compelling support for an unconscious

acquisition of attitudes. Moreover, when evidence supports its possibility, it is best indicated on self-reported evaluations, not IAEs. This is consistent with classic social psychological work supportive of the role of unconscious processes in self-reported judgments and evaluations (e.g., Nisbett & Wilson, 1977; Schwarz & Clore, 1983, Strack, Martin & Stepper, 1988; Wilson, Dunn, Bybee, Hyman & Rotondo, 1984).

Turning to efficiency, research indicates that a concurrent task at the time of learning impairs attitude acquisition as measured by self-reported evaluations (Dedonder, Corneille, Yzerbyt, & Kuppens, 2010; Field & Moore, 2005; Mierop, Hütter, & Corneille, 2017; Pleyers, Corneille, Yzerbyt, & Luminet, 2009). This effect of load has been demonstrated in the conditioning of both familiar (Pleyers et al., 2009) and unfamiliar stimuli (Dedonder et al., 2010). It has also been evidenced in different sensory modalities (e.g., flavor-flavor conditioning) and on behavioral indicators of evaluation (drink pick-up latency: Davies, El-Deredy, Zandstra, & Blanchette, 2012). Finally, a recent study indicates a lack of efficient evaluative conditioning effects, even when the individual conditioned and unconditioned stimuli (USs) are correctly encoded. This suggests that the encoding of CS-US *pairs* in explicit memory is a critical but non-efficient process (Mierop, Maurage, & Corneille, in press).

Turning to controllability, Gawronski, Balas, and Creighton (2014) employed a self-report measure and an evaluative priming measure after an evaluative conditioning paradigm that asked participants to promote or prevent the influence of the USs on the evaluation of the CSs. They showed that effects on self-reported evaluations were controllable, while effects on the evaluative priming measure were uncontrollable. Using processing tree modeling, however, Hütter and Sweldens (2018) collected evidence that uncontrollable learning processes may contribute to attitude acquisition even when assessed with self-reported evaluations. Therefore, uncontrollability as a feature of automatic learning is not specifically

evidenced on IAEs (see also Section V). In addition, a recent replication of the evidence

obtained by Hütter and Sweldens (2018) in an instruction-based procedure challenges an

interpretation of uncontrollable learning in terms of association formation (Corneille, Mierop,

Stahl, & Hütter, 2019).

In sum, although there is some indication that attitudes may be acquired automatically,

current evidence suggests that IAEs do not offer privileged access to automatic learning

effects and that effects can be explained based on a non-associative learning theories.

III.3.b.2. Implicit-as-slowly-acquired-associations

Many dual-process theories assume that IAEs originate from the direct experience of

events, passively registered by a slowly and incrementally learning System 1 (e.g., McConnell

& Rydell, 2014; Petty, Tormala, Briñol, & Jarvis, 2006; Rudman, 2004; Rydell & McConnell,

2006, Smith & DeCoster, 2000; Strack & Deutsch, 2004; Wilson, Lindsay & Schooler, 2000).

Consider this quotation:

"Implicit attitudes and stereotypes may be acquired over many years from language

and social experiences that cumulatively construct an overlearned repertoire of cultural

expertise rooted in even more thousands of hours of experience than those invested by

medical doctors, virtuoso musicians, or world-class athletes in their professional training. This

unsought cultural expertise, condensed into stereotypes and attitudes, may be as difficult or

impossible to unlearn as are musical, medical, and athletic expertise." (Greenwald & Banaji,

2017, pp. 866-867).

According to the implicit-as-indirect conceptualization supported by these authors,

however, there is no cogent reason why IAEs should reflect the way in which attitudes were

acquired. Turning to the implicit-as-associative conceptualization, the problem is primarily

empirical. Specifically, contrary to the slow-paced view, several studies have demonstrated

that IAEs can be acquired quickly and changed based on the communication of symbolic

information. For instance, providing participants with mere instructions about the nature of stimulus pairings in an evaluative conditioning paradigm, in the absence of experience of the pairings, impacts responses on the IAT (De Houwer, 2006). IAEs are also sensitive to procedures in which participants experience a US and infer only later that a CS was present as well (Gast & De Houwer, 2012). Similar instruction-based effects have been documented for the mere exposure paradigm (Van Dessel, Mertens, Smith, & De Houwer, 2017) and approach-avoidance training (Van Dessel, De Houwer, Gast, & Smith, 2015). Evaluative statements have also recently been shown to be *more* effective than the experience of co-occurrences at creating IAEs also in children, who are thought to be more sensitive to low-level learning (Charlesworth et al., 2019).

As to resistance to change, recent studies show that IAT scores (e.g., a positive attitude towards Gandhi) are quickly reversed when inconsistent declarative information is provided about the attitude object (e.g., when participants are told that Gandhi refused medical treatment to his wife, who died from this lack of treatment, whereas Gandhi claimed it when he fell ill; Van Dessel, Ye, & De Houwer, 2018). IAEs also change when prejudice-consistent statements are negated in a short training phase prior to an evaluative priming paradigm (Johnson, Kopp, & Petty, 2018; see also Gawronski, Deutsch, Mbirkou, Seibt, & Strack, 2008) or when new verbal information is provided that elicits a reinterpretation of previous information (Mann & Ferguson, 2015; for a comprehensive discussion of quick changes in IAEs, see Cone, Mann, & Ferguson, 2017). More generally, IAEs are largely dependent on information activated in the immediate context of judgment (e.g., Blair, 2002; Gawronski & Sritharan, 2010; Wittenbrink, Judd, & Park, 2001), depend on currently activated goals (Melnikoff & Bailey, 2017), and show low test-retest reliability when measured on the individual level (e.g., Gawronski, Morrison, Phills, & Galdi, 2017; Payne et al., 2017).

III.3.b.3. Implicit-as-unqualified-associations

Imagine an aversive sound whose termination is signaled by the appearance of a cartoon character. Dual-process theories distinguish between a fast, declarative learning mechanism that infers the positivity of the character and a slow-paced learning mechanism that passively registers the mere and unqualified co-occurrence of the cartoon character with a negative stimulus. While several articles document seeming dissociations between self-reported evaluations and IAEs (Hu, Gawronski, & Balas, 2017; Moran & Bar-Anan, 2013), recent evidence shows that IAEs reflect an influence of relational qualifiers. Van Dessel, De Houwer, and Smith (2018) found that IAT scores are sensitive to approach-avoidance instructions that vary agency (i.e., self-agent versus stimulus-agent instructions). Likewise, IAT scores and evaluative priming scores vary as a function of the relation between CSs and USs in an evaluative conditioning paradigm (Zanon, De Houwer, & Gast, 2012). Moreover, recent research shows that the results of an IAT strongly depend on the target categories contrasted in that measure (Bading, Stahl, & Rothermund, 2020).

In sum, a growing body of research demonstrates that IAEs are sensitive to the same learning factors that influence self-reported evaluations. This does not mean that systematic differences cannot be observed between tasks that vary in the level of automaticity they involve. In particular, evaluative outcomes will likely differ when respondents report a speeded evaluation or introspect at length on the reasons for their evaluation. However, these differences need not be explained based on an associative learning theory, nor do they require a reference to the "implicit" construct.

More generally, it is also important to distinguish between the associative learning assumptions versus associative expression assumptions of dual-process theories of attitudes (see Hütter & Rothermund, 2020, for an integrative framework). For instance, the APE model assumes that IAEs result from retrieval processes that reflect the mere activation of memory

content without validation of this activated content (e.g., Gawronski & Bodenhausen, 2006, 2011). While automatic evaluations are thought to be observed under suboptimal conditions (e.g., limited time and resources), self-reported evaluations require more favorable conditions for deliberative processing. These expression assumptions may be valid, while the learning assumptions may not be (for a discussion, see Corneille & Stahl, 2019), resulting in a lack of dissociative evidence between learning mode and expression mode. The MODE model (e.g., Fazio, 2007) makes related expression assumptions while entirely disregarding the "implicit" construct.

III.3.c. Conclusions for the implicit-as-associative conceptualization

The associative conceptualization of "implicit" has failed to receive conclusive empirical support in virtually all its stated assumptions. Specifically, (i) evidence for automatic learning is not best obtained on IAEs, (ii) IAEs are not particularly sensitive to a slow form of learning, nor (iii) to the learning of unqualified relations. The reason why we recommend discontinuing the usage of the "implicit" terminology under this third conceptualization is that, in addition to creating confusions with concurrent conceptualizations that are not associative in nature, "implicit" relates here to a set of strongly debated associative learning assumptions (Corneille & Stahl, 2019). We also encourage researchers to reflect on whether association formation theories are best suited for their research question relative to, for instance, propositional theories.

III.4. Hybrid definitions

III.4.a. Implicit as indirect and automatic

One regularly comes across a hybrid definition that restricts implicit attitudes and evaluations to measures that are indirect *and* characterized by features of automaticity. This may be illustrated with Nosek et al. (2011): "The label 'implicit' is applied to a variety of procedures and processes that share a common theme: they are not direct, deliberate,

controlled, and intentional self-assessments." (p. 153). This definition circumvents some issues inherent to the first two conceptualizations. It is more constraining than the "implicit-as-automatic" definition as it allows for indirect measurement only. This is captured in the following statement: "The signature feature of implicit measurement procedures is that they assess mental content indirectly" (Nosek et al., p. 153). It is, however, more process-loaded than the "implicit-as-indirect" definition (i.e., it is not purely procedural).

It is unclear, however, what "implicit measurement procedures" indicate. Schimmack (2019) rightly claims that the IAT is a method in search of a construct. But this critique readily applies to many other tasks falling into the "implicit" category. Nosek et al. (2011) conceded almost a decade ago: "Knowledge about what implicit measures measure is less mature than knowledge about what they do." (p. 156). The authors point out that measures of this sort reduce conscious control and that the "indirect assessment relieves the requirement that the respondent is able to report the associated mental content." (p. 153). These measures are also thought to "assesses mental content without requiring awareness of the relation between the response and the measured content" (p. 153). However, and importantly, these measures also do not *preclude* awareness so that IAEs cannot readily be interpreted as unaware.

III.4.b. Implicit as driven by affective gut reactions

Finally, some researchers define implicit evaluations as responses driven by affective gut reactions. This definition is hybrid in that it combines mental and physiological assumptions. For instance, Gawronski and Bodenhausen (2011) note: "Applied to the distinction between implicit and explicit evaluations, we argue that the overall valence of the concepts that are activated in response to a given object determines the evaluative quality of an individual's affective gut reaction to that object, which in turn drives responses on measures of implicit evaluations" (p. 62).

Although we do not believe it was Gawronski and Bodenhausen's (2011) intention, the reference to "affective gut reactions" suggests that IAEs may be distinctly anchored in physiological reactions. Tracing back to the pioneering work by James (1884), "gut reactions" classically refers to bodily grounded emotions. Hence, linking explicit evaluations to higher-order inferential processes that characterize high-order propositional thinking, as opposed to implicit evaluations grounded in "gut reactions", suggests that the latter may be more bodily grounded; typically in visceral states, as proposed in classic (Lange, 1885; Schachter & Singer, 1962) and more contemporary (Barrett, 2006, Barrett & Lindquist, 2008; Critchley & Nagai, 2012; Damasio, 1994, 1999; Prinz, 2014, 2012) theories of emotions. Attaching the qualifier "affective" to "gut reaction" further suggests a dissociation between high-order cognitive versus more bodily grounded or physiological processes. The AMP (Payne, Cheng, Govorun, & Stewart, 2005) is a typical instance of an evaluative task that subscribes to the implicit-as-driven-by-affective-gut-reactions as it assumes affect as a driving force (for a semantic account, see Blaison, Imhoff, Hühnel, Hess, & Banse, 2012).

That "gut reactions" may be understood not just in the colloquial sense of "spontaneous" but in a more bodily grounded sense is evident when considering actual manipulations of the "gut feeling" construct in implicit social cognition research. For instance, Lee, Lindquist, and Payne (2018) recently used the following instructions for examining the link between fear and racial biases: "*My gut feelings toward Blacks reflect fear*," "*My gut feelings toward Blacks reflect anxiety.*" This indicates that reference to "gut feelings" in attitude research does not unambiguously refer to the colloquial usage of this concept.

Consistent with the opening quotation stating that gut-reactions *drive* (and so are conceptually extrinsic to) IAEs, physiological measures are currently used to study the bodily substrates of evaluations (e.g., Lieberman, 2007; Stanley, Phelps, & Banaji, 2018). They are

the *explanans* rather than the *explanandum*. Whereas they were originally considered indirect measures of evaluations (e.g., Fabrigar, Krosnick, & MacDougall, 2005), more recent research suggests that the interpretation of physiological measures as indicators of evaluations is mitigated (e.g., Cunningham, Packer, Kesek, & Van Bavel, 2009) and that they await further psychometric validation (e.g., scaling: see e.g., Blanton & Jaccard, 2015). As a result, physiological measures are excluded from recent methodological chapters on implicit social cognition (e.g., Gawronski & De Houwer, 2014) and recent research method sections in social and affective neuroscience are agnostic to the "implicit attitudes and evaluations" constructs (e.g., Berkman, Cunningham, & Lieberman, 2014).

Nevertheless, linking behavioral evaluative measures to their physiological substrates is an important research endeavor. It is indisputable that behavioral responses require neurophysiological activity. It is theoretically unclear, however, why distinct neurophysiological activity should underly self-reported evaluations on the one hand and IAEs on the other hand. As a result, it may be questioned whether the neuroscience of "implicit prejudice" (Amodio, 2014) is warranted. Current research efforts are aimed at clarifying this issue in the context of the IAT (e.g., Marini, Banaji, & Pascual-Leone, 2018), and it remains to be seen whether results generalize to other IAEs.

Finally, it should be noted that self-reported evaluations are sensitive to gut reactions, too. Gawronski and Bodenhausen (2011) add: "In many cases, people may use their affective gut reaction to an object as a basis for an endorsed evaluative judgment about that object, such that they may simply report the evaluative quality of their gut response on measures of *explicit evaluations*." (Gawronski & Bodenhausen, 2011, p. 62). Hence, the absence or presence of "gut reactions" *per se* is irrelevant to the characterization of a response as indicating IAEs. The notion that affective reactions offer a basis for self-reported evaluations

is supported by both classical (e.g., Crites, Fabrigar, & Petty, 1994) and more recent (e.g., Rocklage & Fazio, 2016) research.

III.4.c. Conclusions for hybrid definitions

We recommend refraining from using the construct of "gut reactions" in conceptualizing the "implicit" construct in attitude research. This is because "gut-reactions" can be understood in a colloquial sense (i.e., spontaneous) or in a sense that ties this concept to research on emotions, which assumes that affective reactions are grounded in visceral states. We also recommend not using the implicit-as-indirect-and-automatic conceptualization, as it prevents a coherent interpretation of task outcomes. It is a pragmatic "(…) 'lumping' strategy that prioritizes the rapid accumulation and comparison of evidence" (Nosek, Hawkins, & Frazier, 2011, p. 153), without allowing a clear interpretation of this evidence.

IV. Current ambiguities in the definition and measurement of the implicitness construct hamper progress in attitude research

Conceptual confusions between the nature of a task, the nature of the processes driving responses on that task, the nature of the evaluations captured by that task, and the nature of the attitude underlying evaluations pave the way for misunderstandings and unwarranted inferences in attitude research. For instance, in an influential article, Dovidio and Gaertner (2004) interpreted IAT outcomes as indicating that "(…) the vast majority of white Americans harbor unconscious negative associations about blacks" (Dovidio & Gaertner 2004, p. 20). Such overgeneralizations are pervading theory and practice well beyond psychological research. To illustrate, the company Pizza Hut recently launched an eye-tracking measure to uncover unconscious preferences for pizza topping in their consumers (Henderson, 2014). Based on previous research, however, the conclusion that IAT or eye-tracking measures indicate unconscious representations is unjustified.

This sort of confusion has important implications for designing social interventions (for a recent analysis of common pitfalls in interpreting implicit biases, see Gawronski, 2019). The IAT is regularly used in the context of social intervention on unconscious biases, for example, in addressing gender inequalities in STEM subjects (e.g., Carnes et al. 2015; Devine et al. 2017). Consistent with the assumptions discussed earlier, the rationale is that the IAT reveals unconsciously held knowledge that has slowly accumulated through exposure to a structurally biased environment. If relying on these assumptions, an intervention may miss important points; for instance; that the bias may be conscious (e.g., Gawronski, 2019; Gawronski et al., 2006; Hahn et al., 2014), or changed quickly (e.g., Cone et al., 2017), or that gender imbalance in STEMs actually increases with national gender equality – in our current example, with much smaller proportions of females in STEM subjects actually observed in countries such as Finland and Norway than Algeria and the United Arab Emirates (Stoet & Geary, 2018). Relying on faulty assumptions conveyed by the sort of conceptual confusions reviewed here can distract practitioners from more effective intervention programs.

As a second illustration, we may consider theorization in personality psychology positing conflicts arising from consciously and unconsciously held attitudes of opposite valences about the self: "Whereas explicit self-esteem is often defined as conscious feelings of self-liking, self-worth, and acceptance (e.g., Brown, 1993; Kernis, 2003; Rosenberg, 1965), implicit self-esteem is typically believed to consist of nonconscious, automatic, and overlearned self-evaluations (Greenwald & Banaji, 1995; Pelham & Hetts, 1999)." (Zeigler-Hill, 2006, p. 120). An important question here is how clinical psychologists can adequately treat patients who are presumably torn between conflicting self-views. Does it make sense to counteract patients' unconscious negative self-view by extensively training them in self-positive associations? If so, should this training be implemented in a slow-paced way?

On a more theoretical level, a focus on mental processes as an explanation for differences between measures may distract researchers from asking questions about the nature of the measures or elicitation techniques that could equally be responsible for these differences. Let us discuss one recent example to illustrate this point. Moran and Bar-Anan (2013) introduced a relational evaluative conditioning paradigm. That is, the CSs were paired with pleasant or unpleasant sounds. However, some CSs started those sounds (i.e., they were played before the onset of the US) while others stopped them (i.e., they were played before the offset of the US). The self-report measure demonstrated an interaction effect: the influence of the sounds on the CS evaluation was qualified by the relationship between the CS and the sounds. In an IAT, however, only a main effect of the pleasantness of the sounds was observed. This is consistent with the implicit-as-unqualified-associations conceptualization. Hu et al. (2017) noted that this demonstration was to date the "most compelling evidence for dual-process accounts" (p. 19).

Moran and Bar-Anan (2013), however, warned the reader: "Perhaps an effect of relational information on automatic evaluation could be detected with other measures or with other comparisons." (p. 750). Bading et al. (2020) proceeded to that examination. In a series of experiments, these authors showed that the IAT reflects any distinction between CSs that helps participants solve the task more efficiently. For instance, when the response options opposed CSs that were paired positively with CSs that were paired negatively, the IAT score reflected this distinction as in the original work by Moran and Bar-Anan (2013). However, when participants had to sort CSs that started a sound and CSs that stopped a sound, the IAT score reflected this distinction. Thus, the IAT indicated that participants learned the starting versus stopping relation. And, finally, when participants categorized CSs that had a positive meaning (i.e., they started a positive sound or stopped a negative sound) using one key and

CSs that had a negative meaning (i.e., they started a negative sound or stopped a positive sound) using the other key, the IAT score also reflected this particular distinction.

The conclusion from this example is that if researchers do not consider the influence of the measurement procedure on the measurement results, they can miss important theoretical insights that, in the case of Bading et al. (2019), (1) demonstrate the great flexibility of the human mind, (2) relate attitude scores to ecological explanations rather than intrapsychic ones (Fiedler, 2014), (3) challenge the notion that IAEs indicate unqualified learning of co-occurrences.

V.   Processing tree models

Multiple processes are driving responses on any given task. In many regards, then, comparing outcomes from an IAT, an AMP, or a self-report task is like comparing apples and oranges. By quantifying the unique contribution of different processes to task performance, processing tree modeling allows for a more precise and accurate study of evaluative phenomena. In this section, we discuss first how processing tree models can help achieve a finer-grained conceptualization and measurement of the mental processes involved in evaluative responses. Then, we discuss how these models may help address attitude formation questions.

V.1. Processing tree models as precision-enhancement tools for conceptualization and measurement

Recently, researchers have proposed to minimize issues arising from the comparison of outcomes between distinct evaluative tasks by developing variants of the same task that keep the general structure constant. For example, Payne, Burkley, and Stokes (2008; see also Van Dessel, Cone, Gast, & De Houwer, 2020) have designed two variants of the affect misattribution procedure that differ in whether participants are explicitly asked to rate the racial prime. Hence, the tasks are supposed to differ only in whether they assess relatively

intentional versus unintentional evaluation of the prime. However, such variants may still differ in multiple ways. For instance, the degree of attention that participants pay to the prime will differ and also the degree to which they are motivated to respond in a socially desirable way (cf. Payne et al., 2008, Study 4). The ultimate goal should thus be to keep the task perfectly constant while separating the different processes contributing to performance on that task. Processing tree modeling can achieve that goal.

Processing tree models are a research framework whose influence on attitude research has increased in recent years (for a review, see Hütter & Klauer, 2016). Based on the notion that no measure is process-pure, this approach disentangles the contribution of several processes to performance in a given task. Let us consider the congruent block of an IAT, in which positive (negative) targets share a key with positive (negative) attributes. Deliberate processes such as detecting the correct response and non-deliberate processes such as the automatic activation of an attitude should both contribute to correct responses in this block. In the incongruent block, however, deliberate processes facilitate correct responses, while activated attitudes facilitate incorrect responses. Based on this rationale, Conrey, Sherman, Gawronski, Hugenberg, and Groom (2005) specified a processing tree model for the IAT and other indirect measures. The quadruple process ("Quad") model considers four processes that contribute to performance on these measures (hence its name): association activation (AC), stimulus discrimination (D), overcoming bias (OB) and guessing (G; actually, a collection of processes that lead to unsystematic responses). While the AC parameter is conceptualized as an automatic process, the D and OB parameters are conceptualized as deliberate processes. Hence, from performance in a single task this model quantifies processes that vary in their relevance to the constructs of evaluation and automaticity.

Conrey and colleagues (2005) showed that the Quad model offers a suitable account of responses on an IAT and other evaluative tasks (see Sherman et al., 2008, for an overview;

but see also Payne & Bishara, 2009). An alternative conceptualization of IAT performance is laid out in Meissner and Rothermund's (2013) ReAL model. This model also separates association activation (*A*) from other processes contributing to task performance (recoding *Re* and label-based discrimination *L*) and has proven useful in a series of validation experiments. As illustrated by the two different accounts of IAT performance laid out in the Quad and ReAL models, however, the mere fit of these models is not necessarily informative on the validity of the conceptualization. Each parameter requires validation in order to be assigned psychological meaning (see Hütter & Klauer, 2016).

V.2. Processing tree models as precision-enhancement tools for investigating learning effects

Processing tree models are also useful for testing learning assumptions in attitude research. Challenging the implicit-as-associative view, research that relied on these procedures indicates that evidence for automatic attitude learning (i) is found on self-reported evaluations, (ii) is found in learning paradigms that provide symbolic information about attitude objects, and it indicates that evidence for unqualified attitude learning (iii) does not seem to originate from slow-paced association formation.

Relevant to the "implicit-as automatically-acquired" conceptualization, Hütter et al. (2012) developed a multinomial processing tree model that dissociates memory-based learning from memory-independent learning (i.e., evaluative learning in the absence of memory for the source of this evaluation). This model was validated in the context of an evaluative conditioning paradigm and supported the contribution of unconscious (i.e., memory-independent; but see Gawronski & Walther, 2012; Mierop et al., 2017) processes to self-reported evaluations. This evidence is inconsistent with the view that IAEs have a special relation to associative learning, when defining it as an automatic (here: unconscious) learning process.

Furthermore, the question remained whether this finding required the experience of the stimulus pairings, or whether it could also be established based on a mere verbal description of the pairings. Hütter and De Houwer (2017) examined this question in the context of an instruction-based learning paradigm. Challenging an association formation account (De Houwer & Hughes, 2016), their studies demonstrated that instructed procedures allow for memory-independent learning.

Similar findings were recently obtained when considering another criterion of automaticity: controllability. Hütter and Sweldens (2018) found evidence for the contribution of uncontrollable attitude formation on self-reported evaluations. Again, however, this finding was replicated in a learning procedure in which participants were merely instructed about the task and never watched actual stimulus pairings (Corneille, Mierop, Stahl, & Hütter, 2019).

Smith, Calanchini, Hughes, Van Dessel, and De Houwer (2020) compared Quad model parameters between an instructed and an experienced variant of three central evaluative learning paradigms: mere exposure, evaluative conditioning, and approach-avoidance training. The Activation parameter was similarly influenced by experience-based procedures and instruction-based procedures in all paradigms. Thus, whether a learning procedure was completed versus merely instructed does not provide information on the nature of the processes involved in the task.

Turning to the "implicit-as-unqualified-associations" conceptualization, Kukken, Hütter, and Holland (2020) developed a processing tree model to separate effects of the implication of a pairing (e.g., stopping an aversive sound has a positive implication) from effects of the pairing (e.g., the co-occurrence with an aversive sound irrespective of its relationship with the CS) on evaluative self-reports. The parameters for attitude acquisition based on meaning versus pairing were dissociable in validation studies and consistently larger

than zero. Hence, self-reported evaluations, not just IAEs, can indicate learning of unqualified relations.

Furthermore, Heycke and Gawronski (2019) demonstrated that the effect of the pairing in a relational paradigm was unrelated to the number of repetitions of the CS-US pairs, while the effect of the relation between US and CS increased with an increasing number of repetitions. This pattern, again obtained on self-reported evaluations, challenges the notion that association formation, if defined as an unqualified registration of stimulus co-occurrences, corresponds to a slow-paced process (even though the results can be accommodated when introducing additional assumptions; see Heycke & Gawronski, 2019).

V.3. Limitations of processing tree models

While we recommend processing tree models as precision enhancement tools, it is important to note that they come with their own set of assumptions that require careful validation work (e.g., Hütter & Klauer, 2016; Klauer, 2006; Klauer, Dittrich, Scholtes, & Voss, 2015). They also often require more complex research designs that sometimes appear artificial (e.g., Hütter et al., 2012). Moreover, virtually all of our criticism also applies to the interpretation of model parameters. For instance, they should not be taken as indicators of automatic processes unless they have been validated using manipulations of automaticity features (e.g., Hütter & Sweldens, 2018). Researchers should be precise which automaticity feature(s) they refer to when they label a parameter "automatic." The parameters also should not be assumed to constitute measures of associative processes without a clear reflection on what "associative" means and whether other learning theories could also explain the existence of a parameter or variations of that parameter (e.g., Corneille et al., 2019).

V.4. Conclusions for processing tree modeling

The processing tree modeling framework encourages a much more precise conceptualization on mental processes and their measurement by overcoming problems

inherent to task comparisons. It also allows gaining insights into attitude formation questions. We strongly recommend, however, running validation studies when a processing tree approach is employed (cf. Hütter & Klauer, 2016).

VI.   General Discussion

The implicit construct is conceptualized on three levels in attitude research, the procedural level (i.e., implicit-as-indirect), the functional level (i.e., implicit-as-automatic), and the mental theory level (i.e., implicit-as-associative). We additionally discussed two hybrid definitions. We offered a critical review of limitations for each level and argued that the co-existence of conflicting definitions hampers communication, theory, and practice. In this General Discussion, we discuss points of attention for future research.

VI.1. Recommendations *if* using the "implicit" terminology

For the many reasons discussed in this review, we recommend discontinuing the usage of the "implicit" terminology in attitude research and in research inspired by it. As we have argued above, the term is not only delusive but also unnecessary. The procedural definition is misleading in its process-loaded version and is theoretically irrelevant in its process-free version. As to the functional definition, "automatic" should be preferred to "implicit" as it would reduce the risk of confusions with concurrent conceptualizations of the "implicit" construct and prevent that an undue weight is given to the unconscious feature of automaticity. Finally, the mental theory definition endorses associative learning assumptions that have been challenged in recent research.

At the very least, if researchers wish to keep using the "implicit" terminology, we think it is important that they actively prevent conceptual and interpretational confusion that this choice entails. This can be achieved by systematically (i) stating one's definition of the construct, (ii) explaining the reasons underlying task selection, and (iii) proceeding to a cautious interpretation of task outcomes that is consistent with the conceptualization and

rationale of the task. This recommendation goes against that from prominent authors in the field who promote conceptual blurring in the use of the construct. For instance, Greenwald and Lai (2019) mentioned that their readers "(…) should be unencumbered by a need to choose among definitions of "implicit." (p. 25.4).

We would like to repeat here that often the "implicitness" terminology is not just misleading, but also unnecessary. The MODE model, for instance, set the ground for important theoretical and methodological developments while entirely disregarding the "implicit" attitude terminology. More generally, evaluative measures developed in social cognition research over the last two decades are useful for advancing theoretical knowledge. However, using these tasks should not systematically imply using the "implicit" terminology. For instance, Berger, Hütter, and Corneille (2019) recently used sequential priming procedures to examine whether ambivalent attitudes are characterized by the joint and unintentional activation of opposite valences. This could not be done by using self-reported evaluations as people likely sequentially activate and intentionally retrieve different types of knowledge when completing items assessing their positive and negative attitudes towards the attitude object separately. At no point, however, did the authors need to rely on the "implicit" terminology for investigating their research questions or for reporting their results.

VI.2. IAEs cannot be distinctly induced. This is consequential.

A critical consequence of not being able to create IAEs distinctly is that research on "implicit attitudes and evaluations" remains correlational. The most conclusive evidence, however, generally stems from an experimental approach to psychological questions. If seeking strong theoretical and empirical advances, future research relying on the "implicitness" terminology should be able to create rather than merely measure IAEs. More specifically, IAEs should be induced experimentally on a variety of tasks (not just the IAT, if one seeks to go beyond IAT research) in the absence of similar evaluative changes in self-

reported evaluations (and vice versa). As we stressed in Section V, creating *within* a task disjunctive patterns on processing tree parameters would be even more convincing.

If aiming at testing dual-learning theories, additional requirements should be met. First, dissociative patterns should not allow for concurrent explanations by a non-learning account (e.g., an episodic memory account should be ruled out; see e.g., Stahl & Aust, 2018). Second, they should be immune to alternative, non-associative learning explanations (e.g., a propositional learning account; Corneille et al., 2019; Hütter & De Houwer, 2017).

If this research were to succeed, but experimentally induced changes in IAEs would prove inconsequential for behavior (and there is evidence supporting this possibility; Forscher et al., 2017; Lai et al., 2014), then questions would arise concerning the practical significance of IAEs. By extension, the significance of social intervention programs inspired by an implicit-as-associative conceptualization (e.g., slow-paced learning, experiential learning, unintentional learning, subliminal learning) should also be questioned.

VI.3. Potential criticisms of our recommendations

One may object that research on IAEs has been thriving despite high conceptual heterogeneity, and that all that matters in science is how a concept is operationalized, not how much consensus exists in its definition. We respectfully disagree with this view. First, if research on IAEs has been thriving, it has also carried on important overgeneralizations and misinterpretations that could have been avoided through enhanced conceptual clarity. Second, conceptual imprecision likely leads to measurement imprecision, which is detrimental to research advances. Third, conceptual imprecision can feed conceptual replication issues if researchers wrongly assume that two different tasks (e.g., an IAT and an eye-tracking measure) tackle a common evaluative phenomenon. Fourth, and complementarily, it may also feed confirmatory biases, such as not seeking replication on a task merely because it does or does not share the same label as another task. Fifth, as illustrated by many examples in this

review, enhanced conceptual precision encourages researchers to test new research questions that can provide important insights into mental processes. Sixth, how thriving that research has been should be somewhat toned down when (i) considering that theoretical assumptions underlying the (associative) use of this terminology have received little support by empirical research (see Section III), and (ii) considering current uncertainties about the practical significance of IAEs, in particular when it comes to designing diversity or implicit bias training (e.g., Greenwald & Lai, 2019). IAEs may occasionally improve behavioral prediction (for recent discussions, see Kurdi et al., 2019; Meissner et al., 2019). Yet, this does not mean they are causal factors, nor that reliance on an "implicit" conceptualization is required for explaining their predictive value.

The reader may also object that concepts are inherently fuzzy and that efficient scientific communication requires conceptual compromises. We fully agree with this view and are not recommending that the level of conceptualization becomes fine-grained to such level of detail that researchers cannot speak to each other anymore beyond their micro-conceptual niches. We are convinced, however, that there is room for enhancement in terminological usage. For instance, if endorsing the "implicit-as-automatic" view, it does not seem unrealistic to expect researchers to systematically state what feature of automaticity is relevant to their attitude research question, and to select a task and interpret scores on this task accordingly. This is often not done despite earlier calls for enhanced precision (De Houwer et al., 2009). Turning to the "automatic" terminology would also prevent confusion that may arise from concurrent conceptualizations of the "implicit" construct.

VI.4. Relation to single and dual-process theories of the mind

It should be evident at this point that the explicit-implicit opposition in attitude research has been largely inspired by dual-process theories that contrast thinking and behavior under optimal versus suboptimal conditions. We would like to make it clear that, just because

we take a stance against the "implicit" terminology in attitude research does not mean that we discourage research into automatic processes, including unconscious ones. We believe, however, that the investigation of social cognition and behavior under suboptimal conditions is not helped by relying on the "implicitness" construct. Researchers should also carefully consider which features of automaticity are of primary interest for addressing their research questions, as well as the processing stages and mental contents they apply to.

Not only should unconsciousness not be conflated with other features of automaticity, but it should also be related to a specific mental content that arises at a specific stage. For instance, unconsciousness may concern the source of an attitude, the impact this attitude has on one's behavior (Bargh, 1992b), the existence of the attitude, the process through which the attitude is formed, or the awareness that this attitude is measured (Sweldens, Tuk, & Hütter, 2017). We have discussed the utility of processing tree models in this context: these models allow identifying and separating the contribution of discrete mental processes (including "associations") to task outcomes. We acknowledged and discussed why processing tree models are not a panacea. At the very least, however, they constrain researchers to conceptualize their research question much more precisely.

More generally, given the current theoretical and empirical state of the field, we would like to encourage researchers to consider the explanatory value of single-process theories such as propositional (e.g., De Houwer et al., 2019) or episodic memory theories (Stahl & Aust, 2018) that do not distinguish between Type-1 and Type-2 processes.

VI.5. The « implicitness » terminology promotes connectedness at the expense of accuracy

The implicitness terminology certainly has some value. It unites a diversity of research efforts interested in the formation and expression of evaluations under suboptimal conditions. In doing so, it connects attitude research to the even broader dual-process theories of the

mind. These goals are important. However, reliance on this terminology also causes a false sense of understanding. In the end, one cannot be sure we are all speaking the same language. Are implicit evaluations meant to be indirectly assessed, automatic, associative, or unconscious? Does "implicit" refer to representations, processes, measures, or responses? Are "implicit measures" meant to capture affective reactions? Should a specific measure be considered implicit or not, and why?

While relaxed constructs help unite and connect research efforts and are useful for gaining people's attention to unfamiliar and complicated matters, it is not well-suited for seeking accurate knowledge. There is disagreement on how implicit attitudes and evaluations should be defined (therefore preventing a shared understanding of what they are) and measured (therefore preventing the establishment of accurate knowledge). Critically too, IAEs are assumed but have not been shown to have a distinct learning basis. The nature, measurement, and induction of "implicit attitudes and evaluations" await further and finer conceptual and empirical analysis. We have hardly understood what we measure with different attitudinal measures and why they often diverge (e.g., Blanton & Jaccard, 2015; Schimmack, 2019). Implicitness claims typically stem from dissociations between measures that may have diverse sources such as the structure of the task, differences in reliability, or the lack of conceptual correspondence (e.g., Friese, Hofmann, & Schmidt, 2009; Gawronski, 2019). Given this perplexing state of affairs, it is unsurprising that "theorizing about implicit cognition is relatively unsophisticated at this time." (Kurdi et al., 2019, p. 14).

The "implicitness" terminology can also enhance connectedness on a more political level. As Banks and Fords (2009) explained when discussing research on unconscious biases: "The invocation of unconscious bias levels neither accusation nor blame, so much as it identifies a quasi-medical ailment that distorts thinking and behavior. (…) The unconscious bias claim thus facilitates a consensus that the race problem persists." (Banks & Fords, 2009,

p.1054, parenthesis added). The "implicit" conceptualization (coined in the present example in its "unconscious bias" version, but the same analysis would easily apply to other features of automaticity such as unintentionality or uncontrollability) serves a commendable social function. However, political goals may occasionally conflict with scientific advances. Furthermore, as the same authors argue, it is debatable whether IAEs research ultimately best serves social changes: "(…) the rhetoric of unconscious bias is so compelling that people are likely to accept it as the goal of racial reform and, consequently, to push the theory in directions that siphon energy away from problems of substantive inequality and that may be undesirable in their own right." (Banks & Fords, 2009, p. 1054).

VI.6. Conclusion

With the present review, we would like to encourage critical thinking about the implicitness construct and its measurement in attitude research and research inspired by it. We hope that, ultimately, the clarification and recommendations offered here will help our research community develop a univocal scientific language, advance theories and measurement, and set the stage for more effective intervention programs.

References

Amodio, D. M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, *15*(10), 670-682.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or ''Would Jesse Jackson 'fail' the Implicit Association Test?'' *Psychological Inquiry, 15*(4), 257-278. https://doi.org/10.1207/s15327965pli1504_01

Bading, K., Stahl, C., & Rothermund, K. (2020). Why a standard IAT effect cannot provide evidence for association formation: The role of similarity construction. *Cognition and Emotion, 34*(1), 128-143. doi: 10.1080/02699931.2019.1604322

Banaji, M. R., Lemm, K. M., & Carpenter, S. J. (2001). The social unconscious. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 134-158). Malden, MA: Blackwell Publishers, Inc.

Banks, R. R., & Ford, R. T. (2008). (How) Does Unconscious Bias Matter: law, Politics, and Racial Inequality. *Emory Law Journal*, *58*, 1053.

Bargh, J. A. (1992a). The ecology of automaticity: Toward establishing the conditions needed to produce automatic processing effects. *American Journal of Psychology*, *105*(2), 181-199.

Bargh, J. A. (1992b). Being unaware of the stimulus versus unaware of its effect: Does subliminality per se matter to social psychology. In R. F. Bornstein & T. S. Pittman (eds.), *Perception Without Awareness*. New York: Guilford Press. pp. 236-258.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. *Handbook of Social Cognition*, *1*, 1-40.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58.

Barrett, L. F., & Lindquist, K. A. (2008). The embodiment of emotion. *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*, 237–262.

Berkman, E. T., Cunningham, W. A., & Lieberman, M. D. (2014). Research methods in social and affective neuroscience. *Handbook of research methods in social and personality psychology*, *Handbook of research methods in social and personality psychology* (2nd ed., pp. 123–158). New York, NY, USA: Cambridge University Press.

Berger, K. T., Hütter, M., & Corneille, O. (2019). Investigating attitudinal ambivalence via sequential priming: Evidence for the simultaneous and unintentional activation of opposite evaluations. *Journal of Experimental Psychology: General.*

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242-261. https://doi.org/10.1207/S15327957PSPR0603_8

Blaison, C., Imhoff, R., Hühnel, I., Hess, U., & Banse, R. (2012). The affect misattribution procedure: Hot or not? *Emotion*, 12(2), 403. DOI: 10.1037/a0026907.

Blanton, H., & Jaccard, J. (2015). Not so fast: Ten challenges to importing implicit attitude measures to media psychology. *Media Psychology*, *18*(3), 338-369. https://doi.org/10.1080/15213269.2015.1008102

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. (2015). Toward a Meaningful Metric of Implicit Prejudice. *Journal of Applied Psychology, 100*(5), 1468-1481. https://ssrn.com/abstract=2564298

Brownstein, M., Madva, A., & Gawronski, B. (in press). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science*.

Carnes, M., Devine, P. G., Manwell, L. B., Byars-Winston, A., Fine, E., Ford, C. E., ... & Palta, M. (2015). Effect of an intervention to break the gender bias habit for faculty at one

institution: a cluster randomized, controlled trial. *Academic medicine: journal of the Association of American Medical Colleges*, *90*(2), 221.

Charlesworth, T. E., Kurdi, B., & Banaji, M. R. (2019). Children's implicit attitude acquisition: Evaluative statements succeed, repeated pairings fail. *Developmental Science*. https://doi.org/10.1111/desc.12911

Cone, J., Mann, T. C., & Ferguson, M. J. (2017). Changing our implicit minds: How, when, and why implicit evaluations can be rapidly revised. *Advances in Experimental Social Psychology, 56*, 131-199. https://doi.org/10.1016/bs.aesp.2017.03.001

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-model of implicit task performance. *Journal of Personality and Social Psychology, 89,* 469-487. https://doi.org/10.1037/0022-3514.89.4.469

Corneille, O., Mierop, A., Stahl, C., & Hütter, M. (2019). Evidence suggestive of uncontrollable attitude acquisition replicates in an instructions-based evaluative conditioning paradigm: Implications for associative attitude acquisition. *Journal of Experimental Social Psychology*, *85*, https://doi.org/10.1016/j.jesp.2019.103841

Corneille, O., & Stahl, C. (2019). Associative attitude learning: A closer look at evidence and how it relates to attitude models. *Personality and Social Psychology Review*, 23, 161-189. https://doi.org/10.1177/1088868318763261.

Critchley, H. D., & Nagai, Y. (2012). How emotions are shaped by bodily states. *Emotion Review*, *4*(2), 163–168.

Crites Jr, S. L., Fabrigar, L. R., & Petty, R. E. (1994). Measuring the affective and cognitive properties of attitudes: Conceptual and methodological issues. *Personality and Social Psychology Bulletin*, *20*(6), 619-634.

Cummins, J., Hussey, I., & Hughes, S. (2019). *The AMPeror's New Clothes: Performance on the Affect Misattribution Procedure is Mainly Driven by Awareness of Influence of the Primes*. DOI: 10.31234/osf.io/d5zn8

Cunningham, W. A., Packer, D. J., Kesek, A., & Van Bavel, J. (2009). Implicit measurement of attitudes: A physiological approach. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 485-512). New York: Psychology Press.

Damasio, A. R. (1994). Descartes' error and the future of human life. *Scientific American*, *271*(4), 144–144.

Damasio, A. R. (1999). *The feeling of what happens: body and emotion in the making of consciousness*. New York Times Book Review.

Davies, S. R., El-Deredy, W., Zandstra, E. H., & Blanchette, I. (2012). Evidence for the role of cognitive resources in flavour–flavour evaluative conditioning. *The Quarterly Journal of Experimental Psychology*, *65*(12), 2297-2308. https://doi.org/10.1080/17470218.2012.701311

Dedonder, J., Corneille, O., Bertinchamps, D., & Yzerbyt, V. (2014). Overcoming correlational pitfalls: Experimental evidence suggests that evaluative conditioning occurs for explicit but not implicit encoding of CS–US pairings. *Social Psychological and Personality Science*, *5*(2), 250-257. https://doi.org/10.1177/1948550613490969

Dedonder, J., Corneille, O., Yzerbyt, V., & Kuppens, T. (2010). Evaluative conditioning of high-novelty stimuli does not seem to be based on an automatic form of associative learning. *Journal of Experimental Social Psychology, 46*(6), 1118-1121. https://doi.org/10.1016/j.jesp.2010.06.004

De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA: Sage Publishers.

De Houwer, J., & Hughes, S. (2016). Evaluative conditioning as a symbolic phenomenon: On the relation between evaluative conditioning, evaluative conditioning via instructions, and persuasion. *Social Cognition*, *34*(5). Special Issue: Conceptual, Theoretical, and Methodological Challenges in Evaluative Conditioning Research, 480-494. Doi: 10.1521/soco.2016.34.5.480.

De Houwer, J., & Moors, A. (2007). How to Define and Examine the Implicitness of Implicit Measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 179-194). New York, NY, US: The Guilford Press.

De Houwer, J., & Moors, A. (2010). Implicit measures: similarities and differences. In Bertram Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: measurement, theory, and applications* (pp. 176–193). New York, NY, USA ; London, UK: Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347.

De Houwer, J., Van Dessel, P., & Moran, T. (in press). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology*.

Devine, P. G., Forscher, P. S., Cox, W. T., Kaatz, A., Sheridan, J., & Carnes, M. (2017). A gender bias habit-breaking intervention led to increased hiring of female faculty in STEM departments. *Journal of Experimental Social Psychology*, *73*, 211-215.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), Advances in experimental social psychology (Vol. 36, pp. 1-51). San Diego, CA: Academic Press.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on psychological science*, *8*(3), 223-241. https://doi.org/10.1177/1745691612460685

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology, 23,* 75-109. https://doi.org/10.1016/S0065-2601(08)60318-4

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603-637.

Fiedler, K. (2014). From intrapsychic to ecological theories in social psychology: Outlines of a functional theory approach. *European Journal of Social Psychology, 44*(7), 657-670. https://doi.org/10.1002/ejsp.2069

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the Implicit Association Tests. *Basic and Applied Social Psychology*, *27*(4), 307-316. https://doi.org/10.1207/s15324834basp2704_3

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology*, *17*(1), 74-147.

Fiedler, K., & Hütter, M. (2014). The limits of automaticity. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 497-513). New York: Guilford Press.

Field, A. P., & Moore, A. C. (2005). Dissociating the Effects of Attention and Contingency Awareness on Evaluative Conditioning Effects in the Visual Paradigm. *Cognition and Emotion, 19*(2), 217-243. https://doi.org/10.1080/02699930441000292

Forscher, P. S., Lai, C. K., Axt, J. R., Ebersole, C. R., Herman, M., Devine, P. G., & Nosek, B. A. (2019). A meta-analysis of procedures to change implicit measures. *Journal of Personality and Social Psychology, 117*(3), 522-559.

Friese, M. & Hofmann, W. & Schmitt, M. (2009). When and why do implicit measures predict behavior? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, *19*, 285-338. https://doi.org/10.1080/10463280802556958

Gast, A., & De Houwer, J. (2012). Evaluative conditioning without directly experienced pairings of the conditioned and the unconditioned stimuli. *The Quarterly Journal of Experimental Psychology*, *65*(9), 1657-1674. http://dx.doi.org/10.1080/17470218.2012.665061

Gawronski, B. (2019). Six lessons for a cogent science of implicit bias and its criticism. *Perspectives on Psychological Science, 14,* 574-595.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692-731. http://dx.doi.org/10.1037/0033-2909.132.5.692.

Gawronski, B., & Bodenhausen, G. V. (2011). The Associative-Propositional Evaluation Model: Theory, Evidence, and Open Questions. *Advances in Experimental Social Psychology*, *44*, 59-127. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2014). The associative-propositional evaluation model: Operating principles and operating conditions of evaluation. *Dual-process theories of the social mind*, 188-203.

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*, *13*, e28024. https://doi.org/10.5964/spb.v13i3.28024

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

Gawronski, B., & Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in social psychology* (pp. 29-55). New York, NY: Taylor & Francis.

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

Gawronski, B., & Walther, E. (2012). What do memory data tell us about the role of contingency awareness in evaluative conditioning? *Journal of Experimental Social Psychology, 48*(3), 617-623. https://doi.org/10.1016/j.jesp.2012.01.002

Gawronski, B., Balas, R., & Creighton, L. A. (2014). Can the formation of conditioned attitudes be intentionally controlled? *Personality and Social Psychology Bulletin*, *40*(4), 419-432. https://doi.org/10.1177/0146167213513907

Gawronski, B., Deutsch, R., Mbirkou, S., Seibt, B., & Strack, F. (2008). When "just say no" is not enough: Affirmation versus negation training and the reduction of automatic

stereotype activation. *Journal of Experimental Social Psychology*, *44*(2), 370-377.

https://doi.org/10.1016/j.jesp.2006.12.004

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes

unconscious? *Consciousness and cognition*, *15*(3), 485-499.

https://doi.org/10.1016/j.concog.2005.11.007

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell

us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological

Science*, *2*, 181–193.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of

implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology

Bulletin, 43,* 300-312. https://doi.org/10.1177/0146167216684131

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-

esteem, and stereotypes. *Psychological Review, 102,* 4-27.

Greenwald, A. G., & Banaji, M. R. (2017). The implicit revolution: Reconceiving the

relation between conscious and unconscious. *American Psychologist*, *72*(9), 861-871.

http://dx.doi.org/10.1037/amp0000238

Greenwald, A. G., & Lai, C. K. (2019). Implicit Social Cognition. *Annual Review of

Psychology.* https://doi.org/10.1146/annurev-psych-010419-050837

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., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009).

Understanding and using the Implicit Association Test: III. Meta-analysis of predictive

validity. *Journal of Personality and Social Psychology*, *97*(1), 17-41. DOI:10.1037/a0015575

Hahn, A., & Gawronski, B. (2018). Implicit social cognition. In J. T. Wixted (Ed.), *The Stevens' handbook of experimental psychology and cognitive neuroscience, Volume 4* (4th ed., pp. 395-427). New York, NY: Wiley.

Hahn, A., Judd, C. M., Hirsh, H. K., & Blair, I. V. (2014). Awareness of implicit attitudes. *Journal of Experimental Psychology: General*, *143*(3), 1369-1392. doi:10.1037/a0035028

Henderson, J. M. (2014). Eye-tracking technology aims to take your unconscious pizza order. *The Conversation,* on December 09th 2014.

Heycke, T., Aust, F., & Stahl, C. (2017). Subliminal influence on preferences? A test of evaluative conditioning for brief visual conditioned stimuli using auditory unconditioned stimuli. *Royal Society Open Science*, *4*(9), 160935.

Heycke, T., & Gawronski, B. (2019). Co-occurrence and relational information in evaluative learning: A multinomial modelling approach. *Journal of Experimental Psychology: General.* http://dx.doi.org/10.1037/xge0000620

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385. https://doi.org/10.1177/0146167205275613

Högden, F., Hütter, M. & Unkelbach, C. (2018). Does Evaluative Conditioning depend on awareness? Evidence from a Continuous Flash Suppression paradigm. *Journal of Experimental Psychology Learning Memory and Cognition*, 44, 1641-1657.

Hu, X., Gawronski, B., & Balas, R. (2017). Propositional Versus Dual-Process Accounts of Evaluative Conditioning: I. The Effects of Co-Occurrence and Relational Information on Implicit and Explicit Evaluations. *Personality and Social Psychology Bulletin*, *43*(1), 17-32. https://doi.org/10.1177/0146167216673351

Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2018). When people co-occur with good or bad events: Graded effects of relational qualifiers on evaluative conditioning. *Personality and Social Psychology Bulletin*, 38(4). https://doi.org/10.1177/0146167218781340

Hummel, J. E. (2010). Symbolic versus associative learning. *Cognitive Science*, *34*, 958-965.

Hütter, M., & De Houwer, J. (2017). Examining the contributions of memory-dependent and memory-independent components to evaluative conditioning via instructions. *Journal of Experimental Social Psychology*, *71*, 49-58. https://doi.org/10.1016/j.jesp.2017.02.007

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116-159. https://doi.org/10.1080/10463283.2016.1212966

Hütter, M., & Rothermund, K. (2020). Automatic processes in evaluative learning. *Cognition & Emotion, 34*(1), 1-20.

Hütter, M., & Sweldens, S. (2018). Dissociating controllable and uncontrollable effects of affective stimuli on attitudes and consumption. *Journal of Consumer Research, 45,* 320-349. https://doi.org/10.1093/jcr/ucx124

Hütter, M., Sweldens, S., Stahl, C., Unkelbach, C., & Klauer, K. C. (2012). Dissociating contingency awareness and conditioned attitudes: Evidence of contingency-unaware evaluative conditioning. *Journal of Experimental Psychology: General, 141*(3), 539-557. DOI: 10.1037/a0026477

Ito, T. A., Friedman, N. P., Bartholow, B. D., Correll, J., Loersch, C., Altamirano, L. J., & Miyake, A. (2015). Toward a comprehensive understanding of executive cognitive

function in implicit racial bias. Journal of Personality and Social Psychology, 108(2), 187. http://dx.doi.org/10.1037/a0038557

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541. https://doi.org/10.1016/0749-596X(91)90025-F

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. *Varieties of memory and consciousness: Essays in honour of Endel Tulving*, 391-422.

James, W. (1884). What is an emotion? *Mind*, *9*(34), 188–205.

Johnson, I. R., Kopp, B. M., & Petty, R. E. (2018). Just say no! (and mean it): Meaningful negation as a tool to modify automatic racial attitudes. *Group Processes & Intergroup Relations*, *21*(1), 88-110. https://doi.org/10.1177/1368430216647189

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, *4*(6), 533-550. https://doi.org/10.1111/j.1745-6924.2009.01164.x

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71,* 1–31.

Klauer, K. C., Dittrich, K., Scholtes, C., & Voss, A. (2015). The invariance assumption in process-dissociation models: An evaluation across three domains. *Journal of Experimental Psychology: General, 144,* 198-221.

Klauer, K. C., & Teige-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology, 43*(4), 648-655. https://doi.org/10.1016/j.jesp.2006.06.003

Kukken, N., Hütter, M. & Holland, R. (2020). Are there two independent evaluative conditioning effects in relational paradigms? Dissociating the effects of CS-US pairings and their meaning. *Cognition & Emotion, 34*(1), 170-187.

Kurdi, B., Morris, A., & Cushman, F. A. (2020). The role of causal structure in implicit cognition. *PsyArXiv. February*, *1*.

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomezsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*, 74, 569-586. http://doi.org/10.1037/amp0000364.

Lai, C. K., Marini, M., Lehr, S. A., Cerruti, C., Shin, J.-E. L., Joy-Gaba, J. A., Ho, A. K., Teachman, B. A., Wojcik, S. P., Koleva, S. P., Frazier, R. S., Heiphetz, L., Chen, E. E., Turner, R. N., Haidt, J., Kesebir, S., Hawkins, C. B., Schaefer, H. S., Rubichi, S., Sartori, G., Dial, C. M., Sriram, N., Banaji, M. R., & Nosek, B. A. (2014). Reducing implicit racial preferences: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology: General, 143*(4), 1765-1785.

Lange, C. G. (1885). The mechanism of the emotions. *The classical psychologists*, 672-684.

Lieberman, M. D. (2007). The X-and C-systems: The neural basis of reflexive and reflective social cognition (pp. 290-315). In E. Harmon-Jones & P. Winkelman (eds.), *Fundamentals of Social Neuroscience*. New York: Guilford.

Mann, T. C., & Ferguson, M. J. (2015). Can we undo our first impressions? The role of reinterpretation in reversing implicit evaluations. *Journal of Personality and Social Psychology*, *108*(6), 823-849. 10.1037/pspa0000021.

Mann, T. C., Kurdi, B., & Banaji, M. R. (2019). How effectively can implicit evaluations be updated? Using evaluative statements after aversive repeated evaluative

pairings. *Journal of Experimental Psychology: General*.

http://dx.doi.org/10.1037/xge0000701

Marini, M., Banaji, M. R., & Pascual-Leone, A. (2018). Studying implicit social

cognition with noninvasive brain stimulation. *Trends in cognitive sciences*.

https://doi.org/10.1016/j.tics.2018.07.014.

McConnell, A. R., & Rydell, R. J. (2014). The Systems of Evaluation Model: A dual-

systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual*

process theories of the social mind (pp. 204-217). New York: Guilford.

Meissner, F., & Rothermund, K. (2013). Estimating the contributions of associations

and recoding in the Implicit Association Test: The ReAL model for the IAT. *Journal of*

Personality and Social Psychology, *104*(1), 45-69. doi: 10.1037/a0030734

Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., & Rothermund, K. (2019).

Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations,

and sophisticated solutions. *Frontiers in Psychology*, *10*. 10.3389/fpsyg.2019.02483.

Melnikoff, D. E., & Bailey, A. H. (2017). Preferences for moral vs. immoral traits in

others are conditional. *Proceedings of the National Academy of Sciences of the United States*

of America,115: E592–E600.

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in*

cognitive sciences, *22*(4), 280-293. https://doi.org/10.1016/j.tics.2018.02.001

Mierop, A., Hütter, M., & Corneille, O. (2017). Resource availability and explicit

memory largely determine evaluative conditioning effects in a paradigm claimed to be

conducive to implicit attitude acquisition. *Social Psychological and Personality Science*, *8*,

758-767.

Mierop, A. Maurage, P., & Corneille, O. (in press). Cognitive Load Impairs Evaluative Conditioning, even when Individual CS and US Stimuli are Successfully Encoded. *International Review of Social Psychology*.

Moors, A. (2016). Automaticity: Componential, causal, and mechanistic explanations. *Annual Review of Psychology*, *67*, 263-287. 10.1146/annurev-psych-122414-033550

Moors, A., & De Houwer, J. (2006). Automaticity: a theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297- 326.

Moors, A., & De Houwer, J. (2007). What is automaticity? An analysis of its component features and their interrelations. In J. A. Bargh (Ed.), *Social psychology and the unconscious: The automaticity of higher mental processes* (pp 11-50). New York: Psychology Press.

Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion*, *27*(4), 743-752. https://doi.org/10.1080/02699931.2012.732040.

Morrison T.G., Kiss M. (2017) Modern Racism Scale. In: Zeigler-Hill V., Shackelford T. (eds) *Encyclopedia of Personality and Individual Differences*. Springer, Cham.

Newell, B. R., & Shanks, D. R. (2007). Recognising what you like: Examining the relation between the mere-exposure effect and recognition. *European Journal of Cognitive Psychology, 19*(1), 103-118. https://doi.org/10.1080/09541440500487454

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231.

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science*, *16*(2), 65-69. https://doi.org/10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in cognitive sciences*, *15*(4), 152-159. https://doi.org/10.1016/j.tics.2011.01.005

Olson, M., & Fazio, R. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12,* 413-417.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171-192. doi: 10.1037/a0032734

Payne, B. K., & Bishara, A. J. (2009). An integrative review of process dissociation and related models in social cognition. *European Review of Social Psychology*, *20*(1), 272-314. https://doi.org/10.1080/10463280903162177

Payne, B. K., Burkley, M., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology, 94,* 16-31.

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*(3), 277-293. DOI: 10.1037/0022-3514.89.3.277

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233-248. https://doi.org/10.1080/1047840X.2017.1335568

Petty, R. E., Tormala, Z. L., Brinol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: an exploration of the PAST model. *Journal of Personality and Social Psychology*, *90*(1), 21-41. DOI: 10.1037/0022-3514.90.1.21

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only

when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 130-144. DOI: 10.1037/0278-7393.33.1.130

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology: Animal Behavior Processes, 35*(2), 279-285. DOI: 10.1037/a0013429

Prinz, J. (2004). *Gut reactions: A perceptual theory of emotion*. Oxford University Press.

Prinz, J. (2012). *The conscious brain*. Oxford University Press.

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

Régner, I., Thinus-Blanc, C., Netter, A., Schmader, T., & Huguet, P. (2019). Committees with implicit biases promote fewer women when they do not believe gender bias exists. *Nature Human Behaviour, 3*, 1171, 1179. doi:10.1038/s41562-019-0686-3 .

Rocklage, M. D., & Fazio, R. H. (2015). The Evaluative Lexicon: Adjective use as a means of assessing and distinguishing attitude valence, extremity, and emotionality. *Journal of Experimental Social Psychology, 56*, 214-227.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*(2), 79-82. https://doi.org/10.1111/j.0963-7214.2004.00279.x

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: a systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*(6), 995-1008. DOI:10.1037/0022-3514.91.6.995

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69*(5), 379.

Schimmack, U. (2019). The Implicit Association Test: A Method in Search of a Construct. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619863798

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of Personality and Social Psychology*, *45*(3), 513.

Sherman, J. W. (2009). Controlled influences on implicit measures: Confronting the myth of process-purity and taming the cognitive monster. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures* (pp. 391-426). Hillsdale, NJ: Erlbaum.

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., & Groom, C. J. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*(2), 314-335. doi=10.1037%2F0033-295X.115.2.314

Smith, C. T., Calanchini, J., Hughes, S., Van Dessel, P., & De Houwer, J. (2020). The impact of instruction- and experience-based evaluative learning on IAT performance: A Quad model perspective. *Cognition and Emotion, 34*(1), 21-41.

Smith, E. R., & 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*(2), 108-131. https://doi.org/10.1207/S15327957PSPR0402_01

Stahl, C., & Aust, F. (2018). Evaluative conditioning as memory-based judgment. *Social Psychological Bulletin*, *13*, e28589.

Stahl, C., & Bading, K. C. (2019). Evaluative conditioning of pattern-masked nonwords requires perceptual awareness. *Journal of Experimental Psychology. Learning, Memory, and Cognition*.

Stahl, C., Haaf, J., & Corneille, O. (2016). Subliminal evaluative conditioning? Above-chance CS identification may be necessary and insufficient for attitude learning. *Journal of Experimental Psychology: General*, *145*(9), 1107-1131.

Stanley, D., Phelps, E., & Banaji, M. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science*, *17*(2), 164-170. https://doi.org/10.1111/j.1467-8721.2008.00568.x

Steffens, M. C. (2004). Is the implicit association test immune to faking? *Experimental Psychology*, *51*(3), 165-179.

Stoet, G., & Geary, D. C. (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, *29*(4), 581-593.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220-247. https://doi.org/10.1207/s15327957pspr0803_1

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: a nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*(5), 768-777.

Sweldens, S., Tuk, M., & Hütter, M. (2017). How to study consciousness in consumer research, A Commentary on Williams and Poehlman. *Journal of Consumer Research, 44*(2), 266-275.

Teige-Mocigemba, S., & Klauer, K. C. (2008). 'Automatic' evaluation? Strategic effects on affective priming. *Journal of Experimental Social Psychology, 44,* 1414-1417. https://doi.org/10.1016/j.jesp.2008.04.004

Teige-Mocigemba, S., Penzl, B., Becker, M., Henn, L., & Klauer, K. C. (2016). Controlling the "Uncontrollable": Faking effects on the Affect Misattribution Procedure. *Cognition & Emotion, 30,* 1470-1484. 10.1080/02699931.2015.1070793

Van Dessel, P., Cone, J., Gast, A., & De Houwer, J. (2020). The impact of valenced verbal information on implicit and explicit evaluation: The role of information diagnosticity, primacy, and memory cueing. *Cognition & Emotion, 34*(1), 74-85. https://doi.org/10.1080/02699931.2019.1594703

Van Dessel, P., De Houwer, J., & Smith, C. (2018). Relational information moderates approach-avoidance instruction effects on implicit evaluation. *Acta Psychologica, 184,* 137-143. https://doi.org/10.1016/j.actpsy.2017.03.016

Van Dessel, P., De Houwer, J., Gast, A., & Smith, C. T. (2015). Instruction-Based Approach-Avoidance Effects. *Experimental Psychology*, *62(3),* 161-169. https://doi.org/10.1027/1618-3169/a000282

Van Dessel, P., De Houwer, J., Roets, A., & Gast, A. (2016). Failures to change stimulus evaluations by means of subliminal approach and avoidance training. *Journal of Personality and Social Psychology*, *110(1), 1-15*. DOI: 10.1037/pspa0000039

Van Dessel, P., Mertens, G., Smith, C. T, & De Houwer, J. (2017). The mere exposure instruction effects: Mere exposure instructions influence liking. *Experimental Psychology, 64*(5), 299-314. https://doi.org/10.1027/1618-3169/a000376

Van Dessel, P., Ye, Y., & De Houwer, J. (2018). Changing Deep-Rooted Implicit Evaluation in the Blink of an Eye: Negative Verbal Information Shifts Automatic Liking of Gandhi. *Social Psychological and Personality Science*, https://doi.org/10.1177/1948550617752064

Vargas, P., Sekaquaptewa, D., & von Hippel, W. (2007). Armed only with paper and pencil: "Low-tech" measures of implicit attitudes. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit Measures of Attitudes* (pp 103-124). New York, NY: Guilford Press.

Wang, M. Y., & Chang, H. C. (2004). The mere exposure effect and recognition memory. *Cognition & Emotion, 18*(8), 1055-1078. https://doi.org/10.1080/02699930341000374

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

Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude–behavior consistency. *Journal of Personality and Social Psychology*, *47*(1), 5.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. Journal of Personality and Social Psychology, 81(5), 815-827.

Zanon, R., De Houwer, J., & Gast, A. (2012). Context effects in evaluative conditioning of implicit evaluations. *Learning and Motivation*, *43*(3), 155-165. https://doi.org/10.1016/j.lmot.2012.02.003

Zeigler-Hill, V. (2006). Discrepancies between implicit and explicit self-esteem: Implications for narcissism and self-esteem instability. *Journal of Personality*, *74*(1), 119-144. https://doi.org/10.1111/j.1467-6494.2005.00371.x