2

MEASURING AND MODELING IMPLICIT COGNITION

Samuel A.W. Klein and Jeffrey W. Sherman

The widespread use of indirect measures in the psychological sciences, particularly social psychology, was motivated by multiple concerns with direct, self-report measures. First, self-presentational motives may lead respondents to misrepresent their true attitudes and beliefs, particularly when responding to questions about socially sensitive topics, such as race. Second, people sometimes have limited access to their own attitudes and beliefs (Greenwald and Banaji 1995). As such, people may be willing to respond accurately but are unable to do so. Lastly, even if people are able to accurately introspect about their attitudes and beliefs, they may not be sufficiently motivated to do so (Hahn and Gawronski 2019). Indirect measures were designed to circumvent these potential problems by inferring attitudes and beliefs from behavior, rather than by directly asking people to report them. Toward that end, in some cases, indirect measures conceal the purpose of the measure or possess properties that make them resistant to manipulation. The advent of indirect measures has led to an explosion of empirical research and theoretical advances in the study of people's cognition and behavior. At the same time, there remain significant conceptual and methodological problems that impede the use and interpretation of indirect measures. In this chapter, we will discuss these problems and a solution via the use of formal models that disentangle cognitive processes underlying implicit cognition data.

Definitional Issues

There are now a great variety of indirect measures of implicit cognition, including the Implicit Association Test (IAT; Greenwald et al. 1998), evaluative priming tasks (e.g., Fazio et al. 1995), Go/No Go Association Task (GNAT; Nosek and Banaji 2001), the Weapons Task (Payne 2001), the First Person Shooter Task (Correll et al. 2002), the Affect Misattribution Procedure (Payne et al. 2005), and the Stereotype Misperception Task (Krieglmeyer and Sherman 2012). These, among other indirect measures, vary on a long list of features, including the time between presentations of various stimuli (stimulus onset asynchrony), the ambiguity of target information, when judgments can be made within each trial, and many more (see Gawronski and Brannon 2018 for a more comprehensive review of indirect measures and their varying features). Perhaps the most basic conceptual problem concerns the name we use to describe these measures. Most commonly, they are referred to as implicit measures. One problem with this label is that the term "implicit" has multiple meanings in experimental psychology. In some

 DOI: 10.4324/9781003014584-4

traditions, the term refers to automatic processes that occur without awareness or intention, cannot be controlled, and are highly efficient. In other traditions, the term refers to processes that merely operate unconsciously. When people use the term "implicit measure," it is not clear which of these features are being invoked (for a review, see Payne and Gawronski 2010). The term often suggests features of the measures that they may not possess. For example, although "implicit" often implies lack of awareness of the measure's purpose, people are well aware of the intent of the IAT after a few trials. Moreover, among the plethora of measures available, there is great variation in the extent to which they possess these various processing attributes. One feature that is common to all measures labeled "implicit" is that they infer attitudes from task performance rather than by directly asking participants to report them. Therefore, we refer to any measure that indirectly assesses psychological attributes via task performance as an "indirect measure" (versus "direct measure").

The same definitional problem applies to the constructs presumed to be revealed by the measures. Most commonly, they are referred to as implicit attitudes or evaluations. However, the same ambiguities surrounding the meaning of "implicit" apply. It is not clear whether the implication of "implicit" is that people are unaware of the evaluations, that they are formed and used without intention, that they cannot be controlled, or that they operate efficiently. In fact, which of these features apply is dependent on both the means by which the evaluation was measured (i.e., which indirect measure is used) and the subject of the evaluation (e.g., race, age, fruit, dogs, etc.). For these reasons, ideally, implicit evaluations would, instead, be referred to as indirect evaluations. However, given the extent to which the term "implicit bias" has saturated academic and popular culture, we reluctantly retain the term "implicit evaluation." It is simply too impractical to change. Still, note that our use of the term is intended to signify only that the evaluations are implied by performance on an indirect measure rather than explicitly provided on a direct measure. In that sense, the evaluations are, indeed, implicit in the given responses.

We prefer the term "evaluation" over "attitude." Whereas evaluations are assessment outcomes that may be based on a variety of sources, "attitudes" in this context implies the existence of distinct mental representations (implicit and explicit attitudes) that exist in our mind and are uniquely accessed by direct and indirect measures. One problem with viewing indirectly measured outcomes as Things is that it discourages a deeper understanding of the constructive nature of responses on indirect measures. Those evaluations are not mental apples waiting to be picked by an indirect measure; they are constructed from a variety of sources and processes in the act of responding to the task demands. For example, the stimulus must be attended to and interpreted, the correct or intended response must be determined, and the response must be made, which requires mental/physical coordination, self-regulation, motor action, and so on (e.g., Sherman et al. 2008). Thus, implicit evaluation is something we *do*, not something we *have* (De Houwer 2019). The view of indirect measure outcomes as Things also implies a level of situational and temporal stability that has not held up to scrutiny (Gawronski et al. 2017). Indeed, implicit evaluations show considerable contextual and temporal variability, and are quite malleable in response to interventions (e.g., Brannon and Gawronski 2017). In contrast, when viewed as measure-induced constructed evaluations, the expectation of stability is significantly diminished.[1]

A further implication of the Thing view is that the outcomes of different indirect measures should correlate strongly with another. After all, if they are all measuring the same Thing in our heads, then there should be high correspondence among them. However, correlations among indirect measures are modest, at best (Bar-Anan and Nosek 2014). These low correlations are indicative of the fact that the demands imposed by indirect measures and the processes recruited to meet those demands differ among measures in many ways. The outcomes of these measures reflect the impact of these differing demands and processes in completing the tasks. Thus, it is not a simple matter of measuring the Thing sitting in our heads: Implicit evaluations are constructed in real time.

Conceptual Challenges in Measuring and Interpreting Implicit Cognition

In one way or another, the main conceptual issues with indirect measures all can be traced to the framing of indirect (versus direct) measures in terms of dual-process models of human cognition. Dual-process models propose that there are two distinct types of mental processes that characterize human cognition (Sherman et al. 2014). Whereas automatic processes occur without awareness or intention, cannot be controlled, and are highly efficient, controlled processes operate with awareness and intention, can be controlled, and require cognitive resources. Upon the introduction of indirect measures, the distinction between indirect and direct measures was mapped onto dual process models, with indirect measures tapping automatic processes and direct measures tapping controlled processes. With this mapping, indirect measures and their outcomes were endowed with the presumed features of automatic processes. First and foremost, this implied that the properties and outcomes of indirect measures are qualitatively distinct from those of direct measures. Further, it implied that indirect measures and their outcomes would be characterized by the operation of specific types of processes that operated under specific types of conditions and were based on specific types of mental representations. None of these implications were directly tested, and their assumption has greatly impacted how indirect measures are understood and how their outcomes are explained.

We have already described how use of the term "implicit" to denote features of automaticity has clouded thinking about implicit measures and evaluations. Here, we will describe four ways in which features of indirect measures and implicit evaluations associated with automaticity have been conflated, leading to conceptual and explanatory uncertainty. First, the processes that underlie implicit cognition tend to be conflated with the conditions under which those processes operate. Second, indirect measures are often assumed to operate under a unique set of conditions. Third, measures (i.e., indirect vs. direct) are often conflated with the constructs they are designed to reveal. Finally, indirect measures have been conflated with the processes that drive their output (see also Sherman and Klein 2021).

Operating Principles and Operating Conditions

Whereas *operating principles* refer to the qualitative nature of a process or representation, describing what it does (e.g., activation of associations; inhibition), *operating conditions* refer to the conditions under which those processes or representations operate (e.g., Does the process still occur when the person is mentally exhausted?; Sherman et al. 2014). Operating principles and operating conditions are bidirectionally conflated. Sometimes, assumptions or knowledge about operating conditions influences assumptions about the operating principles. Other times, assumptions or knowledge about operating principles drives assumptions about the operating conditions. In neither case are the assumptions warranted.

Knowledge about the conditions under which processes driving implicit cognition occur does not tell us anything about what those processes are (operating principles). For example, knowing that implicit evaluations do not change when participants are under cognitive load does not necessarily mean that the indirect measure is capturing the activation of associations in memory, a central assumption of many theories of implicit cognition (e.g., Gawronski and Bodenhausen

2006; Strack and Deutsch 2004). That is, although the process seems "automatic," the extent of its automaticity does not define what the process does, merely when it may occur. Likewise, knowledge about what the processes are does not tell us anything about when the processes may or may not operate. For example, knowing that a process is inhibitory in nature does not necessarily mean that the process may only operate when people have full processing capacity. Indeed, there is now substantial evidence that a number of processes considered to be controlled in nature nevertheless operate in seemingly highly efficient ways (Calanchini and Sherman 2013).

The key point is that conclusions about the processes contributing to implicit evaluations require independent tests of the nature of those processes. Assessing the role of inhibition in implicit evaluations requires research that specifically examines the role of inhibition. It cannot be inferred from knowing the operating conditions. Likewise, conclusions about the conditions under which a process may influence implicit evaluations require research that specifically examines those conditions. Conclusions about the intentionality, awareness, controllability, and efficiency of a process require direct tests of those features. They cannot be inferred from knowledge about the nature of the process in question (e.g., it is an associative or inhibitory process; Sherman et al. 2008).

Measures and Operating Conditions

The commonly assumed relationship between implicit evaluations and automatic processes in implicit cognition research introduces yet another conflation. Typically, indirect measures are presumed to reflect processes that are automatic in nature, whereas direct measures are presumed to reflect processes that are considered controlled. At this point, such assumptions are no longer tenable. For example, there is growing evidence that people are aware of their implicit evaluations and how they will influence responses on indirect measures (e.g., Hahn et al. 2014). There also is evidence that people can inhibit implicit bias while completing the measures (e.g., Glaser and Knowles 2008; Krieglmeyer and Sherman 2012; Moskowitz and Li 2011; Sherman et al. 2008) and that responses on the measures are influenced by the availability of processing resources, indicating that performance is not entirely efficient (e.g., Conrey et al. 2005; Correll et al. 2002; Krieglmeyer and Sherman 2012).

There may be significant costs in assuming that indirect/direct measures reflect automatic/controlled processes, particularly given that there are many differences between direct and indirect measures that may not be related to the automatic/controlled distinction. An instructional example can be found in the implicit memory literature. For many years, indirect measures of memory were assumed to reflect the automatic influence of memories whereas direct measures of memory were assumed to reflect the intentional use of memory. After years of research built on this assumption, Roediger and his colleagues (e.g., Roediger 1990) observed that performance on indirect measures of memory depended largely on the encoding and retrieval of perceptual features of stimuli, whereas performance on direct measures of memory largely depended on the encoding and retrieval of conceptual (meaning) features of stimuli. When the type of processing was equated between direct and indirect measures of memory, the dissociations disappeared. Thus, the dissociations were between perceptual and conceptual memory rather than between implicit and explicit memory. In the same way, indirect and direct measures of evaluation differ along many dimensions, such as the use of visual stimuli in indirect but not direct measures. When such differences in the structural properties of the tasks are reduced, the correspondence between implicit and explicit responses rises (Payne et al. 2008). Again, claims about the operating conditions of a measure must be independently established with careful empirical work. They cannot be assumed based on the type of measure.

Measures and Constructs

The confounding of measures and operating conditions directly implicates a corresponding confound between measures and constructs. The assumption that indirect/direct measures reflect automatic/controlled processes forms the basis for the claim that implicit and explicit evaluations are, in fact, qualitatively distinct constructs. Again, the central problem is that the measures (and their associated constructs) differ in many ways beyond the extent to which responses are relatively implicit or explicit. Thus, differences between implicit and explicit evaluations may reflect differences in the extent to which those evaluations are implicit/explicit or they may reflect other differences in the procedural demands of indirect and direct measures.

Measures and Operating Principles

Measures also are often conflated with operating principles. Whereas indirect measures are assumed to capture mostly associative processing, direct measures are assumed to capture mostly inhibitory or propositional processing (Fazio 1995; Gawronski and Bodenhausen 2006; Strack and Deutsch 2004). However, just as is the case with conflating measures with operating conditions or constructs, the problem is that indirect and direct measures differ in many ways. To the extent that direct and indirect measures differ in their structural features (e.g., the use of visual images) the processes invoked in responding to those measures also will differ. As such, the responses may reflect structural features of the measures that have nothing to do with the presumed operating principles.

Another problem with conflating measures with particular operating principles is that no measure is process-pure. That is, no measure, indirect or direct, reflects the operation of a single type of process. Though assumed to primarily reflect associative processes, a plethora of other types of processes have been shown to influence responses on indirect measures, including the inhibition of associations, the accurate identification of stimuli, intentional coding strategies, attributional processes, and response biases (Sherman and Klein 2021). Ultimately, the interaction of many processes determines responses, and measure outcomes cannot reveal, on their own, the nature of the underlying processes that produced the outcomes. Conclusions about operating principles must be established directly through empirical work and cannot be inferred from operating conditions or measures.

As a concrete example, consider the finding that aging is associated with increased implicit racial bias. Typically, this effect would be attributed to the fact that older people have more biased associations. However, such differences may also be related to changes in executive function associated with aging. Indeed, the second author's own research showed that bias associated with aging was related to failures of self-regulation rather than to differences in underlying associations (see *Applying the Quad Model* section for elaboration; Gonsalkorale, Sherman et al. 2009). Or consider the fact that younger and older people have been observed to demonstrate similar degrees of implicit anti-aging bias. One might conclude that negative attitudes about aging are so pervasive that even older people possess them. However, aging was, in fact, associated with less negative associations with older people. At the same time, aging also was associated with a weaker ability to regulate the expression of said negative associations. In effect, these two processes cancelled each other out. Older and younger people responded similarly because even though older people had less negative associations, they were less able to control them (Gonsalkorale et al. 2014). In the next section, we will describe how one can simultaneously measure associations and the ability to overcome them.

Process Modeling of Indirect Measures

Following the theoretical and methodological concerns outlined in the first part of this chapter, we turn to formal mathematical modeling as a solution that has seen a steep rise in popularity over the last two decades (Hütter and Klauer 2016; Ratcliff et al. 2016; Sherman et al. 2010). The purpose of formal modeling is to identify the processes underlying indirect task performance, measure those processes, and describe the ways in which they interact to produce responses. To do so, formal models propose a set of parameters representing the hypothesized processes (e.g., activation of association; inhibition of associations) and a set of equations that describe the ways in which the processes interact and constrain one another. Solving for the parameters yields estimates of the extent to which they contribute to responses.

This technique offers a number of important advantages. First, because formal models input data from a single task, differences in the natures or extents of the processes cannot be due to differences in the features of measures. As described earlier, when different measures are used to assess different operating conditions, constructs, or operating principles, it is always possible that the observed differences are related to differences in the features of the measures that have nothing to do with the proposed constructs, operating conditions, or operating principles. Second, inherent in the use of formal models is the assumption that multiple processes interact to drive outcomes. Thus, the measures are not assumed to reflect only one process. Third, constructing a model requires the use of an explicit theory about which processes contribute to performance and the manner in which those processes interact with one another. Therefore, many of the key assumptions underlying conceptual process models of implicit cognition can be directly tested via formal modeling techniques. Finally, competing models that identify different processes or different relationships among the processes can be compared in terms of their ability to fit the data. In essence, this is a means of comparing the validity of different theories of implicit cognition.

Though formal modeling provides a means for proposing and testing the operating principles that determine implicit evaluations, it is important to note that the psychological meanings of the parameters must be independently established with empirical research. If we want to claim that a model parameter represents the inhibition of associations, we need to empirically demonstrate that the parameter responds the way inhibition should. As stated earlier, operating principles cannot be assumed; they must be tested. Likewise, any claims about the conditions under which the parameters operate must be established independently. If we want to claim that a parameter is dependent on the availability of cognitive resources, we need to show that empirically.

A wide variety of formal modeling techniques have been used toward these ends, including signal detection, process dissociation, diffusion models, and multinomial tree models. Though a full discussion of these different types of models is beyond the scope of this chapter (for comprehensive reviews, see Hütter and Klauer 2016; Ratcliff et al. 2016; Sherman et al. 2010; Wixted 2020), we will present one example in some detail. Specifically, we present an overview of the development and use of the Quadruple Process model (Quad model; Conrey et al. 2005: Sherman et al. 2008), which was initially advanced to account for performance on the IAT. We show how the Quad model enhances our understanding of the processes that drive performance on indirect measures, the manner in which those processes contribute to individual differences in implicit evaluations, and the meaning of contextual variations and malleability in indirect task performance. We also demonstrate how the model helps to explain relationships between implicit evaluations and behavior.

The Quadruple Process Model

The Quad model proposes four distinct processes that interact to produce responses on indirect measures of evaluation. The model proposes parameters for the activation of associations (AC), the detection of correct responses to target stimuli (D), the overcoming of biased associations in favor of correct responses (OB), and a general response bias (G). The structure of the Quad model as applied to the IAT is depicted as a processing tree in Figure 2.1. In the tree, each path represents a likelihood. Processing parameters with lines leading to them are conditional on all preceding parameters. For instance, OB is conditional on both AC and D. The conditional relationships described by the model form a system of equations that predicts the numbers of correct and incorrect responses on different trial types (e.g., compatible and incompatible trials). For example, there are three ways in which an incorrect response can be returned on a trial in which Black and 'pleasant' share a response key for a person with pro-White bias. The first is the likelihood that biased associations between 'Black' and 'unpleasant' are activated (AC), detection succeeds (D), and OB fails (1 – OB), which can be represented by the equation AC × D × (1 – OB). The second is the likelihood that the biased associations are activated (AC) and detection fails (1 – D), which can be represented by the equation AC × (1 – D). The third is the likelihood that biased associations are not activated (1 – AC), detection fails (1 – D), and a bias toward guessing 'unpleasant' (1-G) produces an incorrect response, which can be represented by the equation (1 – AC) × (1 – D) × (1 – G). As such, the overall likelihood of producing an incorrect response on such a trial is the sum of these three conditional probabilities: [AC × D × (1 – OB)] + [AC × (1 – D)] + [(1 – AC) × (1 – D) × (1 – G)]. The respective equations for each item category (i.e., White faces, Black faces, pleasant words, and unpleasant words in both trial types) are then used to predict the observed proportions of errors in a given data set. The model's predictions are compared to the actual data to determine the model's ability to account for the data. An estimate of statistical fit is computed for the difference between the predicted and observed responses. To best approximate the model to the data, the parameter values are changed through estimation methods (e.g., maximum likelihood) until they produce

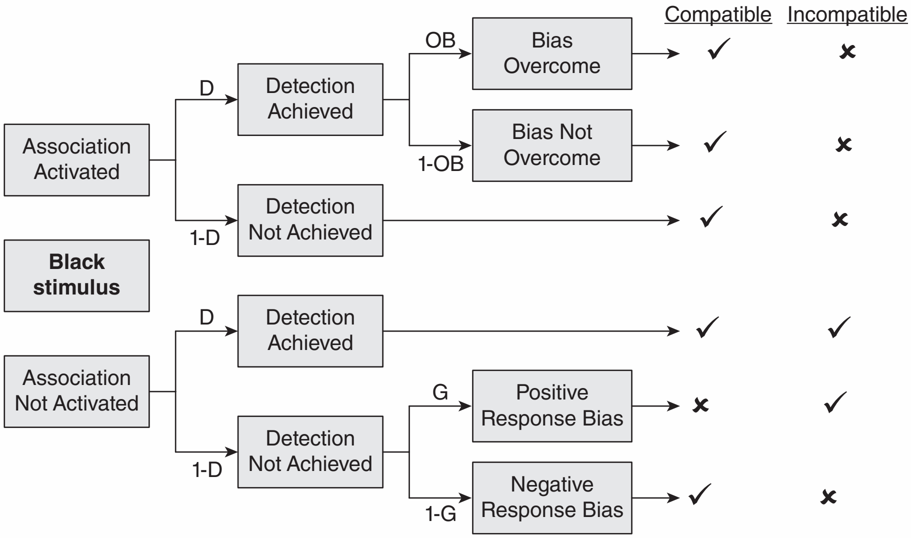

Figure 2.1 The quadruple process model

a minimum possible value of misfit between the observed and predicted responses. The final parameter values that result from this process are interpreted as relative levels of the processes.

Applying the Quad Model

The Quad model has provided consistently good fit across data sets (Conrey et al. 2005; Sherman et al. 2008) and has been applied toward understanding a number of fundamental questions about implicit evaluation (Sherman et al. 2008).

Individual Differences

One central question concerns the interpretation of differences in implicit evaluations among respondents. What does it mean when two individuals or two groups of people differ in implicit evaluations? Traditional approaches to understanding implicit evaluation would suggest that such differences must be due to differences in the underlying evaluative associations among individuals. Application of the Quad model has shown that, indeed, sometimes that is the case, as in differences in pro-White bias between White and Black respondents (Gonsalkorale et al. 2010). However, in other cases, modeling has shown group differences to depend not on the strength of an underlying bias, but on the likelihood of overcoming it. For example, the observation that older individuals display more negative implicit evaluations of some lower-status groups has commonly been understood to represent the presence of more negative associations due to prejudicial social norms during their youth. However, application of the Quad model has shown that greater bias among the elderly is based not on the involvement of more biased associations, but on the likelihood that bias is overcome (Gonsalkorale, Sherman et al., 2009, 2014).

Contextual Variation and Malleability of Bias

Another central question is how to explain changes in implicit evaluation across contexts or in response to interventions. Traditional approaches would suggest that such changes must be due to differences in the underlying evaluative associations across contexts or interventions. Application of the Quad model has shown that, indeed, sometimes that is the case, as in reductions in implicit evaluative race bias when an IAT includes pictures of positive Black and negative White persons (e.g., Gonsalkorale et al. 2010), among respondents focused on a common ingroup identity (Scroggins et al. 2016), among respondents with greater intergroup contact (Rae et al. 2020), or among respondents who have suffered a temporary blow to self-esteem (Allen and Sherman 2011). However, in other cases, modeling has shown such variation to depend not on the strength of an underlying bias, but on the likelihood of overcoming it. For example, the observation of greater evaluative race bias among intoxicated respondents corresponds not with alterations in activated associations but with the likelihood of effectively regulating the influence of those associations (Sherman et al. 2008). The likelihood of overcoming bias similarly accounts for reduced implicit race evaluation when an IAT presents Black and White persons in positive and negative contexts, respectively (for a review, see Calanchini et al. 2021).

Predicting Behavior

A third fundamental question concerns the extent to which implicit evaluations predict behavior. Specifically, when implicit evaluations predict behavior, which component processes of the evaluation are responsible? Traditional approaches would suggest that variations in underlying associations direct behavior and are responsible for the evaluation-behavior link. However, it may be more complicated than that. In one study (Gonsalkorale, von Hippel, et al. 2009), after interacting with a Muslim research confederate (i.e., a research assistant acting the part), White participants completed an anti-Muslim GNAT. As well, the confederate rated how much he enjoyed his interaction with each participant. Results showed that the more negative the subjects' implicit evaluations of Muslims on the GNAT, the less the confederate enjoyed interacting with them. Application of the Quad model to the GNAT data showed that the confederate's liking of the subjects was not predicted solely by the extent of the subjects' negative association with Muslims. Rather, the confederate's liking of the subjects was predicted by an interaction between the subjects' anti-Muslim associations and the likelihood that they overcame them in performing the GNAT. Specifically, when the subjects' associations with Muslims were only moderately negative, the confederate's liking of the subjects did not depend on the likelihood that they overcame their bias on the GNAT. In contrast, when the subjects had strongly negative associations with Muslims, the confederate liked them to the extent that they successfully overcame their bias when performing the GNAT. These findings indicate that the ability to regulate implicit evaluations may play an important role in a person's direct behaviors with members of a group toward whom they are biased.

Summary

In this section, we briefly described the application of one multinomial model, the Quad model, toward understanding key questions in the implicit evaluation literature. Beyond providing a more detailed understanding of the fundamental meaning and basis of implicit responses, modeling deepens understanding of individual differences in implicit evaluations, variability/malleability of implicit evaluations, and evaluation-behavior links. The common understanding of implicit evaluation explains all of these effects by reference to the activation and application of biased associations stored in memory. Modeling shows that, in many cases, these effects are driven by a variety of processes, and sometimes do not involve associations at all. Other research has shown that some of these processes have nothing to do with a specific attitude object, per se, but, rather, represent domain-general cognitive skills. For example, the extent of Detection and Overcoming Bias in the Quad model in evaluations of one domain correlates robustly with the extent of those processes in other domains (Calanchini and Sherman 2013). This indicates that these processes are not tied to specific domains but rather assess general cognitive abilities that influence responses across domains.

These observations made possible by modeling also have important implications for designing interventions to alter implicit evaluations. Traditionally, the assumption has been that such efforts must be targeted toward and are effective to the extent that they change the evaluative associations in people's heads. However, modeling work shows that interventions that alter general cognitive abilities may also be effective in changing implicit evaluations. For instance, training people to more accurately identify stimuli or more effectively inhibit routinized responses may be effective means of implicit bias reduction. As an example, Calanchini et al. (2013) demonstrated that Detection in the Quad model is responsive to training: Participants who completed a counter-prejudicial training task subsequently had higher levels of Detection. Those participants also demonstrated less IAT bias than control participants, suggesting that increases in Detection may be tied to diminished bias. Given the domain-generality of these cognitive skills, developing these abilities may have relative broad payoffs across attitude domains (Calanchini et al. 2014).

Conclusion

Many of the initial foundational assumptions of work on implicit cognition have proven to be problematic. Unfounded assertions regarding the nature of indirect measures, the constructs that they assess, the processes that generate responses on the measures, and the conditions under which those processes operate have significantly complicated efforts to measure, characterize, and understand implicit cognition. Those who use indirect measures for research and those who consume that research should be aware of these complications. For researchers, one potential remedy to these problems is the use of mathematical models combined with careful research to validate any claims about mechanisms and the conditions under which they operate. In addressing these issues, we hope that this chapter helps to provide a framework for the future of doing and understanding implicit cognition.

Related Topics

Chapters 1, 3, 6, 7, 8, 25, 27, 28, 31

Note

1 A popular critique of implicit cognition research is that its data often lacks stability. Removing this stability expectation by viewing the data as measure-induced constructed evaluations strongly devalues the weight of that criticism, as it is founded on a view that separate implicit and explicit Things are sitting in the mind. Although outside the scope of this chapter, see Brownstein et al. (2020) for a more in-depth discussion.

References

Allen, T. J., and Sherman, J. W. 2011. “Ego threat and outgroup derogation: a test of motivated activation versus self-regulatory accounts”. *Psychological Science*, *22*: 331–333.

Bar-Anan, Y., and Nosek, B. A. 2014. “A comparative investigation of seven indirect attitude Measures”. *Behavior Research Methods*, *46*: 668–688. https://doi.org/10.3758/s13428-013-0410-6

Brannon, S. M., and Gawronski, B. 2017. “A second chance for first impressions? Exploring the context-(in) dependent updating of implicit evaluations”. *Social Psychological and Personality Science*, *8*(3): 275–283.

Brownstein, M., Madva, A., and Gawronski, B. 2020. “Understanding implicit bias: putting the criticism into perspective”. *Pacific Philosophical Quarterly*, *101*: 276–307.

Calanchini, J., Gonsalkorale, K., Sherman, J. W., and Klauer, K. C. 2013. “Counter-prejudicial training reduces activation of biased associations and enhances response monitoring”. *European Journal of Social Psychology*, *43*: 321–325.

Calanchini, J., Lai, C. K., and Klauer, K. C. 2021. “Reducing implicit racial preferences: III. A process-level examination of changes in implicit preferences”. *Journal of Personality and Social Psychology*, *121*: 796.

Calanchini, J., and Sherman, J. W. 2013. “Implicit attitudes reflect associative, non-associative, and non-attitudinal processes”. *Social and Personality Psychology Compass*, 7: 654–667. https://doi.org/10.1111/spc3.12053

Calanchini, J., Sherman, J. W., Klauer, K. C., and Lai, C. K. 2014. “Attitudinal and non-attitudinal components of IAT performance”. *Personality and Social Psychology Bulletin*, *40*: 1285–1296.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., and Groom, C. J. 2005. “Separating multiple processes in implicit social cognition: the Quad model of implicit task performance”. *Journal of Personality and Social Psychology*, *89*: 469–487.

Correll, J., Park, B., Judd, C. M., and Wittenbrink, B. 2002. “The police officer’s dilemma: using ethnicity to disambiguate potentially threatening individuals”. *Journal of Personality and Social Psychology*, *83*: 1314–1329. https://doi.org/10.1037/0022-3514.83.6.1314

De Houwer, J. 2019. "Implicit bias is behavior: a functional-cognitive perspective on implicit bias". *Perspectives on Psychological Science*, *14*: 835–840.

Fazio, R. H. 1995. "Attitudes as object-evaluation associations: determinants, consequences, and correlates of attitude accessibility". In R. E. Petty and J. A. Krosnick, eds., *Ohio State University series on attitudes and persuasion, vol. 4: attitude strength: antecedents and consequences*. Hillsdale: Lawrence Erlbaum Associates: 247–282.

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

Gawronski, B., and Bodenhausen, G. V. 2006. "Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change". *Psychological Bulletin*, *132*: 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., and Brannon, S. M. 2018. "Attitudes and the implicit-explicit dualism". In D. Albarracin and B. T. Johnson, eds., *The handbook of attitudes*. London: Routledge: 158–196.

Gawronski, B., and Hahn, A. 2019. "Implicit measures: procedures, use, and interpretation". In H. Blanton, J. M. LaCroix, and G. D. Webster, eds., *Measurement in social psychology*. New York, NY: Taylor and Francis: 29–55.

Gawronski, B., Morrison, M., Phills, C. E., and Galdi, S. 2017. "Temporal stability of implicit and explicit measures: a longitudinal analysis". *Personality and Social Psychology Bulletin*, *43*: 300–312.

Glaser, J., and Knowles, E. D. 2008. "Implicit motivation to control prejudice". *Journal of Experimental Social Psychology*, *44*: 164–172. https://doi.org/10.1016/j.jesp.2007.01.002

Gonsalkorale, K., Allen, T. J., Sherman, J. W., and Klauer, K. C. 2010. "Mechanisms of exemplar exposure effects on implicit attitudes". *Social Psychology*, *41*: 158–168.

Gonsalkorale, K., Sherman, J. W., and Klauer, K. C. 2009. "Aging and prejudice: diminished regulation of automatic race bias among older adults". *Journal of Experimental Social Psychology*, *45*: 410–414. https://doi.org/10.1016/j.jesp.2008.11.004

Gonsalkorale, K., Sherman, J. W., and Klauer, K. C. 2014. "Measures of implicit attitudes may conceal differences in implicit associations: the case of antiaging bias". *Social Psychological and Personality Science*, *5*: 271–278. https://doi.org/10.1177/1948550613499239

Gonsalkorale, K., von Hippel, W., Sherman, J. W., and Klauer, K. C. 2009. "Bias and regulation of bias in intergroup interactions: implicit attitudes toward Muslims and interaction quality". *Journal of Experimental Social Psychology*, *45*: 161–166. https://doi.org/10.1016/j.jesp.2008.07.022

Greenwald, A. G., and Banaji, M. R. 1995. "Implicit social cognition: attitudes, self-esteem, and stereotypes". *Psychological Review*, *102*: 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. K. 1998. "Measuring individual differences in implicit cognition: the implicit association test". *Journal of Personality and Social Psychology*, *74*: 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Hahn, A., and Gawronski, B. 2019. "Facing one's implicit biases: from awareness to acknowledgement". *Journal of Personality and Social Psychology*, *116*: 769–794. https://doi.org/10.1037/pspi0000155

Hahn, A., Judd, C. M., Hirsch, H. K., and Blair, I. V. 2014. "Awareness of implicit attitudes". *Journal of experimental psychology. General*, *143*: 1369–1392. https://doi.org/10.1037/a0035028

Hütter, M., and Klauer, K. C. 2016. "Applying processing trees in social psychology". *European Review of Social Psychology*, *27*: 116–159. https://doi.org/10.1080/10463283.2016.1212966

Krieglmeyer, R., and Sherman, J. W. 2012. "Disentangling stereotype activation and stereotype application in the stereotype misperception task". *Journal of Personality and Social Psychology*, *103*: 205–224. https://doi.org/10.1037/a0028764

Moskowitz, G. B., and Li, P. 2011. "Egalitarian goals trigger stereotype inhibition: a proactive form of stereotype control". *Journal of Experimental Social Psychology*, *47*: 103–116. https://doi.org/10.1016/j.jesp.2010.08.014

Nosek, B. A., and Banaji, M. R. 2001. "The Go/No-Go association task". *Social Cognition*, *19*: 625–666.

Payne, B. K. 2001. "Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon". *Journal of Personality and Social Psychology*, *81*: 181–192.

Payne, B. K., Burkley, M. A., and 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. https://doi.org/10.1037/0022-3514.94.1.16

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

Payne, B. K., and Gawronski, B. 2010. "A history of implicit social cognition: where is it coming from? Where is it now? Where is it going?". In B. Gawronski and B. K. Payne, eds., *Handbook of implicit social cognition: measurement, theory, and applications*. New York: The Guilford Press: 1–15.

Rae, J. R., Reimer, N. K., Calanchini, J., Lai, C. K., Rivers, A. M., Dasgupta, N., . . . Schmid, K. 2020. "Intergroup contact and implicit racial attitudes: contact is related to less activation of biased evaluations but is unrelated to bias inhibition". *PsyArXiv*. https://doi.org/10.31234/osf.io/h4nxd

Ratcliff, R., Smith, P. L., Brown, S. D., and McKoon, G. 2016. "Diffusion decision model: current issues and history". *Trends in Cognitive Sciences*, *20*: 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Roediger, H. L. 1990. "Implicit memory: retention without remembering". *American Psychologist*, *45*: 1043–1056. https://doi.org/10.1037/0003-066X.45.9.1043

Scroggins, W. A., Mackie, D. M., Allen, T. J., and Sherman, J. W. 2016. "Reducing prejudice with labels: shared group memberships attenuate implicit bias and expand implicit group boundaries". *Personality and Social Psychology Bulletin*, *42*: 219–229. https://doi.org/10.1177/0146167215621048

Sherman, J. W., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T. J., and Groom, C. J. 2008. "The self-regulation of automatic associations and behavioral impulses". *Psychological Review*, *115*: 314–335. https://doi.org/10.1037/0033-295X.115.2.314

Sherman, J. W., Gawronski, B., and Trope, Y., eds. 2014. *Dual-process theories of the social mind*. New York: The Guilford Press.

Sherman, J. W., Klauer, K. C., and Allen, T. 2010. "Mathematical modeling of implicit social cognition: the machine in the ghost". In B. Gawronski and B. K. Payne, eds., *Handbook of implicit social cognition: measurement, theory, and applications*. New York: Guilford Press: 156–175.

Sherman, J. W., and Klein, S. A. W. (2021). "The four deadly sins of implicit attitude research". *Frontiers in Personality and Social Psychology, 11*: 60430. https://doi.org/10.3389/fpsyg.2020.604340

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

Wixted, J. T. 2020. "The forgotten history of signal detection theory". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*: 201–233. https://doi.org/10.1037/xlm0000732