

Implicit–Explicit Relations

Brian A. Nosek

University of Virginia

ABSTRACT—*Mental process and mental experience are not the same thing. The former is the operation of the mind; the latter is the subjective life that emerges from that operation. In social evaluation, implicit and explicit attitudes express this distinction. Although it is clear that they are not the same, how they differ is not. Across content domains, implicit and explicit attitude measures show substantial variability in the strength of correspondence, ranging from near zero to strongly positive. Variation in controllability, intentionality, awareness, or efficiency is thought to differentiate implicit and explicit attitudes. Dual-process theories and empirical evidence for moderating influences of implicit–explicit attitude relations provide a framework for comprehending relations between the operation and the experience of the mind.*

KEYWORDS—*implicit–explicit; attitudes; automaticity; Implicit Association Test; consciousness*

Subjective experience is a palpable, compelling, and poor account of mental operations. Introspection is an act of self-assessment, looking inward for answers to important questions: Who am I? Why did I do that? An enduring realization of modern psychology is that such introspection has limited access to mental processes (Nisbett & Wilson, 1977). Rather, humans' conscious experience operates like an intrepid journalist making inferences about a corporation through observable actions and obscured reports from insiders.

It is easy for people to accept that some mental processes governing proprioception (i.e., balance), perception, and language comprehension operate nonconsciously. Somewhat harder to accept is the notion that memory processes have nonconscious components. And, for some people, it seems silly to think of concepts like attitudes, goals, identity, and stereotypes as operating nonconsciously. And yet, modern social psychology proposes that these constructs have active existence distinct from conscious, deliberate, and intentional experience (Greenwald & Banaji, 1995).

Address correspondence to Brian Nosek, University of Virginia, Department of Psychology, Box 400400, Charlottesville, VA 22904; e-mail: nosek@virginia.edu.

The rapidly accumulating literature of implicit influences on social perception, judgment, and action is a consequence of a surge in methodological innovations that manipulate concepts without participants' attention or recognition or that measure concepts without participants' awareness or control. These advances make possible a whole new class of questions such as the following: When will implicit processes predict behavior? How do implicit and explicit mental operations influence one another? What does it really mean for something to be implicit or explicit? How do distinctions between implicit and explicit methods correspond to distinctions between implicit and explicit constructs? This article briefly examines the relationship between implicit and explicit processes with an emphasis on one construct, attitudes, and one implicit measure, the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), because accumulation of knowledge about this construct and this measure has been rapid.

IMPLICIT AND EXPLICIT MEASURES

An explicit response is controllable, intended, made with awareness, and requires cognitive resources. Measures of implicit cognition comprise a heterogeneous set of methods and procedures that differ from measures of explicit cognition by having at least one of the following characteristics: (a) reduced controllability; (b) lack of intention; (c) reduced awareness of the origins, meaning, or occurrence of a response; or (d) high efficiency of processing (Bargh, 1994). Treating these characteristics as a set reflects the lack of evidence for clear distinctions between them rather than a theoretical commitment that they operate the same way.

Assessing explicit attitudes can be simple, such as asking, "Which do you prefer, summer or winter?" Assessing implicit attitudes is usually more indirect. In the IAT, items representing four categories (e.g., summer, winter, pleasant, unpleasant) are sorted as quickly as possible in two different conditions. In one condition, items representing summer and pleasant are categorized with the same response (pressing a certain key on the computer), and items representing winter and unpleasant are categorized with an alternate response (another key). In the second condition, items representing winter and pleasant are categorized with one response and items representing summer

and unpleasant are categorized with the alternate response. The difference in response times between conditions is an indication of association strengths between the concepts and evaluations. Respondents who categorize items more quickly when summer is paired with pleasant (and winter with unpleasant) are said to have an implicit preference for summer compared to winter (try the IAT at <https://implicit.harvard.edu/>).

Explicit and implicit measurement methods have little in common in terms of their procedures. Unlike in explicit attitude assessments, in the IAT, participants do not decide or deliberate about their feelings, they just categorize items as fast as possible. The dramatic difference in methodology leads to reasonable questions of whether and to what degree the assessed constructs are related.

DESCRIPTIVE ACCOUNTS

Evaluations revealed by implicit and explicit measures may have nothing in common, in which case they assess exclusive constructs, and one (implicit measures) might even not be considered an attitude measure. At the opposite extreme, implicit and explicit attitude measures might assess a single construct despite their procedural differences. All differences between measures, in that case, would be attributable to extraneous influences that are irrelevant to attitudes. An intermediate possibility is that implicit and explicit measures assess constructs that are related but distinct. Specifically, implicit and explicit measures might have something in common justifying their shared interpretation as attitude assessments, and something unique justifying the implicit-explicit distinction. This possibility would spur theorists to account for common and distinct influences on implicit and explicit attitude formation and change and for reciprocal influences of implicit and explicit attitudes on each other.

OBSERVED RELATIONS

Evidence of relations between explicit self-report and the IAT supports the third possibility I mentioned: that these tools measure distinct but related constructs. In a study that included measures of attitudes about seven topics (gay-straight, black-white, humanities-science, flowers-insects, Democrats-Republicans, creationism-evolution, fat people-thin people), implicit and explicit attitude measures were moderately related, and the lack of a perfect relationship was not attributable to attitude-irrelevant influences on the measurement instruments (such as average latency of responding on the implicit measures or a tendency to use the extremes of the self-report response scales; Nosek & Smyth, 2007).

In another study, I (Nosek, 2005) randomly assigned Web participants to complete an evaluative IAT and parallel self-report measures for one of more than 50 topics. Smyth and I (Nosek & Smyth, 2007) then analyzed that data with a

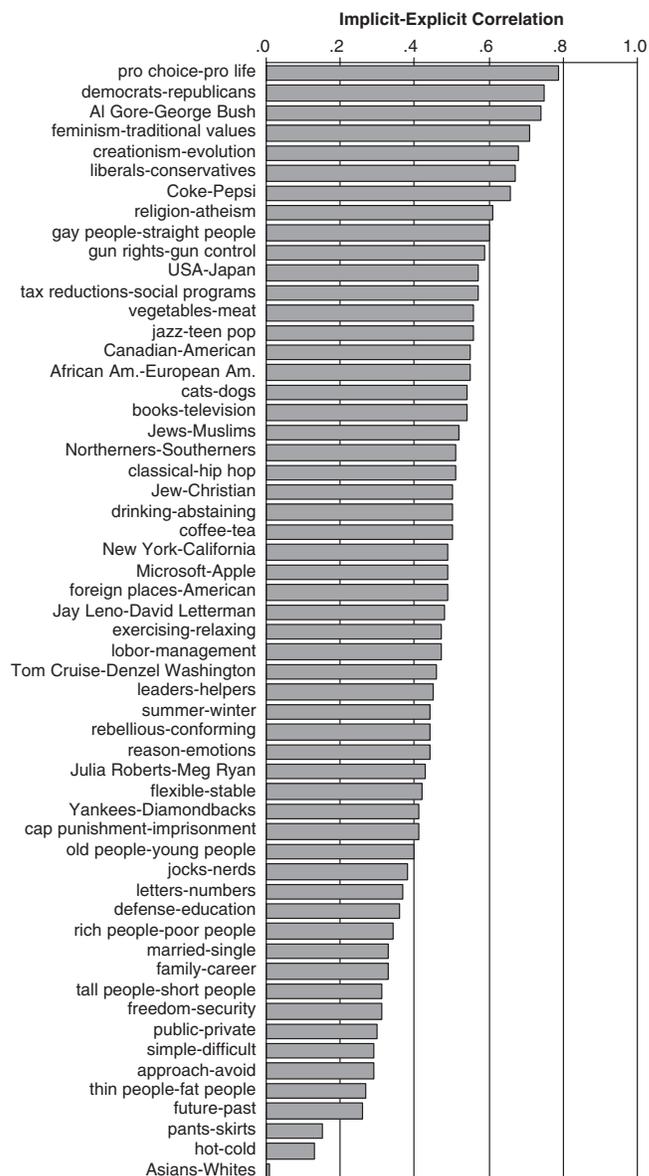


Fig. 1. Correlations between implicit and explicit attitudes for 56 domains (*N* for each was 175–290) ordered from strongest to weakest relation (median correlation = .48). Adapted from Nosek and Smyth (2007) and Nosek (2005).

technique called structural equation modeling. Figure 1 presents the implicit-explicit (hereafter, IE) relations by topic as a correlation. A correlation is an index of the strength of relationship between two concepts: 0 indicates no relationship, and +1 or -1 indicates a perfect relationship. Positive correlations indicate that participants who explicitly reported preferring A to B tended to show a similar preference implicitly.

Correlations varied widely from weakly positive (below .20; e.g., Asians-Whites) to strongly positive (above .75; e.g., pro-choice-pro-life), with a median correlation of .48. We also observed that a two-attitude model fit better than a

single-attitude model for every domain, including those with relatively strong correlations (Nosek & Smyth, 2007). That is, in every case, implicit and explicit attitude measures indicated distinct but related constructs.

Perhaps the most surprising finding from these data is that two measures that have almost nothing in common procedurally actually share quite a bit conceptually. The speed with which people are able to sort items into categories in the IAT is reliably related to their explicit attitudes as revealed in self-assessments.

MODERATORS

The strength of IE correlations varies across topics, and multiple factors account for this variation. I (Nosek, 2005) tested four simultaneous moderators of the relationship between implicit and explicit attitudes. Also, my colleagues and I (Hofmann, Gschwendner, Nosek, & Schmitt, 2005) reviewed the available literature relevant to moderation of IE correlations. This section provides a selective review.

Interpersonal Factors

The historical impetus for development of implicit measures was, in part, concern about the ease of regulating responses on direct self-report measures. By reducing the opportunity for deliberate judgment, implicit measures decrease the likelihood that respondents can hide undesired responses. Thus, implicit measures might assess evaluations that respondents do not want to express because such evaluations violate their self-image (e.g., racial bias in a person who honestly endorses egalitarian values) or because expressing such evaluations may have adverse social consequences (e.g., a negative opinion about a friend's atrocious taste in home decorating).

Self-presentation—that is, altering a response for personal or social purposes—does appear to moderate IE correlations. Those domains for which people report concern about expressing negativity toward a group because of possible social sanction, or because they do not want to have those attitudes, tend to reveal weaker IE correlations than do domains for which people are unconcerned about expressing any negative attitudes.

Conceiving of implicit measures as lie detectors would require that self-presentation be the only moderator of IE correlations and, more extremely, that the measures capture only deliberate hiding of endorsed feelings and not altering of automatic responses that are honestly unwanted. However, self-presentation accounts for only a small portion of the variation in IE correlations. Figure 1 shows, for example, that contrary to the general trend, attitudes toward pants versus skirts elicited weak self-presentation concern but also showed weak IE correspondence (.15). The figure also shows that attitudes toward African Americans versus European Americans elicited strong self-presentation concern but showed relatively strong IE correspondence (.55).

Another interpersonal factor moderating IE correlations is the perceived distinctiveness of one respondent's evaluation compared to those of others. Evaluations that distinguish a person should be easier to recognize and calibrate because that person can contrast him- or herself against the social context. Individual attitudes perceived as more distinct from norms tend to elicit stronger IE correlations than attitudes that are perceived as normative. Perceived distinctiveness, as well as the moderators in the next section, predicted variation in IE relations above and beyond that accounted for by self-presentation.

Intrapersonal Factors

Intrapersonal influences are ones that involve consideration of internal factors, such as amount of personal experience with a particular domain (e.g., people of certain races), or having a clear basis of comparison, such as an opposite (e.g., pro-choice as the opposite of pro-life), to clarify the degree of favorability (dimensionality). Practicing a response leads to its automatization and increased accessibility. Consistently, attitudes that are important or well elaborated tend to elicit stronger IE correlations than ones that are unimportant or infrequently thought about.

Evaluations that conform to a simple, bipolar structure in which liking for a concept (e.g., pro-choice) implies disliking of a second concept (e.g., pro-life) tend to elicit stronger IE correlations than do evaluations that have a unipolar structure. The simple structure of bipolarity, in which the ends of the continuum are mutually reinforcing, may enhance the consistency and efficiency of cognitive processing—one of the hallmarks of automaticity. This moderator has been examined only with the IAT, a measure of relative evaluation, so it is not known whether it will generalize when polarized items are assessed independently (i.e., measuring liking for an item in the absence of its opposite).

Together, these intrapersonal factors account for a significant portion of variation in IE correlations across domains in which implicit attitudes are operationalized with the IAT and explicit attitudes are operationalized with affective ratings of warmth-coldness or pleasantness-unpleasantness.

Other Factors

There are also moderating influences of research measurement, design, and sampling that are not evaluative processes but play an important role in clarifying IE relations. Historically, the prevalence of dissociation between implicit and explicit evaluation has been exaggerated because of a lack of appreciation for the impact of measurement factors. For example, the IAT is a relative measure of evaluation (e.g., summer vs. winter), so explicit measures that assess single evaluations (e.g., summer only) will be at a measurement disadvantage for eliciting relations.

The internal consistency of a measure imposes a theoretical maximum for the degree to which it can be related to other measures. The maximum possible correlation between two measures is equal to the square root of the product of the measures' internal consistency. Across domains in Figure 1, the median correlation was .37 before addressing internal consistency; accounting for internal consistency with structural equation modeling produced an estimated median relation of .48 (Nosek & Smyth, 2007). That estimate corresponds to a 65% increase in shared variance between implicit and explicit measures.

Finally, common sampling biases have led to underestimates of IE correlations. For example, implicit racial-attitude measures often reveal evaluations that suggest an implicit preference for European Americans compared to African Americans that are dissociated from self-reported attitudes. However, most behavioral research is based in university cultures that are hotbeds of egalitarian values. These cultures are characterized by minimal outward expression of social-group biases and, consequently, weak IE correspondence. However, heterogeneous and anonymous sampling of 8,132 Web respondents from a variety of social environments revealed a positive, .47, correlation between implicit and explicit racial attitudes (Greenwald, Nosek, & Banaji, 2003), suggesting that lab work within university cultures might underestimate IE correspondence for social-group biases.

PERSON VERSUS CULTURE

Sometimes an implicit attitude is contrary to an individual's intended, deliberate, endorsed explicit attitude. If the implicit response is actively and honestly rejected, how do researchers make sense of its existence? One perspective suggests that implicit measures are influenced by cultural—that is, *extra-personal*—knowledge, and that this knowledge is distinct from attitudes (Karpinski & Hilton, 2001).

Another interpretation is that implicit evaluation reflects accumulated experience that may not be available to introspection and may not be wanted or endorsed but is still attitudinal because of its potential to influence individual perception, judgment, or action (Banaji, Nosek, & Greenwald, 2004). I favor this interpretation for four reasons. First, as established in the previous two sections, implicit and explicit attitudes are related, and their relationship is moderated by individual factors such as attitude strength. Second, implicit measures relate to attitudes that people declare as their own and not to cultural knowledge that they attribute to others, despite substantial variation in both (Nosek & Hansen, in press). Third, explicit evaluations perceived as similar to cultural norms show weaker IE correlations than do evaluations that are perceived to be distinct; this result is the opposite of what would be expected if knowledge influenced implicit measures (Nosek, 2005). Fourth, both explicit and implicit measures are valid predictors of

individual perception, judgment, and action (Poehlman, Uhlmann, Greenwald, & Banaji, 2006).

DIRECTIONS

Perhaps the most pressing issue for IE-relations researchers is to use the panoply of empirical findings to improve the theories that explain the distinction between implicit and explicit attitudes. Early theory deliberately blended two reasons that implicit responses might differ from explicit ones (Greenwald & Banaji, 1995). Dissociation might occur because people are *unwilling* to report mental contents that they could, in principle, report if they wished to do so. Alternatively, dissociation might occur because people are *unable* to report some of their mental experience because they are not aware of its existence or operation. These reasons for dissociation have not been formally distinguished, in part because it is a difficult problem.

The existence of IE relations might appear to argue against the possibility that people are unable to report on the associations captured by implicit measures. A correlation with self-reported attitudes is a necessary condition for concluding that people are aware of their implicit associations. However, it is a sufficient condition only for a weak form of awareness. An IE correlation supports awareness to the same degree as does a person's ability to report the attitudes of his or her family and friends despite not being able to see into their heads (Nosek, 2005). Conscious awareness usually refers to privileged access to one's own thoughts. Because we can know things about ourselves in the same way we know things about others, by observation, IE correlations are not sufficient to clarify whether individuals know their own "implicit" minds better than an observer could. The fact that people are often surprised by their performance on measures like the IAT suggests that some associations exist outside of awareness. However, this evidence supports a relatively weak form of *unawareness*. People might be able to become aware under the correct circumstances or if given the right motivation or training. In short, questions about who is aware of what are as vexing as they are interesting.

One of the components of automaticity, controllability, usually refers to influencing the *expression* of mental contents. However, researchers also are interested in the controllability of content that gets into the mind in the first place. One motivation for people to dismiss implicit cognitions as cultural knowledge may be to retain a sense of control over what gets into and becomes their minds. Are there distinct implications for mental contents that are acquired deliberately versus those that are acquired passively? Is one more or less "me"? Future research may clarify whether control is more than an outmoded foundation on which to build a sense of self.

The theoretical and practical interest in implicit cognition spurred an implicit-measures design industry. Existing evidence suggests that the varieties of measures bearing the "implicit" label have heterogeneous underlying processes. This review

considered only the IAT; clarifying the underlying cognitive processes across measures will enrich the current understanding of IE relations (Strack & Deutsch, 2004).

Three additional issues are underrepresented in the IE-relations literature: (a) evidence for or against stable individual differences that moderate IE relations; (b) examination of the similarity of moderators across constructs (e.g., attitudes, stereotypes, identity, goals); and (c) the relevance of IE relations for their conjoint or discriminant predictive validity. Investigations of these questions will provide the basis for a comprehensive theory of the relationship between implicit and explicit cognition.

Mental operations and subjective experience of those operations are distinct but related. Understanding the relationship between the mind and humans' experience of it is one of the great projects of psychological science.

Recommended Reading

- Bargh, J.A. (1994). (See References)
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford.
- Greenwald, A.G., & Banaji, M.R. (1995). (See References)
- Wilson, T.D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.
-

Acknowledgments—This research was supported by Grant R01 MH68447 from the National Institutes of Health.

REFERENCES

- Banaji, M.R., Nosek, B.A., & Greenwald, A.G. (2004). No place for nostalgia in science: A response to Arkes & Tetlock. *Psychological Inquiry*, *15*, 279–289.
- Bargh, J.A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R.S. Wyer, Jr. & T.K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1–40). Hillsdale, NJ: Erlbaum.
- Greenwald, A.G., & Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4–27.
- Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
- Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.
- Hofmann, W., Gschwendner, T., Nosek, B.A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology*, *16*, 335–390.
- Karpinski, A., & Hilton, J.L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*, 774–788.
- Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Nosek, B.A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*, 565–584.
- Nosek, B.A., & Hansen, J.J. (in press). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit cognition. *Cognition and Emotion*.
- Nosek, B.A., & Smyth, F.L. (2007). A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. *Experimental Psychology*, *54*, 14–29.
- Poehlman, T.A., Uhlmann, E., Greenwald, A.G., & Banaji, M.R. (2006). *Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity*. Unpublished manuscript, Yale University.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*, 220–247.