Meta-Analytic Use of Balanced Identity Theory to Validate the Implicit Association Test

Dario Cvencek [a1], Andrew N. Meltzoff [a], Craig D. Maddox [a], Brian A. Nosek [b], Laurie A. Rudman [c], Thierry Devos [d], Yarrow Dunham [e], Andrew S. Baron [f], Melanie C. Steffens [g], Kristin Lane [h], Javier Horcajo [i], Leslie Ashburn-Nardo [j], Amanda Quinby [k], Sameer B. Srivastava [l], Kathleen Schmidt [m], Eugene Aidman [n], Emilie Tang [o], Shelly Farnham [p], Deborah S. Mellott [q], Mahzarin R. Banaji [r], Anthony G. Greenwald [o]

[a] Institute for Learning & Brain Sciences (I-LABS), University of Washington, Box 357988, Seattle, WA; [b] Department of Psychology, University of Virginia, Box 400400, Charlottesville, VA, 22904; [c] Department of Psychology, Rutgers University, 53 Avenue E, Piscataway, NJ, 08854; [d] Department of Psychology, San Diego State University, 5500 Campanile Dr., San Diego, CA, 92182; [e] Department of Psychology, Yale University, Box 208205, New Haven, CT 06520; [f] Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC, V6T 1Z4, Canada; [g] Department of Psychology, University of Koblenz-Landau, Fortstr. 7, D-76829 Landau, Germany; [h] Department of Psychology, Bard College, 30 Campus Road, Annandale-on-Hudson, NY, 12504; [i] Department of Psychology, Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049 – Madrid, Spain; [j] Department of Psychology, Indiana University – Purdue University Indianapolis, 402 North Blackford Street, Indianapolis, IN, 46202; [k] Bloomington Counseling and Wellness Services, 205 North College Avenue, Suite 512, Bloomington, IN, 47404; [l] Haas School of Business, University of California Berkeley, 2220 Piedmont Avenue, Berkeley, CA, 94720; [m] Department of Psychology, Southern Illinois University, Life Science II 6502, Carbondale, IL, 62901; [n] School of Psychology, The University of Sydney, Brennan MacCallum Building (A18), NSW, 2006, Australia; [o] Department of Psychology, University of Washington, Box 351525, Seattle, WA, 918195; [p] Third Place Technologies, 1216 E. John St, Seattle, WA, 98102; [q] Seattle, WA; [r] Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, 02138.

Author contributions: All listed authors contributed data; D.C., C.D.M., and A.G.G. analyzed data; D.C., A.G.G., A.N.M., and C.D.M. wrote the paper.

Abstract (143 words)

This meta-analysis evaluated theoretical predictions from balanced identity theory (BIT) and

evaluated the validity of zero points of IAT and self-report measures used to test these predictions.

Twenty-one researchers contributed individual-subject data from 36 experiments (total $N = 12,773$)

that used both explicit and implicit measures of the social–cognitive constructs. The meta-analysis

confirmed predictions of BIT's balance–congruity principle and simultaneously validated

interpretation of the IAT's zero point as indicating absence of preference between two attitude

objects. Statistical power afforded by the sample size enabled the first confirmations of balance–

congruity predictions with self-report measures. Beyond these empirical results, the meta-analysis

introduced a within-study statistical test of the balance–congruity principle, finding that it had

greater efficiency than the previous best method. The meta-analysis's full data set has been publicly

archived to enable further studies of interrelations among attitudes, stereotypes, and identities.

Keywords: attitudes, stereotypes, identity, self-esteem, Implicit Association Test

"Meta-Analytic Use of Balanced Identity Theory to Validate the Implicit Association Test"

Balanced identity theory (BIT) was developed as an account of relationships among several of social psychology's most prominent theoretical constructs—identities, attitudes, stereotypes, and self-esteem. The theory (Greenwald et al., 2002) drew on principles originating in social psychology's consistency theories of the late 1950s, especially Heider's (1958) balance theory, Osgood and Tannenbaum's (1955) congruity theory, Newcomb's (1953) symmetry theory, and Festinger's (1957) cognitive dissonance theory.

BIT's main theoretical devices are (a) its definitions of identities, attitudes, stereotypes, and self-esteem as *associations*[2] involving self, groups, stereotypic attributes, and valence, and (b) the *balance–congruity principle*—the proposition that an association between concepts A and B should strengthen to the extent that each of A and B is associated with the same third concept, C. The balance–congruity principle is a close relative of the concept of *mediated generalization*, first described by Cofer and Foley (1942). Its name acknowledges its additional roots in the affective–cognitive consistency theories of Heider (1958; balance theory) and Osgood and Tannenbaum (1955; congruity theory).

Historical Background

Balance theory and congruity theory sought to explain how naturally arising affective–cognitive configurations induce further affective–cognitive changes. For example, a positive attitude toward a group will not lead to identification if self-esteem is negative, and will occur in proportion to the positivity of self-esteem (i.e., more when self-esteem is strongly positive than when it is moderately or weakly positive). The balance–congruity principle extended those prior theories to explain the collection of implicit self-esteem effects that were identified by Greenwald

---

[2] As explained by Greenwald, Nosek, Banaji, and Klauer (2005), BIT's appeal to associations draws on a long-established "theory-uncommitted" understanding of association that implies nothing more complex than a (physiologically unspecified) link that can allow one mental concept to activate another.

and Banaji (1995), including minimal group effects (Tajfel, Billig, Bundy, & Flament, 1971),

endowment effects (Kahneman, Knetsch, & Thaler, 1990), mere ownership effects (Beggan, 1992),

self-anchoring effects (Cadinu & Rothbart, 1996; Gawronski, Bodenhausen, & Becker, 2007; Roth

& Steffens, 2014), implicit self-referencing (Perkins & Forehand, 2006; 2012), and implicit self–

object linking (Ye & Gawronski, 2016). These are phenomena in which newly created associations

between the self and social or non-social objects produce an associative transfer of self-evaluation

(i.e., self-esteem) to those self-linked objects.

BIT compared to Heider's balance theory. Both BIT and Heider's (1958) balance theory

predict social knowledge to be organized in ways that maintain affective–cognitive consistency: In

balance theory, consistency was conceived in terms of cognitive structures that link a focal person

("p") to other persons ("o") or external objects ("x") via either *sentiment* (attitude) or *unit*

relationships (Heider, 1958). By replacing Heider's distinct sentiment and unit connections with the

more general conception of *association* as the cognitive link between persons and other entities,

BIT was able to expand theoretical scope beyond the attitudes with which balance theory was

concerned, including additionally stereotypes, identities, and self-esteem.

BIT compared to social identity and self-categorization theories. BIT and two other

well-established theories on social identity—Turner's self-categorization theory (SCT; Turner,

Hogg, Oakes, Reicher, & Wetherell, 1987) and Tajfel's social identity theory (SIT; Tajfel, 1982)—

allow consideration of identities in relation to self-esteem. Relationships among self-esteem, group

membership, and in-group attitude are considered by all three theories (BIT, SCT, SIT). The three

theories agree in expecting that persons with a strong in-group identity should have a stronger

positive attitude toward their group (i.e., in-group attitude) than those with weak in-group identity.

A substantial difference between BIT and SIT arises from the difference in the way self-

esteem is conceived in the two theories. SIT conceives self-esteem as a fundamental human need

(Tajfel & Turner, 1979), whereas BIT conceives self-esteem non-motivationally as an association of

the concept of one's self with positive valence. This leads to substantial differences in how the two

theories see the relation of self-esteem to formation of a novel identity. This was investigated in

numerous experiments in which subjects were assigned to one of two previously unknown groups

that differed in meaningless or arbitrary aspects. Tajfel et al. (1971) interpreted the repeated finding

of subjects evaluating their own group more positively than the other group as a cognitive strategy

occurring because subjects could achieve a boost in self-esteem by conceiving their group as the

superior one (see also Abrams & Hogg, 1988; Hogg, 2000). The theoretical expectation was that

this minimal group effect should occur most noticeably in subjects who had relatively low self-

esteem, who would have stronger need for the self-esteem boost. In BIT, the minimal group

phenomenon occurs as a consequence of the association of self with both the novel group and with

positive valence. The effect should be greater for those for whom self-esteem is already high, rather

than for those with low self-esteem. The meta-analysis of 34 studies by Aberson, Healy, and

Romero (2000), examining the relation between self-esteem and in-group bias, favored the

relationship predicted from BIT, rather than the one predicted from SIT (see also Hewstone, Rubin,

& Willis, 2002).

Methods for Evaluating the Balance–Congruity Principle

Implicit and Explicit Measures

This article examines evidence available to assess validity of BIT's balance–congruity

principle in studies using either explicit (direct) or implicit (indirect) measures (cf. Fazio & Olson,

2003). Explicit measures generally use self-report, allowing research subjects to be aware of what is

being investigated. In contrast, implicit measures do not use self-report and do not require the

subject to know the nature of the construct being assessed, which might be an attitude, a stereotype,

an identity, or self-esteem. A recent treatment comparing the two types of measures is available in Greenwald and Lai (2020).

Implicit Association Test (IAT)

The focus of this meta-analysis is on the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). In the 20 years since its initial publication, the IAT has been applied in a diverse array of disciplines including social and cognitive psychology (Axt & Lai, 2019; Critcher & Ferguson, 2016), neuroscience (Mitchell, Ames, Jenkins, & Banaji, 2009; Schindler, Wolff, Kissler, & Brand, 2015), education (Cvencek, Nasir, O'Connor, Wischnia, & Meltzoff, 2015; Devos & Cruz Torres, 2007; Nosek et al., 2009), developmental science (Baron & Banaji, 2006; Cvencek, Meltzoff, & Greenwald, 2011), clinical psychology (Creemers, Scholte, Engels, Prinstein, & Wiers, 2013; Leeuwis, Koot, Creemers, & van Lier, 2015), health psychology (Cooper et al., 2012; Sabin & Greenwald, 2012), and marketing research (Horcajo, Briñol, & Petty, 2010; Trendel, Mazodier, & Vohs, 2018). The wide range of applications provides sufficient research literature from which to review the strength of the evidence for the support of the balance–congruity principle.

The IAT is a method for indirectly measuring the strengths of associations among concepts. In its data-providing *combined-task* blocks, the IAT requires sorting of stimulus exemplars belonging to four different categories using just two response keys, each of which is assigned to two of the four categories. The underlying principle of the IAT is that it is easier to give the same response to items representing categories that are associated in memory than to ones representing categories that are not associated. For example, the assignment of items representing *summer* with *warm* to the same key should be easier for subjects to deal with than the assignment of items representing *summer* and *cold* to the same key.

The two *combined tasks* of a standard IAT most typically include two concept categories (e.g., *summer* and *winter*) and two attribute categories (e.g., *warm* and *cold*). In each combined-task block, there is a strict alternation between concept category exemplars (typically on odd-numbered trials) and attribute category exemplars (on even-numbered trials). In one of the two combined tasks, exemplars of *summer* and *warm* (e.g., images of beaches and words such as "hot" and "sunny") require pressing one of two response keys (generally positioned left and right on a computer keyboard) for a correct response, and exemplars of *winter* and *cold* (e.g., images of snow and words such as "freeze" and "icy") require response with the other key. In the second critical block, exemplars of *summer* and *cold* are sorted with one response key, and exemplars of *winter* and *warm* are sorted with the alternative response key. The faster the responses, the stronger the presumed underlying association between the two categories sharing the same key. For participants who possess the expected stronger associations of concept *summer* with attribute *warm* and *winter* with *cold*, the first sorting task will likely be much easier than the second. Ease of sorting is indexed by the speed of producing correct responses. Most IAT procedures oblige occurrence of a correct response to end every trial; when trials are permitted to end on occurrence of an incorrect response, a time penalty for trials on which errors occurred is applied in the scoring procedure (described by Greenwald, Nosek & Banaji, 2003, p. 214).

Aims of the Research

Despite the centrality of identities, self-esteem, and in-group attitudes in social psychology, there were no studies of affective–cognitive consistency among these constructs prior to formulation and the first tests of balanced identity. This study examines the evidence for affective–cognitive consistency in naturally arising-configurations of these constructs, comparing the strength of that evidence for implicit and explicit measures. The available quantitative evidence was first examined via meta-analytic hypothesis test using the established standard approach for testing

BIT's balance–congruity principle. The meta-analysis additionally allowed testing a novel, within-study meta-analytic method that was found to be more efficient than the already established method. Both approaches included (a) analyses of studies using self-esteem measures alongside studies using other self-concept measures, and (b) subject-level data from each study in the meta-analysis to assure use of the same analysis methods for all studies.

Meta-Analytic Evaluation of the Balance–Congruity Principle

Cvencek, Greenwald, and Meltzoff (2012) found support for balance–congruity principle expectations in a review of studies including about 1,900 subjects. The present meta-analysis increases the number of subjects available for study by a factor of 6. Additionally, by obtaining individual-subject IAT and self-report measures (where available) from authors for all 36 samples reviewed in this article, it was possible to use the same analysis method for all studies. A further contribution of this paper is that we are making the full data set, consisting of 12,773 participants (ranging from young children to adults) across 36 studies publicly available in a widely accessible archive.

Search Method for Locating Balanced Identity Data Sets

PsycINFO, PubMed, and Google Scholar were searched, using as individual keywords/phrases: *cognitive balance*, *cognitive consistency*, *balanced identity*, *IAT*, *Implicit Association Test*, *implicit attitude*, *implicit identity*, *implicit self-esteem*, *implicit stereotype*, *implicit self-concept*, *3 IATs*, and *3 Implicit Association Tests*. The cut-off date for the search was May 31, 2013. PsycINFO was also used to find studies that referenced Greenwald et al. (2002). These searches identified 19 reports. The IAT scores used in this meta-analysis are the exact ones that appear in each of the published reports.[3] Authors of these 19 reports were then contacted in search

---

[3] In a balanced identity analysis of trios of measures, there are eight ($= 2^3$) possibilities for direction of scoring of the three measures. Four of these combinations are ones for which the balance–congruity principle predicts that the product

of additional studies, yielding eight more reports. The resulting 27 reports included 36 independent

samples, with a total of 9,808 subjects providing data for IAT measures and 12,773 providing data

for explicit measures. Table 1 describes the 36 samples. These samples include both male and

female participants, multiple ethnic groups (e.g., Asians, Blacks, Latinos, Whites), and multiple age

groups (e.g., pre- and elementary school children, undergraduates, adults, elderly). In addition (see

Table 1), the studies in this meta-analysis included measures of attitudes toward both social (e.g.,

gender, race, age) and non-social categories (e.g., weight, food), along with multiple stereotypes

(e.g., math, parenthood, achievement).

The research reviewed in this manuscript was done at 16 institutions and was published in

17 peer-reviewed journals.[4] Part of the explanation for there not being more such studies is that

these studies are effort-demanding. Seventeen of the 36 studies required locating and recruiting

non-standard subject populations (see the "Participants" column in Table 1). Authors were also

obliged to create novel IATs for 17 of the 36 studies. Five of the 36 studies included in the meta-

analysis had unpublished data sets.

Use and Evaluation of BIT

Each association in a balanced identity study is embedded in an associative network that

includes many other associations, as indicated in following Equations SG1, SA1, and GA1. The

three associations of a balanced identity research design are represented in these equations by SG

(self–group association), SA (self–attribute association), and GA (group–attribute association).

Each association in the design is embedded in multiple trios of associations (see Figure 1). The

specific concepts included in a study's "focal" trio of measured associations are indicated in the

---

of two of the measures should correlate positively with the third measure. For the other four, expectations are for
negative correlations. By convention, sets of measures are scored with one of the four combinations for which positive
correlations are expected. All data sets in the present meta-analysis were so scored.

[4] Most of these data come from the original researchers who formulated BIT, as well as generations of their
collaborators, students, and post-doctoral fellows.

equations with filled-triangle subscripts. Additional concepts that enter the prediction of the

(criterion) measure on the left side of each equation have numerical subscripts. The *b* coefficients in

the three equations represent weights that, in principle, could be empirically assessed with

regression analyses—if measures were available for all associations in the equation. In practice,

however, measures are not available for associations linking self and group to other attributes (the

other As in SG1), self and attribute to other groups (the other Gs in SA1), or group and attribute to

persons other than self (the Ps in GA1). The ellipsis ("…") that ends each equation acknowledges

the indefinite multiplicity of such additional groups, attributes, and persons that may contribute to

the strength of the criterion association. Nevertheless, the first predictor in each equation (e.g.,

$S_{▲}A_{▲}·G_{▲}A_{▲}$ in Equation SG1) should be the strongest predictor and, consistent with balance–

congruity principle, should be correlated with other predictors in the equation.

$$S_{▲}G_{▲} = b_{▲}·[S_{▲}A_{▲}]·[G_{▲}A_{▲}] + b_1·[S_{▲}A_1]·[G_{▲}A_1] + b_2·[S_{▲}A_2]·[G_{▲}A_2] + b_3·[S_{▲}A_3]·[G_{▲}A_3] + \ldots \qquad \text{(SG1)}$$

$$S_{▲}A_{▲} = b_{▲}·[S_{▲}G_{▲}]·[G_{▲}A_{▲}] + b_1·[S_{▲}G_1]·[G_1A_{▲}] + b_2·[S_{▲}G_2]·[G_2A_{▲}] + b_3·[S_{▲}G_3]·[G_3A_{▲}] + \ldots \qquad \text{(SA1)}$$

$$G_{▲}A_{▲} = b_{▲}·[S_{▲}G_{▲}]·[S_{▲}A_{▲}] + b_1·[P_1G_{▲}]·[P_1A_{▲}] + b_2·[P_2G_{▲}]·[P_2A_{▲}] + b_3·[P_3G_{▲}]·[P_3A_{▲}] + \ldots \qquad \text{(GA1)}$$

The 4-Test Method

Testing the balance–congruity principle requires a statistical method to evaluate the

predictions involving just the three focal variables in Equations SG1, SA1, and GA1 (e.g., $S_{▲}G_{▲} =$

$b_{▲}·[S_{▲}A_{▲}]·[G_{▲}A_{▲}]$ in Equation SG1). Greenwald et al.'s (2002) 4-test method depended on an

assumption that these three associations were measured (at least to a good approximation) on

interval scales that had rational zero points. In the first decade after the 2002 publication, the two

measurement assumptions (interval scales and rational zero point) were plausible and were

consistent with observed data, but they were not yet empirically testable with any precision.

Subsequent accumulation of the data reported in this meta-analysis made possible the development

of more precise empirical tests that are reported in this article. Detailed overview of the 4-test

method, as well as the results of analyses applying the 4-test method to the current data set can be

found in the Supplemental Materials.

An Alternative to the 4-Test Method: Within-Study Meta-Analysis

One limitation of the 4-test method is its cumbersomeness: It requires a computation of 12

statistical tests—four in each of the 2-step regressions for each of three measures predicted by the

product of the other two. A second limitation is that the 4-test method provides no quantitative

indicator of magnitude of confirmation of the balance–congruity principle. A third limitation is an

increase in possibly spurious confirmations when the two product-component predictors (a) are

additively (not multiplicatively) correlated with the criterion *and* (b) both have means deviating

from zero by more than 1 SD in the positive direction (see Greenwald, Rudman, Nosek, & Zayas,

2006, Figure 1E). This last difficulty is due to collinearity of two individual predictors that are

positively correlated with both (a) the regression's criterion measure and (b) their own product.

Despite these three limitations, the 4-test method is superior to the traditional simultaneous multiple

regression method in detecting the presence of a pure multiplicative relationship (see Greenwald et

al., 2006, Figure 2).

Seeking a possibly superior alternative to the 4-test method, this meta-analysis's data were

used to evaluate a new method that combined the three correlation effect sizes produced by Test 1

in each sample and (separately) those produced by Test 2 (see Supplemental Material for details).

For each of the three types of criterion measure (SG, GA, and SA), separately for self-report and

IAT measures, the three effect sizes were transformed to Fisher Z values and then aggregated in a

random-effects, within-study meta-analysis.

The within-study meta-analytic summaries of Tests 1 and 2 were then themselves meta-

analytically combined (separately for IAT and self-report and, within those, separately for Tests 1

and 2) across all studies in the meta-analysis.[5] The within-study meta-analyses used only Tests 1 and 2 because (a) the *r* associated with Test 1 can be interpreted as a basic test of fit of a multiplicative model, and (b) the *pr* associated with Test 2 can be interpreted as an index of fit of a *pure* multiplicative model. The detailed results of the within-study meta-analysis method can be found in the Supplemental Materials. This novel within-study meta-analysis method provides what may be a more efficient indicator of conformity to BIT predictions than the 4-test method provides (see Supplemental Materials for details).

Evaluation of Within-Study Meta-Analysis Method in Comparison With 4-Test Method

The 4-test method provides the existing standard indicator of conformity of data from the three measures of a balanced identity design to the multiplicative model prediction of BIT's balance–congruity principle. Predicting the number of tests passed for each study (maximum of 12), the Test 1 within-study aggregate was entered on Step 1 of a 2-step regression, and the Test 2 within-study aggregate was entered on Step 2. This weighted 2-step regression found that the within-study Test 1 aggregates were *not* significant predictors of total number of tests passed, while the Test 2 aggregates were significant predictors; partial *r* for Test 2 in the second step was $pr = .383$, $t(33) = 2.38$, $p = .02$. The interpretation of these findings for drawing conclusions about usefulness of the within-study meta-analysis method is considered in Discussion.

Comparing Studies Using Self-Esteem Measures With Those Using Other Self-Concept Measures

Available evidence for validity of IAT measures of self-esteem is limited (Bosson, Swann, & Pennebaker, 2000; Greenwald & Farnham, 2000), with some of the strongest evidence coming from empirical tests of the balance–congruity principle. Within the BIT framework, self-esteem can

---

[5] The within-study meta-analytic aggregations for each of Tests 1 and 2 produce (appropriately) only one datum for each study. The variance in these observations across studies was treated as random effects variance—as was done for the other meta-analytic statistical tests reported in this article.

be distinguished from another class of social–cognitive constructs involving the category "self":

self-concepts (or identities). According to the original formulation of BIT, "self-esteem is the

association of the concept of self with a valence attribute," whereas "self-concept [or identity] is the

association of the concept of self with one or more (nonvalence) attribute concepts" (Greenwald et

al., 2002, p. 5). The present meta-analysis affords an opportunity to compare evidence from studies

involving valence (i.e., self-esteem measures as SA measures) with those involving other attributes

(i.e., self-concept measures as SA measures). The available evidence was compared in two ways.

First, the *aggregated mean* outcomes of Tests 1 and 2 of the 4-test method for these two groups of

samples were compared meta-analytically. Second, the *within-study meta-analysis* method was

applied to both self-esteem ($k = 22$) and self-concept ($k = 14$) measures (see Supplemental

Materials).

The results showed that confirmations for self-esteem measures were mostly comparable to

the self-concept measures: For implicit measures, the difference between the weighted aggregate

effect sizes of self-esteem and self-concept measures was statistically significant for Test 2 (p

$=.005$), but not for Test 1, $p > .26$; for self-report measures, this difference was not statistically

significant for either Test 1 or Test 2 (ps $>. 11$). In addition, the level of support for balance–

congruity from studies involving valence (i.e., self-esteem rather than self-concept) was substantial

(see Supplemental Materials for details). Implications of these findings for understanding validity of

IAT and self-report measures of self-esteem are considered further in Supplemental Materials.

Interpretation of Zero Points of Attitude and Stereotype Measures

The IAT's Zero-Point Assumption

The first use of the balance–congruity principle was to test the prediction that a woman

possessing both an association of *self* with *female* (in-group identity) and an association of *self* with

positive valence (positive self-esteem) should be expected also to have a positive in-group attitude

(favorable toward female; Greenwald et al., 2002, p. 10). In this self–gender–valence balanced

identity design, the balance–congruity principle predicts that the association of *female* with *positive*

valence should be strengthened if *both* female and positive valence are associated with *self* (BIT's

"shared first-order link" configuration; Greenwald et al., 2002, p. 6). If *either* the self–female or the

self–positive association is zero, no strengthening is predicted. Further, strengthening should be

greater when *both* associations in the shared first-order link are strong. Tests of this multiplicative

prediction require that the association used in the test are measured on scales that have valid zero

values (Greenwald et al., 2002, p. 11). The dependence of this prediction on a valid interpretation of

the zero point of measures used to test the prediction is a main topic of this article.

Two Types of Zero Points

For attitude measures, two types of zero points can be useful. The more intuitive zero point

is one that indicates *neutrality* (absence of attitudinal valence). An attitude can be said to be absent

or neutral when a person's evaluation of the attitude's object has neither negative nor positive

valence. A 7-point self-report item to assess this understanding of zero might range from a value of

–3 (labeled "strong dislike," indicating negative valence) to +3 (labeled "strong liking," indicating

positive valence), with a midpoint of 0 (labeled "neutral," or "neither like nor dislike"). Evaluation

items using the semantic differential method (Osgood, Suci, & Tannenbaum, 1957) are of this type.

The second type of zero point indicates *indifference* (lack of preference) between two

contrasted attitude objects. In a 7-point self-report item, the end points for an item that can assess

the indifference-meaning of zero might range from –3 (labeled "strongly prefer A relative to B") to

+3 (labeled "strongly prefer B relative to A") with a midpoint of 0 (labeled "equal liking of A and

B"). Items with the indifference-indicating zero point are useful in investigations of choice among available alternatives, such as pre-election polls.[6]

IAT attitude measures allow only the indifference interpretation of zero, indicating no preference. More conceptually stated, the zero value of an attitude IAT (i.e., one in which the attribute category contrast is pleasant vs. unpleasant or good vs. bad) indicates lack of difference in strengths of associations of the contrasted concept categories (e.g., White vs. Black race) with positive or negative valence. This zero value is obtained when a research subject performs equally rapidly on the attitude IAT's two combined tasks. In a stereotype IAT, the IAT's zero value indicates lack of difference in strengths of associations of two contrasted attribute categories (e.g., career vs. family) with the two contrasted concept categories (e.g., female vs. male).

Studies that report data for both IAT and self-report measures of intergroup attitudes or stereotypes typically find greater proportions of respondents showing biases on the IAT measure than on its parallel (i.e., indifference-zero format) self-report measure. For example, in a large study that included measures of attitudes toward White and Black races, approximately 20% more people showed White-race preference on the IAT measure than on the parallel self-report measure (Nosek et al., 2007). Such findings call for an explanation for why IAT and self-report preference measures differ in this fashion. The most favorably regarded explanation for this difference is that IAT and self-report measures are based on different types of mental representation (perhaps associations versus propositions, as suggested by Strack & Deutsch, 2004). A second favorably regarded explanation is that zero points of self-report measures may be distorted by respondents' desires to

---

[6] In self-report measurement of attitudes, a thermometer-type valence format may be administered for each of two alternatives, A and B. The thermometer end anchors might be –5 (extremely cold) and +5 (extremely warm), with 0 labeled "neither warm nor cold". The B minus A difference between the two thermometer responses provides a preference-type measure with a 21-point range, from –10 indicating maximum preference for A to +10 indicating maximum preference for B, and 0 indicating absence of preference. Used in such pairs, thermometer items can assess both the valence-absence and the indifference zero. The indifference zero doesn't require that either item have a zero value. It does require that both items have the same numerical value.

appear unprejudiced (e.g., Greenwald et al., 2002). A third explanation is that zero points of IAT

measures may be distorted due to characteristics of the IAT's procedure (Blanton, Jaccard, Strauts,

Mitchell, & Tetlock, 2015).[7]

Methods to Evaluate Interpretations of the IAT's Zero Point

All three of the preceding paragraph's explanations could be correct. Available empirical

evidence does not rule any of them out. This article evaluates specifically the third explanation—the

one based on presumed invalidity of zero points of IAT measures, which could be tested using a

new method applied to individual-subject data. The method displaces zero points of predictors by

adding or subtracting constants, followed by observing the extent to which tests of BIT's balance–

congruity principle are (or are not) impaired by the displacements. If the IAT or self-report

measures used in these predictors have valid zero points, these zero-point displacements should

impair the previously reported support for the balance–congruity provided by the (already reported)

undisplaced tests of those predictions. Relatedly, increasingly large zero-point displacements should

produce increasingly large impairments of that support.

This article's use of the zero-point displacement method may be appreciated by considering

its relation to the entirely familiar use of multiplication for adjacent side lengths in computing the

area of a rectangle. Like this article's correlational Test 1 for balance–congruity predictions, valid

use of Euclidean geometry to compute rectangle areas from rectangle side lengths requires that

those length measures have valid zero values. Therefore, increasing magnitudes of displacements of

---

[7] Blanton et al. (2015) offered a method to empirically evaluate validity of zero points of attitude measures. Their method examined regressions of IAT attitude measures on correlated self-report or behavioral measures, which they used as univariate predictors. Their test for a valid zero point of the IAT measure was to determine whether the computed intercept (i.e., the IAT value associated with zero of the predictor measure) was at or very near zero. Blanton et al. did not consider the known impact of error of measurement of regression predictors on regression intercept estimates, nor did they consider the role of the magnitude of predictor–criterion correlation as an influence on intercept values. (These statistical problems with their method are described more fully in the Supplemental Materials.)

zero values in side-length measures should increasingly impair the accuracy of area values

computed using those zero-displaced measures.

Figure 2A presents results of applying the zero-point displacement strategy to the meta-

analysis's IAT data. Figure 2A reports the observed data, along with two simulations, one based on

the unrealistic assumption of perfect reliability of measures used in the test, and one based on the

realistic assumption that IAT measures have average reliability of $r = .50$.[8] Figure 2B does likewise

for the meta-analysis's self-report data, using the conservative assumption of test–retest reliability

of $r = .80$. These reliability simulations assume the (likely unrealistic) assumption that the *only*

determinant of Test 1's results is the multiplicative effect predicted by BIT's balance–congruity

principle.

The IAT results showed the expected decline in support as a function of magnitude of zero-

point displacement with a maximum at the value of zero displacement. Remarkably, the curve for

the obtained data was very close to the values expected from the assumption that test–retest

reliability was the only source of noise in the findings. For self-report results, the decline as a

function of displacement was noticeable, but the observed data were quite far from those expected if

test–retest reliability was the only source of noise. Combining (a) the finding that support for the

BIT prediction was generally weaker for self-report than for IAT with (b) the finding that this

support with self-report was much further below expectation based on expected test–retest

reliability, there is strong support for the conclusion that IAT measures come closer than do self-

report measures to measuring the constructs described by balanced identity theory.

---

[8] The estimated reliability of $r = .50$ is the value reported in the meta-analysis of 58 published reports of the IAT's test–retest reliability reported in Table 2 of Greenwald and Lai (2020).

Additional Zero-Point Tests

An additional set of measures provided data for which the balance–congruity principle's

predictions depended on validity of zero points: a *correlation scatterplot prediction* for positive

slope and zero intercept for the regression of the correlation between a subset of two of each study's

three measures (used as criterion) on the mean of the third measure. The correlation scatterplot

prediction was described by Greenwald et al. (2002):

> When [the mean of] any variable in the balanced identity design is polarized toward its high end, the
>
> zero-order correlation between the other two variables should be positive; when any of the variables is
>
> polarized toward its low end, the zero-order correlation between the other two variables should be
>
> negative; and if a variable in the balanced identity design is not polarized, correlations between the other
>
> two variables should not differ from zero (p. 11).

The trio of measures in each study can be sorted into one *self–attribute* association (SA),

one *self–group* association (SG), and one *group–attribute* (GA) association. Correlations between

any two of these should be predicted by the mean of the third. For example, correlations between

SG and SA should be predicted by GA. If GA has a negative mean, the SG–SA correlation should

be negative, if GA has a positive mean the SG–SA correlation should be positive, and if GA has a

mean of zero the SG–SA correlation should be zero. The test of the set of these predictions comes

from examination of the scatterplot in which all such correlations for each type of measure (IAT or

self-report) are plotted as a function of the varying means of the third measures of each trio.[9] This

scatterplot is presented for the meta-analysis's IAT measures in Figure 3A and for the meta-

analysis's self-report measures in Figure 3B.

---

[9] This prediction was explained by Greenwald et al. (2002, p. 10) in their Figure 6 and two accompanying text
paragraphs. In that explanation, the balance–congruity principle and the IAT's zero-point assumption were combined to
predict that "the slope of the regression relation between any two variables (e.g., criterion and Predictor A) is governed
by the level of the third variable (Predictor B). When the third variable is at a high level, the expected relationship
between the first two variables is positively sloped; when the third variable is at a low level, the expected relationship
between the first two is negative."

Figure 3's scatterplots combine, separately for IAT and self-report measures, the scatterplots for $r_{\text{SG–SA}}$ predicted by M_{GA}, $r_{\text{SG–GA}}$ predicted by M_{SA}, and $r_{\text{GA–SA}}$ predicted by M_{SG}. These analyses improve in two ways on the only previous test (available in Figure 8.3 of Cvencek et al., 2012). First, the substantially larger numbers of studies and subjects in the present tests substantially increases power and precision. Second, the new analysis manages the treatment of scoring direction of IAT measures in a way that adds substantially to the statistical power and precision of regression intercept estimates.[10]

For IAT measures, the 19 regression scatterplots, each combining all 108 (= 3 X 36) correlation values (Figure 3A) all had strongly positive slopes, corresponding to correlations between .83 and .85. The 95% confidence intervals for 17 of the 19 regression intercepts included the origin, and the widths of these CIs ranged from .049 to .053 on the Fisher Z scale used for the vertical axis. The scatterplot for the sample with the median of the 19 intercepts (–0.004) is displayed in Figure 3A. The data are remarkably consistent with the pattern expected from the balance–congruity principle's prediction, which is not expected unless the measures being used have valid zero points.

The method of Figure 3A was applied also to the 48 correlations that were available for self-report measures in Figure 3B, producing 19 regressions and selecting the one with the median intercept for display. These 19 regressions also had positive slopes, corresponding to correlations ranging from .49 to .56. Fifteen of the 19 had 95% CIs that included the origin. The widths of those CIs ranged from .121 to .131 on the Fisher Z scale. Figure 3B's data are therefore also consistent

---

[10] The problem addressed for the first time in this analysis is a method of dealing with the partial arbitrariness of direction of scoring measures used in tests of balance–congruity predictions, described previously in Footnote 2. A symmetric distribution of predictor means around zero affords greater precision in estimating regression intercepts. This was achieved for Figure 3's plots by reversing both the sign of the predictor variable and the sign of the correlation between the other two variables for a random half of the samples. Scatterplots were created for 19 iterations of this analysis, with randomization for each sample done independently in each iteration. The scatterplots with the median regression intercepts for IAT and self-report data were selected for presentation in Figures 3A and 3B.

with validity of zero points for the collection of 48 self-report measures used in the present research, but suggest that, even with a measure that has relatively high test–retest reliability, there may be more individual subject variability in proximity of measures' zero points to the desired indifference meaning of zero.

General Discussion

This review quantitatively assessed results obtained in studies of both IAT and self-report measures to evaluate both the balance–congruity principle of balanced identity theory (BIT; Greenwald et al., 2002) and the validity of zero points of IAT measures. In doing so, this review established four new findings.

First, the review found that predictions from BIT's balance–congruity principle are confirmed not only for IAT measures, but also for self-report measures. In every way in which conformity to the balance–congruity principle could be compared for IAT and self-report measures, results revealed stronger confirmation of predictions with IAT than with self-report measures. In retrospect, the previously observed lack of confirmation for self-report measures (Cvencek et al., 2012) is most plausibly attributed to the lesser statistical power available in previously analyzed data sets.

Second, this review developed and reported a within-study meta-analytic test of the balance–congruity principle that is more efficient than the previously standard 4-test method. The within-study method not only reduced 12 statistical tests to two, but provided an index of fit to a pure multiplicative model.

Third, this review reported the first tests of the assumption that zero points of IAT and self-report attitude measures are validly interpretable as indicating absence of preference for one of (i.e., indifference between) two alternative concepts contrasted in the test. This valid zero value is required not only for tests of the balance–congruity principle, but also for meaningful interpretation

of IAT measures.[11] Results of these tests (presented in Figures 2 and 3) consistently supported

validity of the zero values, and did so more strongly for IAT than for self-report measures.

Fourth, this review confirmed the balance–congruity principle for the subset of studies that

used self-esteem IAT measures, separately from confirming it for studies that used self-associations

other than self-esteem. This finding for studies using IAT measures of self-esteem is useful,

considering that the validity of measures of both explicit and implicit self-esteem has been

questioned in the published literature (Baumeister, Campbell, Krueger, & Vohs, 2003; Bosson et al.,

2000; Buhrmester, Blanton, & Swann, 2011).

Pure Multiplicative Model?

The introduction of this article explained that the three associations examined in any test of

the balance–congruity principle are embedded in a larger associative network (see Figure 1) that

includes other associations that can influence strengths of the three focal associations (see Equations

SG1, SA1, GA1). The involvement of each of the three focal associations in multiple triads of

associations, only one of which is assessed in each study in the meta-analysis, necessarily

diminishes the expectation that the data for any individual balanced identity study's trio of

associations should conform in *pure* fashion to the multiplicative form of the balance–congruity

principle's prediction. Nevertheless, the results in Figures 2A and 3A show that the data for IAT

measures (much more than for self-report measures) were quite close to expectations based on the

assumption of a pure multiplicative model. This suggests that the consistency processes theorized in

BIT may be sustained or enhanced by the multiple triads in which any one measure participates.

---

[11] For example, in the absence of valid zero values, a positive value on a self-esteem IAT might not validly indicate positive valence associated with self, and a zero value of a race attitude IAT might not validly indicate absence of automatic preference for either Black or White race.

Usefulness of the Within-Study Meta-Analytic Method for Testing BIT's Balance–Congruity Principle

The within-study meta-analytic method introduced in this report provides an efficient alternative to the previously standard 4-test method for evaluating the balance–congruity principle. This method revealed (a) increased power of Test 2 (compared to the power of that test in the 4-test method), demonstrated by its averaged partial correlation coefficient being statistically significant for 25 of the 36 samples for which implicit measures were available, and (b) Test 2 successfully predicting fit of the pure multiplicative model as indexed by each sample's total number (out of 12) of 4-test method tests passed. Although passage of all 12 tests for 4-test method indicates purity of fit to the multiplicative prediction, it does not provide a quantitative index of strength of the multiplicative relationship—something that *is* provided by Test 2 of the within-study meta-analysis. In future research it will be reasonable to continue the use of the 4-test method, but it should be useful to report the within-study meta-analysis alongside. The two analyses complement one another. The two tests assess fit of data to balance–congruity predictions in complementary and mutually supportive ways.

Sources of Variance in IAT and Self-Report Measures

Interpretations of the present findings depend on understanding how IAT and self-report measures vary across testing occasions. Measures of internal consistency (such as split-half correlations or Cronbach's alpha) estimate the proportion of variance on a single testing occasion that is consistently measured. For the IAT, internal consistency has been found to average $r = .80$ in the meta-analysis of 257 studies located by Greenwald and Lai (2020). The difference between the percentage of variance represented by this internal consistency (64%) and that represented by the same meta-analysis's finding of test–retest reliability of $r = .50$ (25% of variance) indicates that

39% (= 64% − 25%) of variance of IAT measures is attributable to variance across testing

occasions.

Although meta-analytic estimates of internal consistency and test–retest reliability are not

available specifically for the parallel self-report measures used in the meta-analyzed studies, these

can be approximately estimated, respectively, as $r \approx .90$ and $r \approx .80$. Using those numbers, 17% (=

81% − 64%) of variance of these self-report measures can be understood as variance across testing

occasions. If both IAT and self-reports have valid zero points and both function exactly in the

fashion predicted by the balance–congruity principle, the finding for the meta-analysis' observed

self-report data in Figure 3B should be close to the simulation for reliability $r = .80$, in the same

way that the finding for the meta-analysis' IAT data in Figure 3A is close to the simulation based on

its expected reliability of $r = .50$. A plausible interpretation of the close similarity for IAT measures

in Figure 3A is that the IAT measures behave closely in accordance with the balance–congruity

predictions, with only small additional systematic sources of variance; additional variance across

testing occasions is non-systematic, meaning that (on average) it does not create any directional

distortion for IAT measures. For Figure 3B, the substantial gap between the predicted reliability of r

= .80 simulation and the observed self-report data indicates the presence of substantial systematic

influences other than the balance–congruity principle contributing to the observed self-report data.

The logical conclusion is that self-report measures have more systematic sources of artifact than do

IAT measures.

Evidence for Validity of Zero Points of IAT Measures Used in Tests of Balanced Identity

Theory

Because BIT's balance–congruity principle predictions are based on assumed validity of the

zero interpretations of measures used in their tests, those predictions should be confirmed to the

extent that measures used in their tests indeed possess valid zero points. Cvencek et al. (2012) meta-

analyzed 18 studies that reported tests of BIT's balance–congruity principle (54 correlational tests,

three per study). Although Cvencek et al.'s results were consistent with the assumption that IAT

measures had theoretically valid zero points, their analyses had insufficient power for precise tests

for either IAT or self-report measures. The present meta-analysis had substantially greater power

and precision.

Relative to non-displaced measures, zero-displaced IAT and self-report measures reduced

correlations between products of two of the triad of measures in each study with the third measure.

This was demonstrated in examination of the aggregated within-study results of Test 1 of the 4-test

method (see Figures 2A and 2B). Still greater precision was available in the tests that examined

BIT's predictions concerning the correlation between the means of each of the three variables in the

balanced identity design with the numerical value of the correlation between the other two

measures. If zero points of the measures used in this test are valid, the regression of correlations of

pairs of variables on means of the third variable should be positive in slope and should pass through

the origin of the regression plot (i.e., the intercept of the regressions should be at or close to zero).

The expected positive slopes were found for both IAT and self-report measures (Figures 3A and

3B), with the clarity of confirmation of this prediction again being considerably clearer for IAT

measures (Figure 3A) than for self-report measures (Figure 3B).

This article's zero-point validity analyses were conducted using sample-level correlations.

That fact prompted a reviewer to question whether the article's meta-analytic tests say more about

sample-aggregate values of zero points than about individual-respondent zero points. This is not the

case. The first BIT publication made clear that balance–congruity predictions depended on validity

of the zero value at the level of individual subjects (Greenwald et al., 2002, p. 11). Also relevant is

that the present findings were based entirely on analyses that used individual-subject data, never

using multi-subject aggregations as variables in computed correlations.

Why Are Balanced Identity Patterns More Strongly Apparent With IAT Than With Self-Report Measures?

Greenwald et al. (2002, p. 17) suggested two plausible causes for the relatively weak fit to expectations of the balance–congruity principle that they observed with self-report measures. First, subjects might lack *introspective access* to the strengths of the associations they are asked to report. Second, subjects might suppress accurate report on associations to which they have introspective access due to *response factors* such as demand characteristics (Orne, 1962), evaluation apprehension (Rosenberg, 1969), and subject role-playing (Weber & Cook, 1972). As an example, in a balanced identity design involving White–Black contrasts in racial identity (SG), self-esteem (SA), and racial attitudes (GA), White participants who want to appear unbiased might suppress report of an internally known racial bias, instead reporting either no bias or perhaps an out-group preference.

Previous discussions of differences between IAT and self-report measures have focused on results for single IAT or self-report measures. These explanations do not immediately apply to differences in affective–cognitive consistency findings involving relations among trios of measures, as in the balance–congruity tests summarized in this article. Nevertheless, the response factors explanation of IAT–self-report differences provides some basis for expecting *greater* consistency among self-report than IAT measures, to the extent that response factors include conscious reasoning that might increase evidence for consistency. The present findings of greater evidence for consistency with implicit measures, which presumably limit opportunity for conscious reasoning, oppose that suggestion. It is time to consider the possibility that consistency processes may operate outside of conscious awareness. Theory to explain such automatic operation of affective–consistency processes is not yet developed.

One of the noteworthy findings of this report is the confirmation of BIT predictions with self-report measures. This result is in line with general idea of differences between associative and propositional representations (Strack & Deutsch, 2004). According to Strack and Deutsch (2004) "the reflective system is driven by the principle of consistency as it strives to avoid or remedy inconsistencies between its elements" (p. 225). Following this first theoretical effort to account for implicit–explicit divergence in attitudes, a subsequent conceptual model was proposed: the Associative-Propositional Evaluation model (APE; Gawronski & Bodenhausen, 2006). The APE model holds that explicit evaluations are the behavioral outcome of propositional processes, which are defined as the validation of the information implied by activated (automatic) associations (Gawronski & Bodenhausen, 2014). According to the APE model: "cognitive consistency is exclusively a concern of propositional reasoning" (Gawronski & Bodenhausen, 2006, p. 695). However, the APE model was not stated in a way that would account for the strong findings with IAT measures, nor does it generate a prediction that associative measures should outperform propositional measures. Future research will profit from examining conditions under which consistency is more likely to arise from propositional versus associative processes.

Conclusion

BIT's balance–congruity principle predicts that people with positive self-esteem should have positively-valenced attitudes toward concepts or groups that are closely associated with the self. This review provides the strongest evidence yet available for this theoretical expectation of BIT. It also solidifies previous indications that BIT's support is more evident in studies using IAT measures than in those with self-report measures. The within-study meta-analytic strategy introduced in this article to test balance–congruity predictions was found to offer an efficient alternative to the original 4-test method for testing those predictions. Tests that displaced the zero points of IAT measures (prior to using them in tests of predictions involving multiplicative products

of measures) confirmed the assumption that (non-displaced) IAT measures have the valid zero

points that are required in tests of BIT's balance–congruity principle. As a group, the subset of

studies in this meta-analysis that included IAT self-esteem measures confirmed BIT predictions,

comparable to the subset of studies in which IATs measured the associations of self with attributes

other than valence. BIT's balance–congruity principle has now been confirmed when tested either

with IAT or self-report measures.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Aberson, C. L., Healy, M., & Romero, V. (2000). Ingroup bias and self-esteem: A meta-analysis. *Personality and Social Psychology Review*, *4*, 157–173. doi:10.1207/S15327957PSPR0402_04

Abrams, D., & Hogg, M. A. (1988). Comments on the motivational status of self-esteem in social identity and intergroup discrimination. *European Journal of Social Psychology*, *18*, 317–334. doi:10.1002/ejsp.2420180403

*Aidman, E. V., & Carroll, S. M. (2003). Implicit individual differences: Relationships between implicit self-esteem, gender identity, and gender attitudes. *European Journal of Personality*, *17*, 19–37. doi:10.1002/per.465

*Ashburn-Nardo, L. (2010). The importance of implicit and explicit measures for understanding social stigma. *Journal of Social Issues*, *66*, 508–520. doi:10.1111/j.1540-4560.2010.01659.x

Axt, J., & Lai, C. K. (2019). Reducing discrimination: A bias versus noise perspective. *Journal of Personality and Social Psychology*, *117*, 26–49. doi:10.1037/pspa0000153

*Banaji, M. R., Greenwald, A. G., & Rosier, M. (1997, October). *Implicit esteem: When collectives shape individuals*. Paper presented at the Preconference on Self, Toronto, Canada.

*Baron, A. S. (2003). [IAT response latencies and errors]. Unpublished raw data.

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53–58. doi:10.1111/j.1467-9280.2005.01664.x

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem

cause better performance, interpersonal success, happiness, or healthier lifestyles?

Psychological Science in the Public Interest, *4*, 1–44. doi:10.1111/1529-1006.01431

Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect.

Journal of Personality and Social Psychology, *62*, 229–237. doi:10.1037/0022-

3514.62.2.229

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful

metric of implicit prejudice. *Journal of Applied Psychology*, *100*, 1468–1481.

doi:10.1037/a0038379

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of

implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and*

Social Psychology, *79*, 631–643. doi:10.1037/0022-3514.79.4.631

Buhrmester, M. D., Blanton, H., & Swann, W. B., Jr. (2011). Implicit self-esteem: Nature,

measurement, and a new way forward. *Journal of Personality and Social Psychology*, *100*,

365–385. doi:10.1037/a0021341

Cadinu, M. R., & Rothbart, M. (1996). Self-anchoring and differentiation processes in the minimal

group setting. *Journal of Personality and Social Psychology*, *70*, 661–677.

doi:10.1037/0022-3514.70.4.661

Cofer, C. N., & Foley, J. P., Jr. (1942). Mediated generalization and the interpretation of verbal

behavior: I. Prolegomena. *Psychological Review*, *49*, 513–540. doi:10.1037/h0060856

Cooper, L. A., Roter, D. L., Carson, K. A., Beach, M. C., Sabin, J. A., Greenwald, A. G., & Inui, T.

S. (2012). The associations of clinicians' implicit attitudes about race with medical visit

communication and patient ratings of interpersonal care. *American Journal of Public Health*,

102, 979–987. doi:10.2105/AJPH.2011.300558

Creemers, D. H. M., Scholte, R. H. J., Engels, R. C. M. E., Prinstein, M. J., & Wiers, R. W. (2013).

Damaged self-esteem is associated with internalizing problems. *Frontiers in Psychology*, *4*,

152. doi:10.3389/fpsyg.2013.00152

Critcher, C. R., & Ferguson, M. J. (2016). "Whether I like it or not, it's important": Implicit

importance of means predicts self-regulatory persistence and success. *Journal of Personality

and Social Psychology*, *110*, 816–839. doi:10.1037/pspa0000053

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2012). Balanced identity theory: Review of

evidence for implicit consistency in social cognition. In B. Gawronski & F. Strack (Eds.),

Cognitive consistency: A fundamental concept in social cognition (pp. 157–177). New York,

NY: Guilford Press.

*Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool

children confirm self-esteem's role in maintaining a balanced identity. *Journal of

Experimental Social Psychology*, *62*, 50–57. doi:10.1016/j.jesp.2015.09.015

*Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary

school children. *Child Development*, *82*, 766–779. doi:10.1111/j.1467-8624.2010.01529.x

*Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math–gender

stereotypes in Singaporean children. *Journal of Experimental Child Psychology*, *117*, 73–91.

doi:10.1016/j.jecp.2013.07.018

Cvencek, D., Nasir, N. S., O'Connor, K., Wischnia, S., & Meltzoff, A. N. (2015). The development

of math–race stereotypes: "They say Chinese people are the best at math". *Journal of

Research on Adolescence*, *25*, 630–637. doi:10.1111/jora.12151

*Devos, T., Blanco, K., Muñoz, C., Dunn, R., & Ulloa, E. C. (2008). Implicit orientation toward

family and school among bilingual Latino college students. *The Journal of Social

Psychology*, *148*, 449–472. doi:10.3200/SOCP.148.4.449-472

*Devos, T., Blanco, K., Rico, F., & Dunn, R. (2008). The role of parenthood and college education in the self-concept of college students: Explicit and implicit assessments of gendered aspirations. *Sex Roles*, *59*, 214–228. doi:10.1007/s11199-008-9430-6

*Devos, T., & Cruz Torres, J. A. (2007). Implicit identification with academic achievement among Latino college students: The role of ethnic identity and significant others. *Basic and Applied Social Psychology*, *29*, 293–310. doi:10.1080/01973530701503432

*Devos, T., Diaz, P., Viera, E., & Dunn, R. (2007). College education and motherhood as components of self-concept: Discrepancies between implicit and explicit assessments. *Self and Identity*, *6*, 256–277. doi:10.1080/15298860601118710

**Devos, T., Gavin, K., & Quintana, F. J. (2010). Say "adios" to the American dream? The interplay between ethnic and national identity among Latino and Caucasian Americans. *Cultural Diversity and Ethnic Minority Psychology*, *16*, 37–49. doi:10.1037/a0015868

*Dunham, Y., Baron, A. S., & Banaji, M. R. (2007). Children and social groups: A developmental analysis of implicit consistency in Hispanic Americans. *Self and Identity*, *6*, 238–255. doi:10.1080/15298860601115344

*Farnham, S. D., & Greenwald, A. G. (1999, June). *In-group favoritism: Implicit self-esteem × in-group identification*. Paper presented at the 11th annual meeting of the American Psychological Society, Denver, CO.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297–327. doi:10.1146/annurev.psych.54.101601.145225

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

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: Conceptual, empirical, and metatheoretical issues: Reply to Albarracín, Hart,

and McCulloch (2006), Kruglanski and Dechesne (2006), and Petty and Briñol (2006). *Psychological Bulletin*, *132*, 745–750. doi:10.1037/0033-2909.132.5.745

Gawronski, B., & Bodenhausen, G. V. (2014). Implicit and explicit evaluation: A brief review of the associative–propositional evaluation model. *Social and Personality Psychology Compass*, *8*, 448–462. doi:10.1111/spc3.12124

Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it, because I like myself: Associative self-anchoring and post-decisional change of implicit evaluations. *Journal of Experimental Social Psychology*, *43*, 221–232. doi:10.1016/j.jesp.2006.04.001

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychology Review*, *109*, 3–25. doi:10.1037/0033-295X.109.1.3

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology*, *79*, 1022–1038. doi:10.1037/0022-3514.79.6.1022

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*, 419–445. doi: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. doi:10.1037/0022-3514.74.6.1464

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. doi:10.1037/0022-3514.85.2197

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience

asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and

Wentura (2004). *Journal of Experimental Psychology: General*, *134*, 420–425.

doi:10.1037/0096-3445.134.3.420

Greenwald, A. G., Rudman, L. A., Nosek, B. A., & Zayas, V. (2006). Why so little faith? A reply to

Blanton and Jaccard's (2006) skeptical view of testing pure multiplicative theories.

Psychological Review, *113*, 170–180. doi:10.1037/0033-295X.113.1.170

*Gumble, A., & Carels, R. (2012). The harmful and beneficial impacts of weight bias on well-

being: The moderating influence of weight status. *Body Image*, *9*, 101–107.

doi:10.1016/j.bodyim.2011.07.005

Heider, F. (1958). *The psychology of interpersonal relations*. Hoboken, NJ: Wiley.

doi:10.1037/10628-000

Hewstone, R., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Psychology Review*, *53*,

575–604. doi:10.1146/annurev.psych.53.100901.135109

Hogg, M. A. (2000). Social identity and social comparison. In J. Suls & L. Wheeler (Eds.), *The

Plenum series in social/clinical psychology. Handbook of social comparison: Theory and

research* (pp. 401–421). Dordrecht, Netherlands: Kluwer Academic.

*Horcajo, J., Briñol, P., & Petty, R. E. (2010). Consumer persuasion: Indirect change and implicit

balance. *Psychology and Marketing*, *27*, 938–963. doi:10.1002/mar.20367

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1990). Experimental tests of the endowment effect

and the Coase theorem. *Journal of Political Economy*, *98*, 1325–1348. doi:10.1086/261737

*Lane, K. A., Mitchell, J. P., & Banaji, M. R. (2005). Me and my group: Cultural status can disrupt

cognitive consistency. *Social Cognition*, *23*, 353–386. doi:10.1521/soco.2005.23.4.353

Leeuwis, F. H., Koot, H. M., Creemers, H. E., & van Lier, P. A. C. (2015). Implicit and explicit

self-esteem discrepancies, victimization and the development of late childhood internalizing

problems. *Clinical Developmental Psychology*, *43*, 909–919. doi:10.1007/s10802-014-9959-

5

*Mellott, D. S., & Greenwald, A. G. (2000). *But I don't feel old! Implicit self-esteem, age identity

and ageism in the elderly*. Unpublished manuscript, University of Washington, Seattle, WA.

*Meltzoff, A. N., Cvencek, D., Heimann, M., Sundqvist, A., Koch, F.-S., & Maddox, C. D. (2019).

*Gender–career stereotypes in college: Relationship to social cognition, parental leave, and

occupational interests*. Manuscript in preparation.

Mitchell, J. P., Ames, D. L., Jenkins, A. C., & Banaji, M. R. (2009). Neural correlates of stereotype

application. *Journal of Cognitive Neuroscience*, *21*, 594–604. doi:10.1162/jocn.2009.21033

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review*,

60, 393–404. doi:10.1037/h0063098

*Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore

math ≠ me. *Journal of Personality and Social Psychology*, *83*, 44–59. doi:10.1037//0022-

3514.83.1.44

*Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math

engagement and achievement. *American Educational Research Journal*, *48*, 1125–1156.

doi:10.3102/0002831211410683

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . . Banaji,

M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European

Review of Social Psychology*, *18*, 36–88. doi:10.1080/10463280701489053

Nosek, B. A., Smyth F. L., Sriram N., Lindner N. M., Devos T., Ayala, A., . . . Greenwald, A. G.

(2009). National differences in gender–science stereotypes predict national sex differences

in science and math achievement. *Proceedings of the National Academy of Science*, *106*, 10593–10597. doi:10.1073/pnas.0809921106

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783. doi:10.1037/h0043424

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

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review*, *62*, 42–55. doi:10.1037/h0048153

Perkins, A. W., & Forehand, M. R. (2006). Decomposing the implicit self-concept: The relative influence of semantic meaning and valence on attribute self-association. *Social Cognition*, *24*, 387–408. doi:10.1521/soco.2006.24.4.387

Perkins, A. W., & Forehand, M. R. (2012). Implicit self-referencing: The effect of nonvolitional self-association on brand and product attitude. *Journal of Consumer Research*, *39*, 142–156. doi:10.1086/662069

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research* (pp. 279–349). New York, NY: Academic Press.

Roth, J., & Steffens, M. C. (2014). When I becomes we: Associative self-anchoring drives implicit intergroup bias in minimal groups. *Social Psychology*, *45*, 253–264. doi:10.1027/1864-9335/a000169

*Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin*, *27*, 1164–1178. doi:10.1177/0146167201279009

*Rudman, L. A., & McLean, M. C. (2013). [Evidence for Balanced Identity Theory for Black and White Americans]. Unpublished raw data.

Sabin, J. A., & Greenwald, A. G. (2012). The influence of implicit bias on treatment recommendations for 4 common pediatric conditions: Pain, urinary tract infection, attention deficit hyperactivity disorder, and asthma. *American Journal of Public Health*, *102*, 988–995. doi:10.2105/AJPH.2011.300621

Schindler, S., Wolff, W., Kissler, J. M., & Brand, R. (2015). Cerebral correlates of faking: Evidence from a brief implicit association test on doping attitudes. *Frontiers in Behavioral Neuroscience*, *9*, 139. doi:10.3389/fnbeh.2015.00139

*Schmidt, K., & Nosek, B. A. (2015). *Balanced identities: Racial attitudes, racial identity, and self-esteem*. Unpublished manuscript, University of Virginia, Charlottesville, VA.

*Srivastava, S. B., & Banaji, M. R. (2011). Culture, cognition, and collaborative networks in organizations. *American Sociological Review*, *76*, 207–233. doi:10.1177/0003122411399390

*Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, *102*, 947–963. doi:10.1037/a0019920

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

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1–39. doi:10.1146/annurev.ps.33.020182.000245

Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, *1*, 149–178. doi:10.1002/ejsp.2420010202

Tajfel, H., & Turner, J. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S.

    Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey, CA:

    Brooks-Cole.

*Tang, E., & Greenwald, A. G. (2013, May). *Bicultural Conflict Examined with Balanced Cultural

    Identity Measures*. Poster presented at Undergraduate Research Symposium, University of

    Washington, Seattle, WA.

Trendel, O., Mazodier, M., & Vohs, K. (2018). Making warnings about misleading advertising and

    product recalls more effective: An implicit attitude perspective. *Journal of Marketing

    Research*, *55*, 265–276. doi:10.1509/jmr.14.0305

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering

    the social group: A self-categorization theory*. Cambridge, MA: Basil Blackwell.

Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of

    subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, *77*, 273–

    295. doi:10.1037/h0032351

Ye, Y., & Gawronski, B. (2016). When possessions become part of the self: Ownership and implicit

    self-object linking. *Journal of Experimental Social Psychology*, *64*, 72–87.

    doi:10.1016/j.jesp.2016.01.012

Table 1

Characteristics of 36 Independent Samples for Meta-Analysis of Studies Using Balanced Identity Designs

Citation	Participants	n implicit	n explicit	Group concepts	Attribute concepts	Tests passed (implicit / explicit)
Aidman & Carroll (2003)	Females, males	66		FEMALE (male)	PLEASANT (unpleasant)	4 / na
Ashburn-Nardo (2010)	Blacks, Whites	112	113	OWN RACE (other race)	PLEASANT (unpleasant)	12 / 2
Banaji et al. (1997)	Blacks, Whites	61		WHITE (Black)	POSITIVE (negative)	10 / na
Baron (2003)	Minority adolescents	40		BLACK (White)	POSITIVE (negative)	9 / na
Cvencek et al. (2016, Study 1)	Girls, boys	39		BOY (girl)	GOOD (bad)	12 / na
Cvencek et al. (2016, Study 2)	Girls, boys	96		BOY (girl)	GOOD (bad)	4 / na
Cvencek et al. (2016, Study 3)	Girls, boys	60		BOY (girl)	GOOD (bad)	10 / na
Cvencek et al. (2011)	Girls, boys	222	220	BOY (girl)	MATH (reading)	12 / 8
Cvencek et al. (2014)	Girls, boys	155	167	BOY (girl)	MATH (reading)	10 / 12
Devos, Blanco, Muñoz, et al. (2008)	Latinos (whose parents either were or were not high school graduates)	128		FAMILY (school)	PLEASANT (unpleasant)	11 / na
Devos, Blanco, Rico, et al. (2008)	Undergraduates who are not parents	169	169	WOMAN (man)	PARENTHOOD (college education)	12 / 9
Devos & Cruz Torres (2007, Study 1)	Latinos, Whites (whose parents either were or were not high school graduates)	80	80	CAUCASIAN (Latino)	HIGH ACHIEVEMENT (low achievement)	12 / 9
Devos & Cruz Torres (2007, Study 2)	Latinos (whose parents either were or were not high school graduates)	49	49	SIGNIFICANT OTHERS (acquaintances)	HIGH ACHIEVEMENT (low achievement)	12 / 10
Devos et al. (2007, Study 3)	Mothers, females planning to have first child in under / over 4 years	60		COLLEGE EDUCATION (motherhood)	PLEASANT (unpleasant)	12 / na
Devos et al. (2010, Study 2)	Latino U.S. citizens, White U.S. citizens	108	108	CAUCASIAN AMERICAN (Latino American)	AMERICAN (foreign)	8 / 8
Dunham et al. (2007)	Latinos	139		HISPANIC (Black)	GOOD (bad)	7 / na
Dunham et al. (2007)	Latinos	134		HISPANIC (White)	GOOD (bad)	9 / na

Study	Participants	n	n	Group (contrast)	Attribute (contrast)	Tests
Farnham & Greenwald (1999)	Females	65	52	FEMALE (male)	POSITIVE (negative)	12 / 8
Gumble & Carels (2012)	Undergraduates (with either low BMI [< 25] or high BMI [≥ 25])	85		THIN (fat)	GOOD (bad)	6 / na
Horcajo et al. (2010, Study 3)	Undergraduates	22		VEGETABLE (animal)	GOOD (bad)	5 / na
Horcajo et al. (2010, Study 3)	Undergraduates	26		VEGETABLE (animal)	GOOD (bad)	6 / na
Horcajo et al. (2010, Study 4)	Undergraduates	29		VEGETABLE (animal)	GOOD (bad)	7 / na
Horcajo et al. (2010, Study 4)	Undergraduates	25		VEGETABLE (animal)	GOOD (bad)	2 / na
Lane et al. (2005)[a]	Yale undergraduates (in different residential colleges)	224		OWN RESIDENTIAL COLLEGE (other residential college)	GOOD (bad)	12 / na
Lane et al. (2005)[a]	Yale undergraduates	218		YALE (Harvard)	GOOD (bad)	12 / na
Mellott & Greenwald (2000)	Undergraduates, senior citizens	98	83	OLD (young)	POSITIVE (negative)	7 / 8
Meltzoff et al. (2019)	Females, males	96	96	MALE (female)	WORK (family)	12 / 11
Meltzoff et al. (2019)	Females, males	96	96	MALE (female)	WORK (family)	12 / 8
Nosek et al. (2002, Study 2)	Females, males	91		MALE (female)	MATH (arts)	12 / na
Nosek & Smyth (2011)	Females, males	475	5030	MALE (female)	MATH (arts, verbal, or furniture)	6 / 6
Rudman et al. (2001, Study 4)	Females, males	95	95	MALE (female)	POTENT (warm)	10 / 5
Rudman & McLean (2013, Study 1)	Blacks, Whites	206	206	OWN RACE (other race)	POSITIVE (negative)	11 / 3
Schmidt & Nosek (2015)	Asians, Blacks, Hispanics, Whites	5926	6115	WHITE (black)	GOOD (bad)	6 / 6
Srivastava & Banaji (2011)	Research & Development staff, Commercial staff	110	94	RESEARCH & DEVELOPMENT (commercial)	COLLABORATIVE (independent)	4 / 6
Steffens et al. (2010, Study 1)[b]	Girls, boys	140		GIRL (boy)	LANGUAGE (math)	8 / na
Tang & Greenwald (2013)	Asian-born, U.S.-born (Asian undergraduates in U.S.)	63		ASIA (U.S.)	FAMILY LIFE (college life)	12 / na

Note. Participants = participant groups comprising independent samples in each study. Participant groups are identified in terms of the two contrasted identity groups used in the IAT measures; n = number of participants in each independent sample. All studies included measures of associations of: (a) *self* (contrasted with *other*) with a contrasted pair of *group* concepts, (b) *self* with a contrasted pair of *attribute* concepts, and (c) the contrasted pair of *group* concepts with the contrasted pair of *attribute* concepts. Capitalized group and attribute concepts were the ones scored as positive when associated with self. The IAT measure of group–attribute association was scored so that the association between the capitalized group and the capitalized attribute was positive. All studies reported implicit effects as IAT D measures (Greenwald, Nosek, & Banaji, 2003, p. 214), except for the two Lane et al. (2005) studies and Steffens et al. (2010), as described in the text. BMI = Body Mass Index. Tests passed refer to the 12 possible tests of the 4-test method.

Figure 1. Schematic fragment of a social knowledge structure (reproduced with permission of authors from Figure 1 of Greenwald et al., 2002). This structure includes associations corresponding to social psychology's major cognitive (stereotypes and self-concept) and affective (self-esteem and attitude) constructs. Nodes (ovals) represent concepts, and links (lines) represent associative relations. Association strength is indicated by line thickness. The *self-concept* corresponds to the links of the *Me* node to social categories (professor, grandmother) and attributes (intelligent, athletic). *Self-esteem* corresponds to the links of the *Me* node to valence (+ + + or − − −). Analogous to self-concept, *stereotypes* correspond to links between social categories and attributes. Analogous to self-esteem, *attitudes* are links that connect social category nodes to valence nodes (+ + + or − − −).
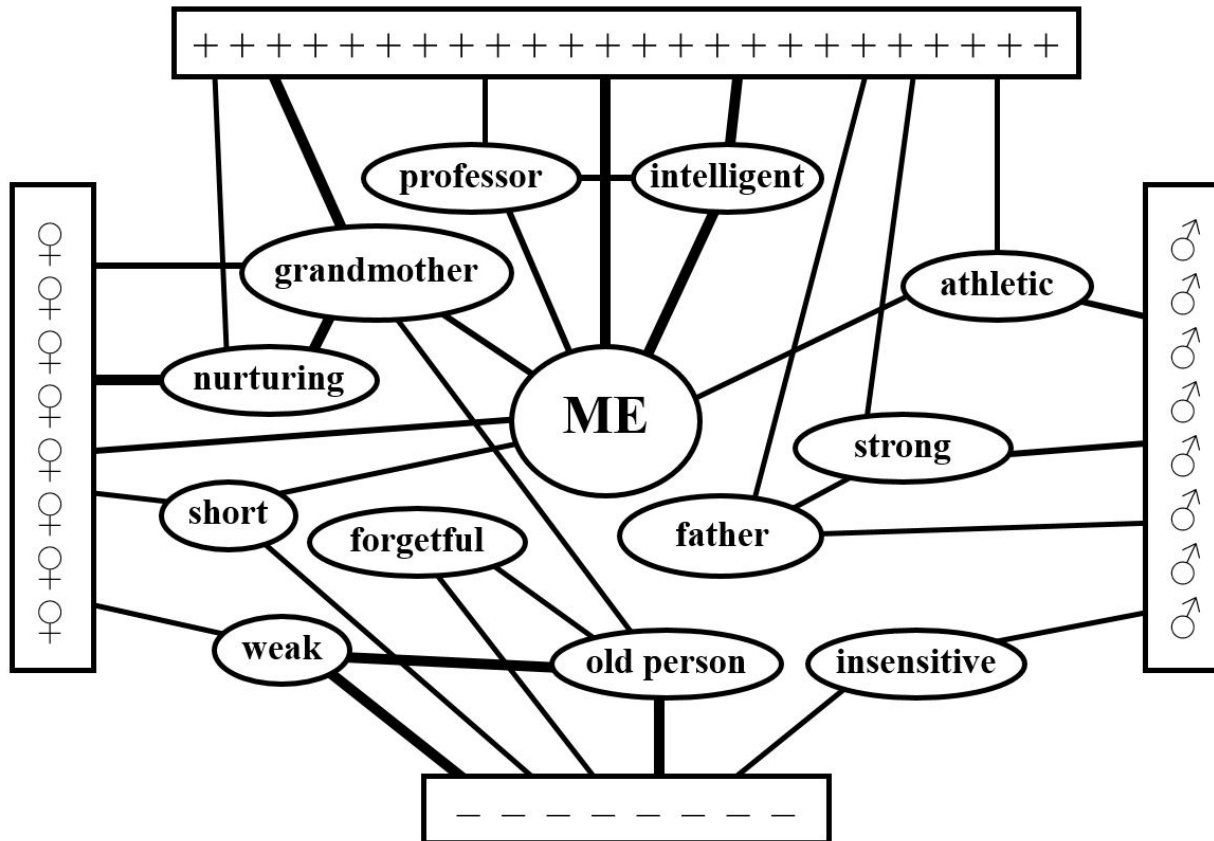
Figure 2. Effects of displacements of zero points of IAT (2A) and self-report (2B) measures on magnitude of confirmations of the balance-congruity principle in Test 1 of the 4-test method. In each panel the dotted line shows results expected if Test 1 is conducted with a perfectly reliable method. The dashed line shows results expected with expected reliability of IAT measures (*r* = .50, Panel 2A) and of self-report measures (*r* = .80, Panel B).
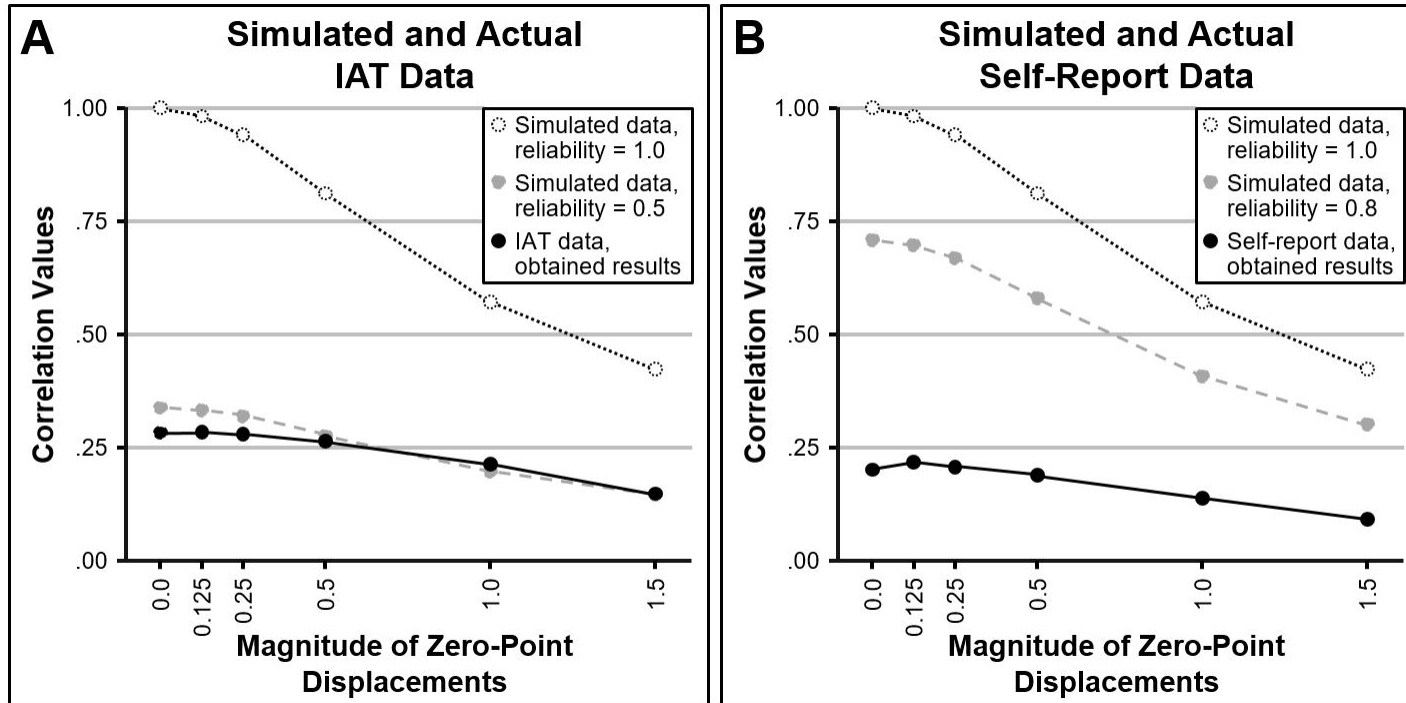
Figure 3. Plots of Fisher Z-transformed correlations between pairs of association strength measures in balanced identity studies, plotted as a function of the third predictor. Plots include regression slopes and their 95% confidence intervals. Distinct data point markers identify the type of correlation between two of the three association measures in each balanced identity designs: self–group (SG, identity), self–attribute (SA, self-esteem or self-concept), and group–attribute (GA, attitude or stereotype). For points representing each type of correlation, the X-axis gives the value (in *SD* units) of the mean of third variable in the design. Data are presented for IAT measures (panel A) and self-report measures (panel B).