Personality and Social Psychology Bulletin
1–16

DOI: 10.1177/0146167220916631
journals.sagepub.com/home/pspb
SAGE

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

Dario Cvencek[1], Andrew N. Meltzoff[1], Craig D. Maddox[1], Brian A. Nosek[2], Laurie A. Rudman[3], Thierry Devos[4], Yarrow Dunham[5], Andrew S. Baron[6], Melanie C. Steffens[7], Kristin Lane[8], Javier Horcajo[9], Leslie Ashburn-Nardo[10], Amanda Quinby[11], Sameer B. Srivastava[12], Kathleen Schmidt[13], Eugene Aidman[14], Emilie Tang[15], Shelly Farnham[16], Deborah S. Mellott[1], Mahzarin R. Banaji[17], and Anthony G. Greenwald[1]

Abstract

This meta-analysis evaluated theoretical predictions from balanced identity theory (BIT) and evaluated the validity of zero points of Implicit Association Test (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

Received February 7, 2019; revision accepted March 4, 2020

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*[1] 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

[1]University of Washington, Seattle, USA
[2]University of Virginia, Charlottesville, USA
[3]Rutgers University, Piscataway, NJ, USA
[4]San Diego State University, CA, USA
[5]Yale University, New Haven, CT, USA
[6]The University of British Columbia, Vancouver, Canada
[7]University of Koblenz–Landau, Germany
[8]Bard College, Annandale-on-Hudson, NY, USA
[9]Universidad Autónoma de Madrid, Madrid, Spain
[10]Indiana University–Purdue University Indianapolis, USA
[11]Bloomington Counseling and Wellness Services, LLC, IN, USA
[12]University of California, Berkeley, USA
[13]Southern Illinois University, Carbondale, USA
[14]The University of Sydney, NSW, Australia
[15]Pacific Northwest University of Health Sciences, Yakima, WA, USA
[16]Google, Seattle, WA, USA
[17]Harvard University, Cambridge, MA, USA

Corresponding Author:
Dario Cvencek, Institute for Learning & Brain Sciences, University of Washington, Portage Bay Building, Box 357988, Seattle, WA 98195, USA.
Email: dario1@uw.edu

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 and Banaji (1995), including minimal group effects (Tajfel et al., 1971), endowment effects (Kahneman et al., 1990), mere ownership effects (Beggan, 1992), self-anchoring effects (Cadinu & Rothbart, 1996; Gawronski et al., 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 nonsocial 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 et al.'s (1987) self-categorization theory (SCT) and Tajfel's (1982) social identity theory (SIT)—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 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 et al. (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 et al., 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 IAT (Greenwald et al., 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 et al., 2009; Schindler et al., 2015), education (Cvencek et al., 2015; Devos & Cruz Torres, 2007; Nosek et al., 2009), developmental science (Baron & Banaji, 2006; Cvencek et al., 2011), clinical psychology (Creemers et al., 2013; Leeuwis et al., 2015), health psychology (Cooper et al., 2012; Sabin & Greenwald, 2012), and marketing research (Horcajo et al., 2010; Trendel et al., 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* and *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 et al., 2003).

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 tests 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 et al. (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. In addition, 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 article is that the full data set, consisting of 12,773 participants (ranging from young children to adults) across 36 studies is being made publicly available in a widely accessible archive.[2]

Search Method for Locating Balanced Identity Data Sets

PsycINFO, PubMed, and Google Scholar were searched, using the following 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 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 nonsocial categories (e.g.,

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, & Dunn (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
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

(continued)

Please remove superscript a

Table 1. (continued)

Citation	Participants	*n* implicit	*n* explicit	Group concepts	Attribute concepts	Tests passed (implicit/explicit)
Lane et al. (2005)	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. (2020)	Females, males	96	96	MALE (female)	WORK (family)	12/11
Meltzoff et al. (2020)	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	5,030	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	5,926	6,115	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 the United States	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 et al., 2003, p. 214), except for the two Lane et al. (2005) studies and Steffens et al. (2010), as described in their respective, original publications. Tests passed refer to the 12 possible tests of the 4-test method. BMI = body mass index; IAT = Implicit Association Test.

weight, food), along with multiple stereotypes (e.g., math, parenthood, achievement).

The research reviewed in this article 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 nonstandard 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 1 to 3. 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 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 Equation 1), self and attribute to other groups (the other Gs in Equation 2), or group and attribute to persons other than self (the Ps in Equation 3). 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., $S_\blacktriangle A_\blacktriangle \cdot G_\blacktriangle A_\blacktriangle$ in Equation 1) should be the strongest predictor and, consistent with the balance–congruity principle, should be correlated with other predictors in the equation.

$$\begin{aligned} S_\blacktriangle G_\blacktriangle = {} & b_\blacktriangle [S_\blacktriangle A_\blacktriangle][G_\blacktriangle A_\blacktriangle] \\ & + b_1 [S_\blacktriangle A_1][G_\blacktriangle A_1] \\ & + b_2 [S_\blacktriangle A_2][G_\blacktriangle A_2] \\ & + b_3 [S_\blacktriangle A_3][G_\blacktriangle A_3] + \ldots \end{aligned} \tag{1}$$

$$\begin{aligned} S_\blacktriangle A_\blacktriangle = {} & b_\blacktriangle [S_\blacktriangle G_\blacktriangle][G_\blacktriangle A_\blacktriangle] \\ & + b_1 [S_\blacktriangle G_1][G_1 A_\blacktriangle] \\ & + b_2 [S_\blacktriangle G_2][G_2 A_\blacktriangle] \\ & + b_3 [S_\blacktriangle G_3][G_3 A_\blacktriangle] + \ldots \end{aligned} \tag{2}$$

$$\begin{aligned} G_\blacktriangle A_\blacktriangle = {} & b_\blacktriangle [S_\blacktriangle G_\blacktriangle][S_\blacktriangle A_\blacktriangle] \\ & + b_1 [P_1 G_\blacktriangle][P_1 A_\blacktriangle] \\ & + b_2 [P_2 G_\blacktriangle][P_2 A_\blacktriangle] \\ & + b_3 [P_3 G_\blacktriangle][P_3 A_\blacktriangle] + \ldots \end{aligned} \tag{3}$$

Please insert a space before AND after the + sign.

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 1 to 3 (e.g., $S_\blacktriangle G_\blacktriangle = b_\blacktriangle \cdot [S_\blacktriangle A_\blacktriangle] \cdot [G_\blacktriangle A_\blacktriangle]$ in Equation 1). 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 two-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 et al., 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

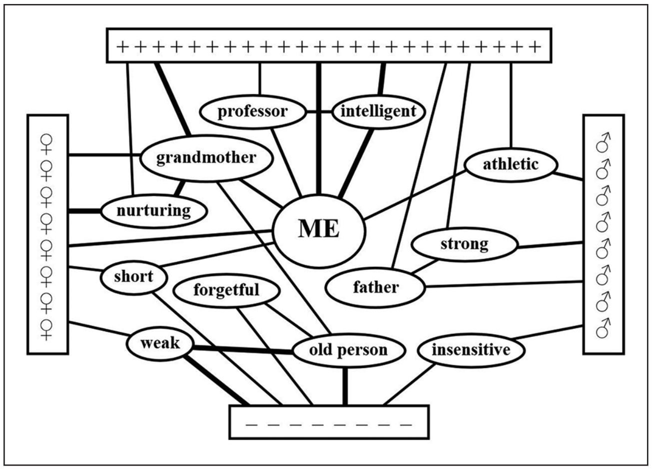

Figure 1. Schematic fragment of a social knowledge structure.
Source. Reproduced with permission of authors from Figure 1 of Greenwald et al. (2002).
Note. This structure includes associations corresponding to social psychology's major cognitive (stereotypes and self-concept) and affective (self-esteem 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 − − −).

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 two-step regression, and the Test 2 within-study aggregate was entered on Step 2. This weighted two-step regression found that the within-study Test 1 aggregates were *not* significant predictors of total number of tests passed, whereas the Test 2 aggregates were significant predictors; partial r for Test 2 in the second step was $pr = .384$, $t(33) = 2.39$, $p = .02$. The interpretation of these findings for drawing conclusions about usefulness of the within-study meta-analysis method is considered in the "General Discussion" section.

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 et al., 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 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 Tes[illegible]report measures, this difference was not statistically significant for either Test 1 or Test 2 ($ps > .10$). In addition, the level of support for balance–congru[illegible]ying valence (i.e., self-esteem rather th[illegible] 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). 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). If *either* the self–female or the self–positive association is zero, no strengthening is predicted. Furthermore, strengthening should be greater when *both* associations in the shared first-order link are strong. Tests of this multiplicative prediction require that the associations used in the test are measured on scales that have valid zero values (Greenwald et al., 2002). The dependence of this prediction on a valid interpretation of the zero points 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 et al., 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 vs. 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 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 et al., 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 support for the balance–congruity provided by the undisplaced tests of those predictions (see Supplemental Materials for relevant tests). 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 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 BIT.

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* association (GA). 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.

Figure 3's scatterplots combine, separately for IAT and self-report measures, the scatterplots for $r_{SG–SA}$ predicted by

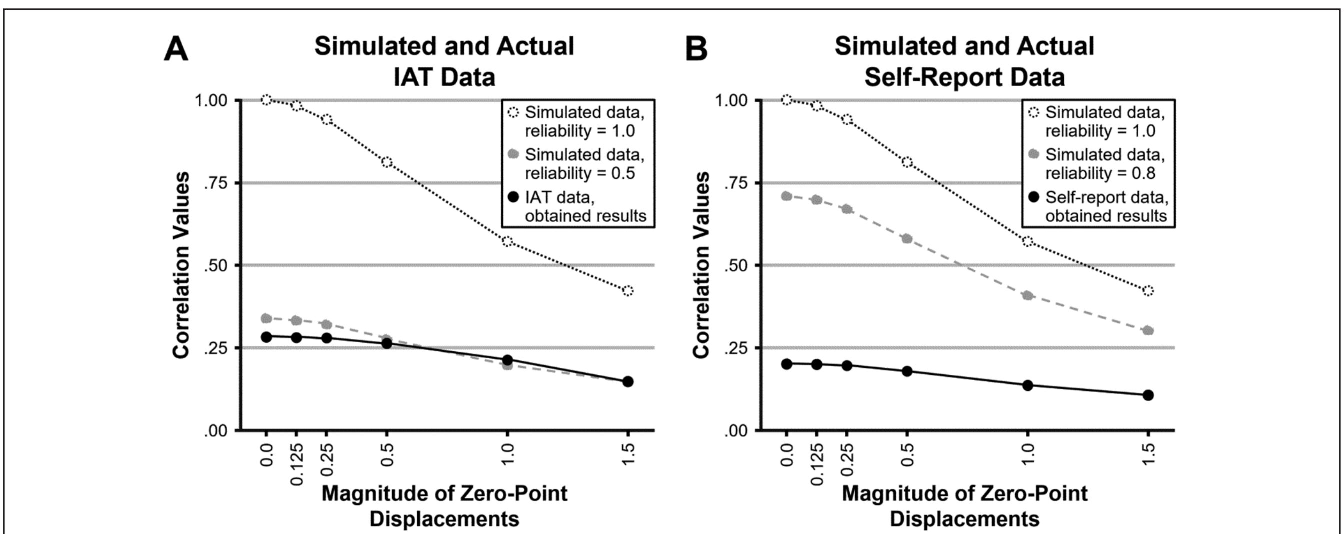

Figure 2. Effects of displacements of zero points of (A) IAT and (B) self-report measures on magnitude of confirmations of the balance–congruity principle in Test 1 of the 4-test method.
Note. In each panel, the dotted line shows results expected if Test 1 is conducted with a perfectly reliable method. The dashed line shows results expected with expected reliability of IAT measures ($r = .50$, (A)) and of self-report measures ($r = .80$, (B)). IAT = Implicit Association Test.

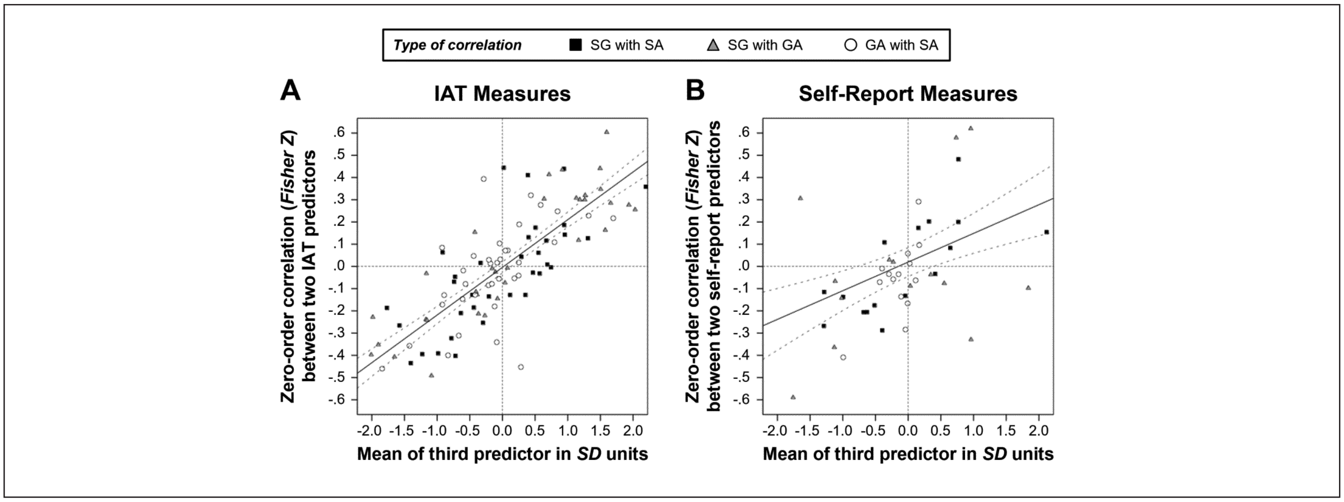

Figure 3. Plots of Fisher Z-transformed correlations between pairs of association strength measures in balanced identity studies, plotted as a function of the mean of third predictor.

Note. 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 e[illegible] self–group (SG, identity), self–attribute (SA, self-esteem or self-concept), and group–attribute (GA, attitude or stereotyp[illegible] ch 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 (A) and self-report measures (B). IAT = Implicit Association Test.

Please insert "the " before the word "third"

M_{GA}, $r_{SG–GA}$ predicted by M_{SA}, and $r_{GA–SA}$ 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 increase 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 × 36) correlation values (Figure 3A) all had strongly positive slopes, corresponding to correlations between .83 and .85. The 95% confidence intervals (CIs) 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 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 substantial 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 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 et al., 2003; Bosson et al., 2000; Buhrmester et al., 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 1–3). 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.

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 with 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 of the 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 (80%) and that represented by the same meta-analysis's finding of test–retest reliability of r = .50 (50% of variance) indicates that 30% (=80% − 50%) 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, 10% (=90% − 80%) 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's 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's 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 the 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 BIT

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 Figure 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 (Figure 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). 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) 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 (APE) model (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.

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.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported in part by a Grants from the National Science Foundation (HRD-1661285 to A.N.M. and SBE-1640889 to D.C.) and by the University of Washington's Implicit Cognition Research Fund.

ORCID iDs

Dario Cvencek https://orcid.org/0000-0002-0073-5862
Andrew N. Meltzoff https://orcid.org/0000-0001-8683-0547
Laurie A. Rudman https://orcid.org/0000-0001-7117-0312
Thierry Devos https://orcid.org/0000-0003-4546-7555
Yarrow Dunham https://orcid.org/0000-0002-4265-4438
Anthony G. Greenwald https://orcid.org/0000-0001-6865-8552

Notes

1. As explained by Greenwald et al. (2005), balanced identity theory's (BIT) 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.
2. All data and analysis files have been archived at: https://osf.io/9w24m/.
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 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 postdoctoral fellows.
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.
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 does not require that either item has a zero value. It does require that both A and B have the same numerical value.
7. Blanton et al. (2015) offered a method to empirically evaluate validity of zero points of attitude measures. Their method examined regressions of Implicit Association Test (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.)
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).
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."

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 Note 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 Figure 3A and 3B.
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.

Supplemental Material

Supplemental material is available online with this article.

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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037/pspa0000153

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

*Baron, A. S. (2003). *IAT response latencies and errors*. Harvard University. [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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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*, Article 152. https://doi.org/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. https://doi.org/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). 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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1080/15298860601115344

*Farnham, S. D., & Greenwald, A. G. (1999, June 3–6). *In-group favoritism: Implicit self-esteem × in-group identification* [Paper presentation]. 11th Annual Meeting of the American Psychological Society, *Denver, CO, United States*.

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

Festinger, L. (1957). *A theory of cognitive dissonance*. 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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1037/0022-3514.79.6.1022

Greenwald, A. G., & Lai, C. K. (2020). Implicit social cognition. *Annual Review of Psychology*, *71*, 419–445. https://doi.org/10.1146/annurev-psych-010419-050837

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.bodyim.2011.07.005

Heider, F. (1958). *The psychology of interpersonal relations*. Wiley. https://doi.org/10.1037/10628-000

Hewstone, R., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Psychology Review*, *53*, 575–604. https://doi.org/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). Kluwer Academic.

*Horcajo, J., Briñol, P., & Petty, R. E. (2010). Consumer persuasion: Indirect change and implicit balance. *Psychology and Marketing*, *27*, 938–963. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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.

*Meltzoff, A. N., Cvencek, D., Heimann, M., Sundqvist, A., Koch, F.-S., & Maddox, C. D. (2020). *Gender-related work/family stereotypes in two countries. University of Washington.* [Manuscript in preparation].

Mitchell, J. P., Ames, D. L., Jenkins, A. (2009). Neural correlates of stereotype a *Cognitive Neuroscience*, *21*, 594–604. https://doi.org/10.1162/jocn.2009.21033

Please print "University of Washington" in regular font, NOT italics.

Newcomb, T. M. (1953). An approach to the study of communicative acts. *Psychological Review*, *60*, 393–404. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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 Sciences of the United States of America*, *106*, 10593–10597. https://doi.org/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. https://doi.org/10.1037/h0043424

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. 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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1086/662069

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact inbehavioral research* (pp. 279–349). 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. https://doi.org/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. https://doi.org/10.1177/0146167201279009

*Rudman, L. A., & McLean, M. C. (2013). *Evidence for balanced identity theory for Black and White Americans*. Rutgers University. [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. https://doi.org/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*, Article 139. https://doi.org/10.3389/fnbeh.2015.00139

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

*Srivastava, S. B., & Banaji, M. R. (2011). Culture, cognition, and collaborative networks in organizations. *American Sociological Review*, *76*, 207–233. https://doi.org/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. https://doi.org/10.1037/a0019920

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

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, *33*, 1–39. https://doi.org/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. https://doi.org/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). Brooks/Cole.

*Tang, E., & Greenwald, A. G. (2013, May). *Bicultural conflict examined with balanced cultural identity measures* [Poster presentation]. Undergraduate Research Symposium, University of Washington, Seattle, WA, United States.

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. https://doi.org/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*. 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. https://doi.org/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. https://doi.org/10.1016/j.jesp.2016.01.012

Cvencek, D., Meltzoff, A. N., Maddox, C. D., Nosek, B. A., Rudman, L. A., Devos, T. Dunham, Y., Baron, A. S., Steffens, M. C., Lane, K., Horcajo, J., Ashburn-Nardo, L., Quinby, A., Srivastava, S. B., Schmidt, K., Aidman, E., Tang, E., Farnham, S., Mellott, D. S., Banaji, M. R., Greenwald, A. G. (in press). Meta-analytic use of Balanced Identity Theory to validate the Implicit Association Test. *Personality and Social Psychology Bulletin*.

Supplemental Material for

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

***Personality and Social Psychology Bulletin* (2020)**
Accepted March, 2020

Supplemental Material Pertaining to the 4-Test Method

Test 1 of the 4-test method examines the regression of a criterion measure of association strength on the multiplicative product of two theoretically specified predictor measures of association strengths. This is done in Step 1 of a 2-step hierarchical regression for each of the three focal measures (SG, SA, and GA) as a criterion. The balance–congruity principle expects the multiplicative product to be a substantial predictor. The entry of a product term on Step 1 of a hierarchical regression differs from the standard procedure for testing product terms (or interaction effects), which is to enter two component variables on Step 1, then enter their product on Step 2. The rationale for the 4-test method's reversal of this standard procedure can be appreciated with a thought experiment using a known pure-multiplicative theoretical model—prediction of the area of a rectangle from length measures of each of two adjacent sides. Using the standard interaction-effect procedure of entering the two length measures separately on Step 1 as predictors of the rectangle's area, each of those two predictors will account for substantial criterion (area measure) variance, leaving relatively little remaining variance to be accounted for when their (theoretically sufficient) multiplicative product is entered on Step 2. This standard test will give no indication that a pure multiplicative model might account for all the predictable criterion variance. In contrast, entry of the multiplicative product on Step 1 will (properly) show that it accounts for 100% of variance, leaving zero variance to be accounted for when the two side-length measures are entered individually on Step 2.

The 4-test method was described further by Greenwald, Rudman, Nosek, and Zayas (2006) in response to a skeptical appraisal of the method provided by Blanton and Jaccard (2006). Considering this past controversy, the zero-point assumption and the regression method are briefly summarized in the main text. The details of the meta-analytical result using the 4-test method are reported below. The critique by Blanton and Jaccard is considered further in this Supplemental Material, along with new evidence relevant to their critique.

Using the 4-Test Method to Test Pure Multiplicative Models

Main text provided three Equations (SG1, SA1, and GA1) which indicated how each association in a balanced identity study is embedded in an associative network that includes many other associations. Three parallel univariate regressions for Test 1—one for each of the three association measures in a balanced identity design—are described by the Equations SG2, SA2, and GA2. Derived from preceding Equations SG1, SA1, and GA1, these equations include only the measures represented by the variables with filled-triangle subscripts in Equations SG1–GA1. In Equations SG2–GA2, b_1 corresponds to $b_\blacktriangle$; b_0 represents summed effects of the unmeasured additional multiplicative predictors in the preceding equations; and "e" combines sources of random error.

$$\mathrm{SG} = \mathrm{b}_0 + \mathrm{b}_1 \cdot \mathrm{SA}{\cdot}\mathrm{GA} + \mathrm{e} \qquad \text{(SG2)}$$

$$\mathrm{SA} = \mathrm{b}_0 + \mathrm{b}_1 \cdot \mathrm{SG}{\cdot}\mathrm{GA} + \mathrm{e} \qquad \text{(SA2)}$$

$$\mathrm{GA} = \mathrm{b}_0 + \mathrm{b}_1 \cdot \mathrm{SG}{\cdot}\mathrm{SA} + \mathrm{e} \qquad \text{(GA2)}$$

Tests 2–4 of the 4-test method are produced by regression Step 2, in which the two association-strength variables that compose the multiplicative predictor on Step 1 are added as individual predictors. If a pure multiplicative model is valid, Step 2 should add zero to the variance explained in Step 1. This is relatively unlikely because of the multiple additional associations that should have impact on the criterion associations in SG2, SA2, and GA2. Appropriately more modest expectations for Step 2 are (a) that the coefficient of the product term will remain positive (Test 2), (b) that the increment in Multiple *R* due to adding the two individual predictors will be non-significant (Test 3), and (c) that the added two predictors, when tested individually in Step 2, will be non-significant (Test 4).[1]

The appropriateness of the 4-test method was contested (by Blanton & Jaccard, 2006) shortly after it was first proposed. In response, Greenwald, Nosek, and Sriram (2006) used simulations to contrast the 4-test method with Blanton and Jaccard's preferred method, which was the standard simultaneous multiple regression (SMR) significance test for a multiplicative predictor on Step 2 of a hierarchical regression. Greenwald et al. found both that the 4-test method was more sensitive to presence of a pure multiplicative model than was SMR, and that SMR suffered reduced power in detecting pure multiplicative models to the extent that means of the predictor variables deviated from their zero values (a frequent property of real data sets). This disagreement notwithstanding, the material below includes reports of results using Blanton and Jaccard's preferred SMR method, which is provided by Test 2 of the 4-test method.

Effect Size Calculations and Aggregation Methods for 4-Test Method Findings

Test 1. The effect size measure was the coefficient of the product term entered at Step 1, converted to an *r* value. This *r* was obtained separately for the three different types of criterion measures (self–group [SG], group–attribute [GA], and self–attribute [SA]) in each study and was done separately for IAT and self-report measures. In computing weighted averages as aggregate effect sizes for Test 1, each *r* was weighted by its inverse variance (n – 3), where *n* is the number of subjects in each independent sample (Hedges & Olkin, 1985).

[1] With association-strength measures, for the predictions of a positive coefficient for the product term in both Steps 1 and 2 to apply the three measures must be scored so that a combination of three positive scores defines a balanced configuration. For example, because the combination of *self* = *female*, *self* = *good*, and *female* = *good* is balanced, the measures could be scored so that each of those three associations has a numerically positive value. However, the three measures could also be scored so that any two of the three associations had negative scores, allowing four distinct scoring combinations to be used with the 4-test method. (See also Footnote 2 of the main text.)

Test 2. The effect size measure was the coefficient of the product term at Step 2, converted to a signed partial correlation (*pr*). These were weighted and aggregated as in Test 1.

Test 3. The effect size measure was derived from the test of significance of increase in variance explained at Step 2. Each *p* value was converted to a dichotomous indicator (significant versus non-significant at $p = .05$, 2-tailed). Aggregated proportions of significant results could be compared to the chance value of .05 by a binomial test.

Tests 4. The effect size measure was (as in Test 3), derived from the significance of the individual predictors added at Step 2. Each *p* value was converted to a dichotomous indicator of significant versus non-significant, which could be tested by a binomial test.

4-Test Results for IAT and Self-Report Measures

Test 1: Multiplicative product term at Step 1. For IAT measures, the weighted average *r* for the 36 Step 1 standardized regression coefficients (95% confidence intervals in parentheses) were: r_{SG} = .330 (±.039), r_{GA} = .315 (±.038), and r_{SA} = .243 (±.040; see Table S1). For the 16 samples for which tests could be done with self-report measures, weighted averaged effect sizes were: r_{SG} = .216 (±.085), r_{GA} = .201 (±.120), and r_{SA} = .190 (±.039; see Table S2).

For each of the 16 samples for which both IAT and self-report measures were available, a difference score (Z_{diff}) was computed by subtracting Fisher Z-transformed effect sizes obtained with self-report measures from those obtained with IAT measures. Weighted aggregate Z_{diff} scores were tested for difference from zero by random effects tests and were significantly greater than zero for SG measures, $Z_{diff\,SG1}$ = .159 (±.108), $p = .004$, GA measures, $Z_{diff\,GA1}$ = .155 (±.129), $p = .019$, and SA measures, $Z_{diff\,SA1}$ = .102 (±.056), $p = .0004$. These findings show that there was generally stronger evidence for the balance–congruity principle in Test 1 with IAT than with self-report measures.

Test 2: Coefficient of product term at Step 2. The partial regression coefficients in Step 2 were significantly positive for both IAT measures: pr_{SG} = .158 (±.041), pr_{GA} = .137 (±.043), and pr_{SA} = .168 (±.035), and self-report measures: pr_{SG} = .086 (±.047), pr_{GA} = .110 (±.045), and pr_{SA} = .105 (±.047). There were no significant differences between corresponding IAT and self-report partial regression coefficients, *p*s > .82.

Test 3: Significance of increase in criterion variance explained at Step 2. The proportion of significant results ($p \leq .05$, 2-tailed) for Test 3 was greater than the expected Type I error rate of 5% for both IAT (Table S1) and self-report (Table S2) measures. For IAT, the proportions of significant results at Test 3 were 22% (8/36) for tests with SG criterion measures, 25% for GA, and 25% for SA. For self-report measures the corresponding proportions were 50% (8/16) for SG, 38% for GA, and 50% for SA. These percentages were all significantly greater than 5% at $p \leq .0001$, 2-tailed. In sum, Test 3 showed some deviation from a pure multiplicative model for both IAT and self-report measures, and this deviation was substantially greater for self-report than for IAT measures.

Test 4: Statistical significance for individual predictors at Step 2. Consistent with the results for Test 3, results for binomial tests showed proportions of significant findings greater than the null value of .05 for both types of measures, with a higher proportion of significant *p* values for self-report measures. For IAT measure, the proportions of significant results at Tests 4 were 14% (5/36) and 19% for tests with SG criterion measures, 17% for both GA tests, and 17% and 19% for SA. For self-report measures, the corresponding proportions were 25% (4/16) and 31% for SG, 44% and 25% for GA, and 25% for both SA tests.

Passing of all 12 tests. The results of the 4-test method repeatedly found that support for BIT's balance–congruity principle is stronger when tested with IAT measures of association strengths than when tested with parallel self-report measures. It is easy to interpret the passing of all 12 tests (four for each of the three criterion measures) as support for a pure multiplicative model, which provides strong support for the balance–congruity principle. This was observed, remarkably, 14 times in the 36 samples for IAT measures, and once in 16 samples for self-report measures (see Table S1). The multiple confirmations of pure multiplicative models for IAT measures are remarkable because the associations in each study are embedded in multiple configurations of trios of associations (see Figure 1 in Main Text). The confirmation of a pure multiplicative model therefore suggests something that no study has yet been ambitious enough to test—the possibility that the unmeasured additional trio configurations of Equations SG1, SA1, and GA1 often, themselves, maintain consistency with one another.

Supplemental Material Pertaining to the Within-Study Meta-Analysis

For IAT data, the meta-analytic aggregate of the 36 within-study meta-analyses of Test 1 yielded a weighted average r of .285 (MOE = .029, $p < 10^{-16}$). For Test 2, the aggregation produced a weighted average pr of .152 (MOE = .035, $p < 10^{-16}$) [2]. For Test 1, 31 (86%) of the individual-study averaged r coefficients were significantly positive, and 25 (69%) of the averaged pr coefficients for Test 2 were significantly positive. Heterogeneity was non-significant for Test 1 (Q = 41.8, df = 35, p = .20) and only weakly significant for Test 2 (Q = 52.2, df = 35, p = .03).

For self-report data, the aggregation of the 16 within-study meta-analyses of Test 1 yielded a weighted average r of .201 (MOE = .065, $p = 3.47 \times 10^{-9}$). For Test 2, the aggregation produced a weighted average pr of .104, (MOE = .039, $p = 2.34 \times 10^{-7}$). For Test 1, 11 (69%) of

[2] Here and elsewhere in Results, a p value of $p < 10^{-16}$ is reported as an inequality because the meta-analysis program used to compute and test weighted average effect sizes (Lipsey & Wilson, 2001) is limited to displaying a minimum p value of 10^{-16}.

the averaged *r* coefficients were significantly positive, as were nine (56%) of the averaged *pr* coefficients for Test 2. Heterogeneity was substantial for Test 1 (Q = 101.8, df = 15, $p = 10^{-14}$), but only weakly significant for Test 2 (Q = 27.2, df = 15, p = .03).

To compare successes of Tests 1 and 2 for IAT and self-report measures, the within-study aggregate tests were examined just for the 16 samples that had both IAT and self-report measures. For both Tests 1 and 2, each of the 16 samples' difference (IAT minus self-report) in the Fisher Z effect size for the within-study aggregate was computed, and these 16 difference scores were aggregated using random effects models. For Test 1, the weighted average difference was 0.124 (MOE = 0.080, p = .002). For Test 2, the weighted average difference was 0.098 (MOE = 0.059, p = .001). These difference tests agreed with the previously described results comparing IAT versus self-report effect magnitudes for Tests 1 and 2 done separately for each of the three types of measures (SG, GA, and SA) as criterion.

Supplemental Material Pertaining to Comparing Studies Using Self-Esteem Measures With Those Using Other Self-Concept Measures

As discussed in the main text, 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 *within-study meta-analysis* method was applied to both self-esteem (k = 22) and self-concept (k = 14) measures. Second, the *aggregated mean* outcomes of Tests 1 and 2 of the 4-test method for these two groups of samples were compared meta-analytically.

Within-Study Meta-Analytic Method

IAT measures. Tests of the weighted aggregate means for Tests 1 and 2 showed that for self-esteem measures, the two weighted aggregate effect sizes were $r = .271$ for Test 1 (MOE = .041, $p < 10^{-16}$) and $r = .105$ for Test 2 (MOE = .038, $p = 10^{-7}$). Heterogeneity was non-significant for Tests 1 and 2 ($Qs \geq 23.8$, dfs = 21, ps ≥ .15). For self-concept measures, the two aggregate effect sizes were $r = .302$ for Test 1 (MOE = .042, $p < 10^{-16}$) and $r = .219$ for Test 2 (MOE = .043, $p < 10^{-16}$). Heterogeneity was non-significant for both Tests 1 and 2 ($Qs \geq 7.91$, dfs = 13, ps ≥ .41). The difference between the weighted aggregate effect sizes of self-esteem and self-concept measures was statistically significant by an independent-samples *t*-test for Test 2, $t(34) = 2.99$, p =.005, but not for Test 1, $p > .26$. For the self-esteem data, 77% of the averaged Test 1 *r* coefficients and 50% of the averaged Test 2 *pr* coefficients were significantly positive. For the self-concept data, 100% of the averaged Test 1 *r* coefficients and 93% of the averaged Test 2 *pr* coefficients were significantly positive.

Explicit measures. For studies with self-esteem measures, analyses showed that weighted aggregate effect sizes were $r = .214$ for Test 1 (MOE = .134, $p = .003$) and $r = .033$ for Test 2 (MOE = .118, $p = .58$). Heterogeneity was significant for both Tests 1 and 2 ($Qs > 13.39$, dfs = 4, ps ≥ .009). For self-concept measures, aggregate effect sizes were $r = .201$ for Test 1 (MOE = .026, $p < 10^{-16}$) and $r = .145$ for Test 2 (MOE = .024, $p < 10^{-16}$). Heterogeneity was non-significant for both Tests 1 and 2 (Qs ≥ 7.56, dfs = 10, $ps \geq .43$). The difference between the weighted aggregate effect sizes of self-esteem and self-concept measures was not statistically significant by an independent-samples *t*-test for either Test 1 or Test 2, ps >. 11. For self-esteem, 60% of the averaged *r* coefficients for Test 1 and 40% of the averaged Test 2 *pr* coefficients were significantly positive. For the self-concept data, 73% of the averaged Test 1 *r* coefficients and 64% of the averaged Test 2 *pr* coefficients were significantly positive.

4-Test Method

As a part of the analyses comparing studies using self-esteem measures with those using other self-concept measures, successes in passing the 4-test method—ranging from 0 to 4 tests passed—were compared.

IAT measures. Using the 4-test method, self-esteem measures (k = 22) passed an average of 2.68 (out of 4) tests and self-concept measures (k = 14) passed an average of 3.29 tests. This difference was not statistically significant by an independent-samples t-test, $t(34)$ = 1.68, p =.102.

Explicit measures. For success in passing the 4-test method, self-esteem measures (k = 5) passed an average of 1.80 (out of 4) and self-concept measures (k = 11) passed an average of 2.73 tests. This difference was not statistically significant by an independent-samples t-test, $t(14)$ = 1.66, p = .119.

Table S1

Effect Sizes for Each of the Four Tests of the 4-Test Method for the 36 Independent Samples Providing Implicit Data

Citation	Criterion Association Measure														
	Self-Group					Group-Attribute					Self-Attribute				
	Test 1	Test 2	Test 3	Test 4a	Test 4b	Test 1	Test 2	Test 3	Test 4a	Test 4b	Test 1	Test 2	Test 3	Test 4a	Test 4b
	r	*pr*	*p*	p_{SA}	p_{GA}	*r*	*pr*	*p*	p_{SG}	p_{SA}	*r*	*pr*	*p*	p_{SG}	p_{GA}
Aidman & Carroll (2003)	.623***	.521***	10^{-7}***	.571	10^{-5}***	.501***	-.301*	10^{-8}***	10^{-6}***	10^{-4}***	.025	.230†	10^{-5}***	.071†	10^{-5}***
Ashburn-Nardo (2010)	.320***	.215*	.293	.124	.935	.371***	.130	.775	.736	.485	.248**	.166†	.349	.185	.727
Banaji et al. (1997)	.578***	.250†	.003***	.072†	.136	.700***	.357**	.964	.813	.868	.267*	.394***	.050*	.025*	.750
Baron (2003)	.102	.157	.567	.315	.556	.122	.134	.837	.566	.916	.133	.155	.598	.321	.969
Cvencek et al. (2016, Study 1)	.615***	.293†	.445	.495	.207	.470***	.259	.164	.152	.090†	.571***	.419**	.371	.816	.181
Cvencek et al. (2016, Study 2)	.321***	-.104	.001***	.074†	2^{-4}***	.323***	-.085	.002***	10^{-4}***	.416	-.007	-.002	.299	.123	.401
Cvencek et al. (2016, Study 3)	.316*	.052	.122	.580	.042*	.321*	.058	.177	.064†	.847	.205	.170	.872	.785	.757
Cvencek et al. (2011)	.214***	.151*	.329	.141	.912	.164*	.170*	.569	.699	.365	.200***	.163*	.079†	.106	.163
Cvencek et al. (2014)	.231***	.142†	.091†	.029*	.680	.179*	.165*	.761	.817	.462	.226***	.137†	.056†	.021*	.407
Devos, Blanco, Muñoz, et al. (2008)	.205*	.093	.872	.636	.961	.207*	.084	.302	.962	.129	.201*	.179*	.099†	.655	.046*
Devos, Blanco, Rico, et al. (2008)	.335***	.296***	.779	.633	.553	.316***	.322***	.522	.664	.333	.340***	.293***	.973	.880	.848
Devos & Cruz Torres (2007, Study 1)	.571***	.236*	.769	.986	.473	.555***	.197†	.441	.228	.612	.273*	.281*	.660	.659	.668
Devos & Cruz Torres (2007, Study 2)	.458***	.351*	.299	.134	.213	.511***	.310*	.461	.216	.368	.522***	.283†	.466	.264	.584
Devos et al. (2007, Study 3)	.299*	.237†	.641	.485	.403	.341**	.307*	.309	.137	.190	.303*	.333*	.343	.178	.417
Devos et al. (2010, Study 2)	.467***	.271**	.469	.227	.627	.414***	.209*	.012*	.764	.003***	.328***	.227*	.018*	.246	.005***
Dunham et al. (2007)	.213*	.170*	.085†	.901	.029*	.162†	.162†	.029*	.028*	.147	.130	.133	.566	.901	.287
Dunham et al. (2007)	.074	.017	.400	.529	.221	.074	.027	.467	.242	.761	.032	.031	.804	.545	.751
Farnham & Greenwald (1999)	.472***	.269*	.483	.252	.607	.445***	.100	.727	.863	.480	.428***	.216†	-.591	.389	.894
Gumble & Carels (2012)	.055	-.038	.515	.768	.337	.191†	-.011	.192	.344	.071†	.150	-.029	.176	.684	.098†
Horcajo et al. (2010, Study 3)	.509*	-.155	.169	.069†	.163	.400†	-.171	.251	.141	.962	.149	-.412†	.038*	.012*	.345

Horcajo et al. (2010, Study 3)	.114	-.285	.224	.538	.089†	.067	-.264	.218	.101	.574	-.279	-.359†	.493	.362	.350
Horcajo et al. (2010, Study 4)	.397*	-.011	.524	.261	.497	.255	-.283	.127	.063†	.124	.270	-.086	.257	.183	.377
Horcajo et al. (2010, Study 4)	.268	-.054	.007**	.002**	.043*	.166	.009	.004***	.194	.003***	-.186	.085	.001***	.002***	.003***
Lane et al. (2005)	.227***	.032	.297	.529	.120	.298***	.140*	.540	.273	.601	.209***	.104	.829	.900	.541
Lane et al. (2005)	.303***	.088	.151	.909	.139	.345***	.136*	.345	.291	.815	.266***	.103	.899	.838	.833
Mellott & Greenwald (2000)	.375***	.019	.224	.089†	.317	.298***	-.112	.007**	.059†	.003***	.403***	.031	.031*	.174	.011*
Meltzoff et al. (2019)	.327***	.186†	.334	.308	.660	.349***	.154	.722	.456	.633	.236*	.219*	.650	.355	.993
Meltzoff et al. (2019)	.349***	.180†	.949	.746	.856	.244*	.109	.471	.508	.309	.222*	.161	.434	.987	.200
Nosek et al. (2002, Study 2)	.407***	.094	.295	.227	.283	.228*	.216*	.511	.344	.828	.433***	.173	.741	.445	.938
Nosek & Smyth (2011)	.309***	.219***	.029*	.056*	.016*	.347***	.265***	10^{-7}***	.005***	10^{-7}***	.380***	.227***	10^{-6}***	.153	10^{-7}***
Rudman et al. (2001, Study 4)	.363***	.390***	.088†	.029*	.589	.355***	.336***	.996	.942	.967	.296***	.343***	.094†	.035*	.772
Rudman & McLean (2013, Study 1)	.442***	.178*	.629	.957	.352	.430***	.194**	.103	.302	.184	.134†	.156*	.427	.849	.195
Schmidt & Nosek (2015)	.342***	.096***	10^{-17}***	.197	10^{-16}***	.342***	.087***	10^{-17}***	10^{-17}***	10^{-5}***	.220***	.142***	.003***	.853	.001***
Srivastava & Banaji (2011)	.135	.231*	.016*	.022*	.052†	.144	.276***	.011*	.003***	.256	.152	.225*	.080†	.025	.724
Steffens et al. (2010, Study 1)	.399***	.305***	10^{-3}***	10^{-4}***	.706	.305***	.230**	.408	.655	.289	.370***	.313***	10^{-4}***	10^{-4}***	.316
Tang & Greenwald (2013)	.317*	.045	.233	.223	.181	.309*	.066	.234	.250	.187	.331**	.086	.233	.339	.145
Average effect size	.330	.158	.315	.364	.360	.315	.137	.352	.358	.397	.243	.168	.317	.399	.436
(95% CI)	(±.039)	(±.041)				(±.038)	(±.043)				(±.040)	(±.035)			
p	10^{-38}	10^{-13}				10^{-38}	10^{-9}				10^{-38}	10^{-38}			
p[Q]	10^{-5}	.0001				.0001	10^{-5}				.0001	.023			

Note. Balanced identity design always includes measures of associations that link the concept of *self* with one *group concept* (e.g., male) and one *attribute concept* (e.g., valence); Effect sizes for Tests 1 and 2 (*r*s) are presented separately for each of the three regressions in which one measure of association strength is always entered as a criterion (e.g., measure of the *self-group* association) and the other two measures as predictors (e.g., measures of *group-attribute* and *self-attribute* associations). Test 1 is always tested at the regression Step 1 and Tests 2–4 are always tested at the regression Step 2. The weighted mean effect sizes at the first regression step (*r*), their 95% confidence intervals (CIs), transformed back to the *r* metric were computed from a random-effects test for Fisher's *Z*-transformed *r* values at Step 1 of a multiple hierarchical regression analysis. Effect sizes for Tests 3 and 4 are reported as average *p* values at Step 2 (see text for details). *pr* = signed, partial correlation coefficient for the product term at Step 2; *p* = *p* values indicating statistical significance of increase in R^2 at Step 2; p_{SG}, p_{GA} and p_{SA} = *p* values indicating statistical significance of individual SG, GA, and SA predictors added at Step 2; *p* [Q] = probability values for fixed-effects test of homogeneity (Hedges & Olkin, 1985). Bold font indicates passed tests. † = $.05 < p \leq .10$; * = $.01 < p \leq .05$; ** = $.005 < p \leq .01$; *** = $p \leq .005$

Table S2

Effect Sizes for Each of the Four Tests of the 4-Test Method for the 16 Independent Samples Providing Explicit Data

Citation	Criterion Association Measure														
	Self-Group					Group-Attribute					Self-Attribute				
	Test 1	Test 2	Test 3	Test 4a	Test 4b	Test 1	Test 2	Test 3	Test 4a	Test 4b	Test 1	Test 2	Test 3	Test 4a	Test 4b
	r	*pr*	*p*	p_{SA}	p_{GA}	*r*	*pr*	*p*	p_{SG}	p_{SA}	*r*	*pr*	*p*	p_{SG}	p_{GA}
Ashburn-Nardo (2010)	**.337*****	-.060	10^{-5}***	**.560**	10^{-4}***	**.411*****	-.187*	10^{-6}***	10^{-6}***	.002***	.094	-.060	.014*	**.565**	.019*
Cvencek et al. (2011)	**.163***	**.108**	**.065**†	.030*	**.679**	.124†	**.117**†	**.166**	**.450**	**.146**	**.209*****	**.142***	.026*	.022*	**.134**
Cvencek et al. (2014)	**.264*****	**.225*****	**.201**	**.134**	**.271**	**.225*****	.223***	**.347**	**.298**	**.240**	**.268*****	**.239*****	**.195**	**.083**†	**.495**
Devos, Blanco, Rico, et al. (2008)	.120	**.007**	**.539**	**.267**	**.632**	.077	**.035**	**.542**	**.849**	**.280**	.118	**.009**	**.354**	**.319**	**.271**
Devos & Cruz Torres (2007, Study 1)	.161	**.165**	**.161**	**.556**	**.238**	.128	**.189**†	**.393**	**.174**	**.984**	.213†	**.146**	**.514**	**.301**	**.748**
Devos & Cruz Torres (2007, Study 2)	**.526*****	-.081	**.375**	**.195**	**.405**	**.501*****	**.208**	**.266**	**.337**	**.171**	.195	**.125**	**-.424**	**.611**	**.404**
Devos et al. (2010, Study 2)	**.398*****	**.137**	.009**	.006**	**.948**	**.218***	**.170**†	**.890**	**.697**	**.909**	**.344*****	**.143**	.002***	.001***	**.977**
Farnham & Greenwald (1999)	-.132	**.082**	.022*	**.120**	**.105**	-.126	**.168**	**.080**†	**.058**†	**.262**	-.002	**.007**	**.138**	**.200**	**.989**
Mellott & Greenwald (2000)	-.001	**.139**	**.159**	**.314**	**.133**	-.002	**.203**†	**.129**	.044*	**.318**	.172	**.184**†	**.415**	**.197**	**.669**
Meltzoff et al. (2019)	.179†	**.174**†	**.767**	**.733**	**.489**	**.207***	**.225***	**.572**	**.357**	**.670**	**.238***	**.233***	**.708**	**.527**	**.578**
Meltzoff et al. (2019)	.114	-.008	**.271**	**.190**	**.713**	.035	**.051**	**.240**	**.531**	**.188**	.176†	**.024**	**.225**	**.193**	**.187**
Nosek & Smyth (2011)	**.221*****	**.138*****	10^{-13}***	10^{-11}***	10^{-4}***	**.166*****	**.168*****	.001***	10^{-4}***	**.683**	**.240*****	**.150*****	10^{-8}***	10^{-9}***	**.919**
Rudman et al. (2001, Study 4)	-.003	**.244***	10^{-4}***	**.571**	10^{-4}***	-.108	**.094**	.006**	.002***	**.365**	.176†	**.127**	**.686**	**.832**	**.470**
Rudman & McLean (2013, Study 1)	**.232*****	-.008	3^{-12}***	**.648**	10^{-12}***	**.404*****	-.098	10^{-13}***	10^{-12}***	10^{-6}***	**.177***	-.176*	10^{-7}***	**.968**	10^{-8}***
Schmidt & Nosek (2015)	**.426*****	**.043*****	10^{-227}***	10^{-14}***	10^{-191}***	**.486*****	**.119*****	10^{-158}***	10^{-154}***	10^{-4}***	**.166*****	**.170*****	10^{-19}***	10^{-19}***	10^{-10}***
Srivastava & Banaji (2011)	**.245***	**.010**	.046*	**.883**	**.075**†	**.289*****	**.061**	.006**	.041*	.015*	-.010	**.003**	.028*	**.910**	.015*
Average *r*	.216	.086	.164	.325	.293	.201	.110	.227	.240	.327	.190	.105	.180	.358	.430
(95% CI)	(±.085)	(±.047)				(±.120)	(±.045)				(±.039)	(±.047)			

p	10^{-6}	.0001	.001	10^{-6}	10^{-38}	10^{-5}
p[Q]	10^{-38}	.0001	10^{-38}	.002	.017	.001

Note. Balanced identity design always includes measures of associations that link the concept of *self* with one *group concept* (e.g., male) and one *attribute concept* (e.g., valence); Effect sizes for Tests 1 and 2 (*r*s) are presented separately for each of the three regressions in which one measure of association strength is always entered as a criterion (e.g., measure of the *self-group* association) and the other two measures as predictors (e.g., measures of *group-attribute* and *self-attribute* associations). Test 1 is always tested at the regression Step 1 and Tests 2–4 are always tested at the regression Step 2. The weighted mean effect sizes at the first regression step (*r*), their 95% confidence intervals (CIs), transformed back to the *r* metric were computed from a random-effects test for Fisher's *Z*-transformed *r* values at Step 1 of a multiple hierarchical regression analysis. Effect sizes for Tests 3 and 4 are reported as average *p* values at Step 2 (see text for details). *pr* = signed, partial correlation coefficient for the product term at Step 2; *p* = *p* values indicating statistical significance of increase in R^2 at Step 2; p_{SG}, p_{GA} and p_{SA} = *p* values indicating statistical significance of individual SG, GA, and SA predictors added at Step 2; *p* [Q] = probability values for fixed-effects test of homogeneity (Hedges & Olkin, 1985). Bold font indicates passed tests. † = $.05 < p \leq .10$; * = $.01 < p \leq .05$; ** = $.005 < p \leq .01$; *** = $p \leq .005$

Supplemental Findings Pertaining to Validity of the IAT's Zero Point

Using a higher precision test than previously available, the main text reported strong confirmations of validity of the IAT's theoretically specified (rational) zero-point location. Presented here is an additional relevant interpretation of the IAT's zero point.

Blanton et al.'s (2015) Test of the IAT's Zero-Point Interpretation

A method of assessing the validity of the IAT's zero-point was proposed by Blanton, Jaccard, Strauts, Mitchell, and Tetlock (2015). However, their method had problems that rendered it unsuitable for that purpose.

On self-report attitude measures, higher numbers typically indicate greater liking or favorableness toward the attitude's object. For example, the numerically high end of a thermometer-format measure of attitude toward a political candidate indicates maximum warmth (i.e., favorability) toward the candidate while the low end indicates maximum coldness (i.e., unfavorability). If the measure is scored from 0 to 10, the middle value (5) may be labeled "neither warm nor cold." This midpoint can be understood as an appropriate zero-point, dividing responses into favorable (>5) and unfavorable (<5) to the candidate. Similarly, the midpoint on the widely used Rosenberg (1965) self-esteem inventory, achieved by agreeing equally with self-praising and self-critical statements, is assumed to separate those who are attitudinally positive versus negative toward themselves.

One obtains a score of zero on an IAT attitude measure by responding equally rapidly in the IAT's two combined tasks. The IAT differs from the single-object thermometer measure described in the preceding paragraph because it includes two attitude objects. A political IAT might compare Candidate A with Candidate B, with zero presumably separating respondents who have more positivity toward A from those who have more positivity toward B. This zero-point is comparable to that for a thermometer-difference measure, in which one responds to a thermometer measure separately for each candidate. The thermometer difference combines these two measures into a relative preference, which produces a zero value when the two candidates have equal thermometer scores.

Blanton and Jaccard (2006) proposed that location of the zero point of IAT measures is "arbitrary" and that "the assumption that the zero point on the IAT measure maps directly onto the true neutral preference [e.g.,] for Whites over Blacks is dubious" (p. 34). Blanton and colleagues (2015) went further to say that the zero point of the race attitude IAT should be placed at a numerically positive value of the IAT's *D* measure.[3] They did not offer a psychological explanation for this presumed displacement of the zero point, but they did propose a statistical regression method to test whether the zero point was displaced in this fashion. Using data they selected to examine with their regression method, they found that the race attitude IAT had an average "right shift" (their term) of the race attitude IAT's zero point of about 1.5 standard deviations above the IAT measure's $D = 0$ value. That estimated average correction would decrease the proportion of people estimated as showing more than slight implicit White preference in the studies they reviewed (pp. 1472–1473) from an average of 83% (using an unaltered IAT *D* measure) to an average of 28%.

In Blanton et al.'s (2015) regression test method, race attitude IAT scores were regressed onto other measures that Blanton et al. believed to have (on average) valid zero points. They expected these analyses to reveal "the mean IAT score one expects to observe among individuals

[3] This assertion applied specifically to the Black–White race attitude IAT, for which a positive *D* score indicates preference for White relative to Black. Scoring direction is arbitrary for IAT measures, at the discretion of researchers. Blanton et al. were not assuming that, if this IAT were scored in the reverse direction, the zero point should be interpreted as indicating preference for Black. If the zero point is displaced from a valid value in this fashion, it would mean that the IAT identifies more persons than it should as possessing a preference for racial White.

who exhibit no behavioral preference for Whites versus Blacks.” In their expectation, an average value of zero for the intercept in this regression should indicate lack of racial preference, meaning that “behavioral neutrality map[s] onto IAT neutrality” (p. 1471).

The “logic model” underlying Blanton et al.’s (2015) regression-intercept method (p.1471) can be unpacked by (a) starting from the formula for the intercept of a bivariate regression and expressing both the IAT measure and its presumed-valid zero-value predictor (X) in standard deviation (SD) units, then (b) using this logic in both the direction tested by Blanton et al. (Equation 3) and in the reverse direction (Equation 4):

$$\text{Intercept}_{\text{IAT}} = M_{\text{IAT}} - r_{\text{X-IAT}} \times M_{\text{X}} \quad (3)$$

where M_{IAT}, M_{X}, and $r_{\text{X-IAT}}$ are (respectively) mean of IAT in SD units, mean of predictor X in SD units, and the product moment correlation between X and IAT (see, e.g., Cohen, Cohen, West, & Aiken, 2003, p. 33, combining their Equations 2.4.3 and 2.4.4).

$$\text{Intercept}_{\text{X}} = M_{\text{X}} - r_{\text{X-IAT}} \times M_{\text{IAT}} \quad (4)$$

Equations 3 and 4 can be solved to find values of M_{X} and M_{IAT} that will produce zero intercepts in both directions of regression by (a) setting both intercepts to 0 and (b) setting $r_{\text{X-IAT}}$ to values observed in the various data sets analyzed by Blanton et al. (2015). The solutions will yield values for M_{IAT} and M_{X} that should produce the desired zero values of intercepts in both directions of regression, testing Blanton et al.’s logic model. Values of $r_{\text{X-IAT}}$ for the 37 data sets in Blanton et al.’s Table 6 ranged from $r = .07$ to $r = .53$. Using either of those extreme values or any values between those, the simultaneous-equation solution is that both M_{X} and M_{IAT} must equal zero. That is, values of zero for both M_{X} and M_{IAT} allow zero intercepts to be observed in both directions. Only when $r_{\text{X-IAT}}$ approaches 1.0 can zero intercepts in both directions be observed with nonzero values of M_X and M_{IAT}, but in this case, nonzero values of the two means must be numerically equal.

Data (generously provided by Hart Blanton) for the 37 regression analyses summarized in Blanton et al.'s (2015) Table 6 were used to compute individual-study intercepts for both directions of regression. In the direction reported by Blanton et al. (regression of IAT on predictor), the weighted average intercept in SD units was 0.51, not at all close to zero. Applying Blanton et al.'s logic, 0.51 is the mean IAT score (corresponding approximately to an IAT *D* measure of 0.20) that one expects to observe among individuals who have no explicit attitude preference for Whites relative to Blacks.

Applying the regression method in the reverse direction produced a weighted average intercept of −0.01, which calls for interpretation (applying the same logic) as the mean explicit race attitude that one expects to observe among individuals who exhibit no IAT preference for Whites versus Blacks. Applying Blanton et al.'s logical model, this very close-to-zero result indicates that the IAT's zero point *is* located at an appropriate rational-zero value.

This juxtaposition of two mutually inconsistent conclusions from regression analyses computed in both directions from the same data set is, in actuality, not paradoxical. The statistics of regression intercepts oblige that, unless a regression involves two perfect measures (i.e., both test–retest reliabilities = 1.0) and a perfect correlation ($r_{X\text{-}IAT}$ = 1.0) between the two measures, the two intercepts will not be identical when the direction of regression is reversed. The data chosen by Blanton et al. were very far from meeting either the reliability criterion of perfection or the correlation criterion of perfection, obliging the conclusion that the reasoning described as the logic explaining their choice of method was not consistent with the mathematics of bivariate regressions conducted with imperfect measures.[4]

Evidence for Construct Validity of IAT-Measured Implicit Self-Esteem

IAT measures of self-esteem do not correlate highly either with self-report measures of self-esteem or with other implicit measures of self-esteem (Bosson, Swann, & Pennebaker, 2000; Buhrmester, Blanton, & Swann, 2011). Explicit self-esteem measures have also been faulted for weakness of evidence for their construct validity (Baumeister, Campbell, Krueger, & Vohs, 2003; see also Krueger, Vohs, & Baumeister, 2008; Swann, Chang-Schneider, & McClarty, 2007). The present data afforded an opportunity to examine validity of both IAT and self-report measures of self-esteem in studies of a theory (BIT) that predicts correlations involving self-esteem measures. The meta-analysis's data provided stronger evidence, in the form of larger observed effect sizes for Tests 1 and 2 (of both the 4-test method and the within-study meta-analysis method), of validity for IAT-measured implicit self-esteem than for self-report measures of explicit self-esteem. In addition, the finding that balance–congruity effects hold in studies involving self-esteem IATs (when these are analyzed as a separate group) provided the first such demonstration in analyses of balanced identity studies. It therefore provides some of the best evidence available for nomological validity of IAT self-esteem measures.

Even while supporting construct validity of IAT-measured implicit self-esteem, the meta-analytic results showed that this evidence with self-esteem measures was, in some instances, somewhat weaker than with IAT-measured self-associations involving attributes other than valence. A possible explanation (although one not testable in the meta-analysis) follows from the

[4] The average difference between means of the IAT measure and predictor in the 37 analyses of Blanton et al.'s Table 6 was 0.48 SD units, with IAT measures indicating greater White preference than did their predictors. If both IAT measures and their predictors are assumed to have valid zero points, this substantial difference between their means is a strong indication that the two measures do not measure identical constructs.

theorized centrality of self and valence in balanced identity theory. Valence is so extensively connected to identity-relevant concepts in BIT's Social Knowledge Structure (see Figure 1 in Main Text) that its associations with group and attribute concepts should have more added associative influences than do non-valence attributes. Balanced identity studies involving novel concepts that have had no chance to develop associations other than those that are experimentally established may provide an opportunity to obtain stronger confirmation of predictions involving valence associations.

References

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

Blanton, H., & Jaccard, J. (2006). Tests of multiplicative models in psychology: A case study using the unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *113*, 155-166. doi:10.1037/0033-295X.113.1.155

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

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Greenwald, A. G., Nosek, B. A., & Sriram, N. (2006). Consequential validity of the Implicit Association Test: Comment on Blanton and Jaccard (2006). *American Psychologist*, *61*, 56–61. doi:10.1037/0003-066X.61.1.56

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

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Krueger, J. I., Vohs, K. D., & Baumeister, R. F. (2008). Is the allure of self-esteem a mirage after all? *American Psychologist*, *63*, 64–65. doi:10.1037/0003-066X.63.1.64

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

Swann, W. B., Jr., Chang-Schneider, C., & McClarty, K. L. (2007). Do people's self-views matter? Self-concept and self-esteem in everyday life. *American Psychologist*, *62*, 84–94. doi:10.1037/0003-066X.62.2.84