From *Cognitive Consistency: A Fundamental Principle in Social Cognition.*
Edited by Bertram Gawronski and Fritz Strack.

8

Balanced Identity Theory

Review of Evidence for Implicit Consistency in Social Cognition

Dario Cvencek
Anthony G. Greenwald
Andrew N. Meltzoff

Balanced identity theory was originally formulated as "A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept" (Greenwald et al., 2002). In this review, we used a new name—"Balanced Identity Theory" (BIT). Aside from this name change, the underlying theory is unchanged.

BIT has roots in three major mid-20th-century theories of cognitive–affective consistency: congruity theory (Osgood & Tannenbaum, 1955), cognitive dissonance theory (Festinger, 1957), and balance theory (Heider, 1958). As described by Greenwald et al. (2002), balanced identity theory rests on three assumptions. First, *social knowledge* is defined as knowledge of persons (including self), groups, and their attributes (including valence) that can be represented as a network of associations using node (concept) and link (association) diagrams such as that in Figure 8.1. Second, the self is a central entity in the associative knowledge structure and is represented as a node that is highly connected in the structure. Third, positive and negative valence can be represented as nodes in the associative structure, permitting (for example)

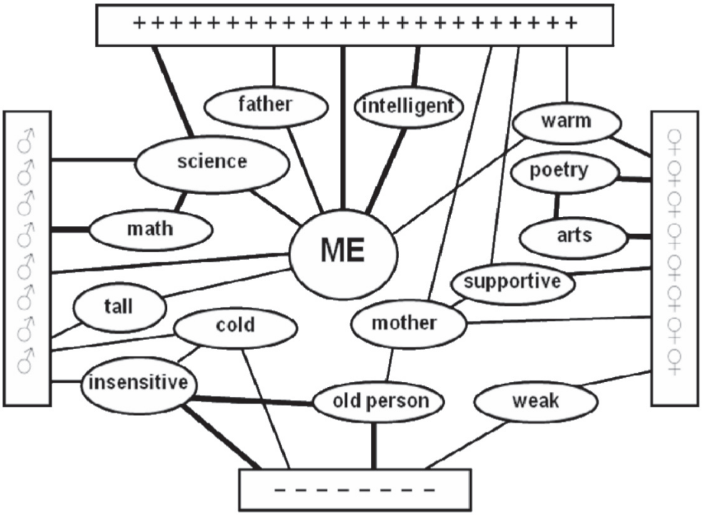


FIGURE 8.1. A social knowledge structure (SKS) of a young male assistant professor (adapted from Greenwald et al., 2002). This structure includes associations corresponding to social psychology's major affective (self-esteem and attitude) and cognitive (stereotypes and self-concept) constructs. Concepts are represented as nodes (ovals), and associative relations are represented by links (lines). Line thickness indicates association strength. The *self-concept* includes the links of the *Me* node to concepts that correspond to social categories (scientist, father) and attributes (intelligent, warm). *Self-esteem* includes the links—either direct or mediated through the self-concept—of the *Me* node to valance (+++ or –––). Analogous to self-concept, *stereotypes* are links between nodes that represent social categories and attributes. Analogous to self-esteem, *attitudes* are links, either direct or mediated through components of a stereotype, that connect social category nodes to valence nodes (+++ or –––).

the representation of self-esteem as connections of the self node to positive or negative valence nodes.

Figure 8.1, which is adapted with minor variations from Greenwald et al.'s (2002) Figure 1, displays a hypothetical social knowledge structure (SKS). Although the nodes represented in the figure comprise a small portion of any actual SKS, they suffice to illustrate the theory's representations of self-concept, self-esteem, stereotype, and attitude.

To describe expected relations among self-esteem, self-concept, stereotypes, and attitudes, BIT posits three principles that constrain associative strengths within associative structures such as SKS (Figure 8.1). This chapter focuses on the first of these, the balance–congruity principle, which has been the focus of empirical testing. Its statement, which is quoted here from the original article (Greenwald et al., 2002, p. 6), required preliminary definition of a property of associative structures.

> ***Definition 1: Shared first-order link.*** When each of two nodes is linked to the same third node, the two are said to have a shared first-order link.

> ***Principle 1: Balance–congruity.*** When two unlinked or weakly linked nodes share a first-order link, the association between these two should strengthen.

The balance–congruity name of this principle acknowledges its relation to central principles of both Heider's (1946, 1958) balance theory and Osgood and Tannenbaum's (1955) congruity theory. In the structure of Figure 8.1, given the existing links of *Me–male* (an identity) and *math–male* (a stereotype), application of the balance–congruity principle to the shared first order links of *Me* and *math* to *male* should establish or strengthen a link between *Me* and *math* (a self-concept).

The other two principles, each also accompanied by definition of a characteristic of the SKS, were formulated as follows (Greenwald et al., 2002, p. 6):

> ***Definition 2: Bipolar opposition of nodes.*** Two nodes that share fewer first-order links than expected by chance are said to be bipolar-opposed.

In the example shown in Figure 8.1, two prominent pairs of bipolar-opposed nodes in the SKS are those for valence (positive, negative) and gender (male, female). One other bipolar pair—cold and warm—represented in Figure 8.1 could easily be extended to include other bipolar pairs, such as tall–short, strong–weak, and intelligent–stupid.

> ***Principle 2: Imbalance–dissonance.*** The network resists forming new links that would result in a node forming first-order links with both of two bipolar-opposed nodes.

Principle 2 is named to acknowledge its debt to both Heider's (1958) balance theory and Festinger's (1957) dissonance theory. The resistance to new links embodied in the imbalance–dissonance principle is theoretically necessary to oppose the otherwise inevitable effect of the balance–congruity principle, in conjunction with environmental influences, to produce links among all pairs of nodes.

Situations that involve sustained external pressure toward an imbalanced configuration call for an additional principle that can avoid the sustained operation of opposing principles. The third principle (Greenwald et al., 2002, p. 6) provides this:

> ***Definition 3: Pressured concept.*** When the operation of the balance–congruity principle is causing a concept to develop links to two bipolar-opposed nodes, the concept is said to be a pressured concept.

> ***Principle 3: Differentiation.*** Pressured concepts tend to split into subconcepts, each linked to one of the two pressuring bipolar-opposed nodes.

As described earlier, the first of the three principles has been the focus of empirical testing. There has been no interest in empirically testing the imbalance–dissonance and differentiation principles. Consequently, this review focuses only on empirical research that has tested predictions generated from the balance–congruity principle.

RELATIONS TO OTHER CONSISTENCY THEORIES

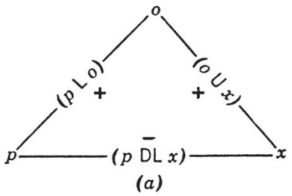
In the sections below, we draw upon the original theoretical statement to provide a brief comparison of BIT to several other consistency theories, both classical and contemporary. For a more complete review that includes a detailed discussion of theoretical similarities and differences, please see the original article (Greenwald et al., 2002).

Relation to Heider's Balance Theory

As described by Greenwald et al. (2002), the main similarities between balanced identity theory and Heider's (1946) balance theory are visible in Heider's diagrams of balanced and imbalanced configurations (reproduced here as Figure 8.2). Heider's diagrams contain representations that correspond to each of BIT's three principles: The balance–congruity principle is represented in the balanced structures *b*–*d*, the imbalance–dissonance principle in diagram *a*, and the differentiation principle in diagram *e*. The main difference between the two theories is that Heider distinguished *unit* (association) from *sentiment* (liking) links, in contrast to BIT's use of only one association type. The use of only one association type suggests that social psychology's cognitive and affective constructs are even more closely interrelated than previously conceived. Heider's discovery that many social relations can be described using just the unit and sentiment relations was a remarkably effective theoretical simplification. BIT incorporates an even more radical simplification: It collapses the distinctions between both (1) person and other concepts, and (2) unit and sentiment relations, with the goal of obtaining even broader theoretical scope (Greenwald et al., 2002). This broader scope follows from the theory's ability to account for social cognitions corresponding to attitude, stereotype, self-esteem, and self-concept using just one type of the link (i.e., an association between two concepts; see Figure 8.1).

Relation to Social Identity and Self-Categorization Theories

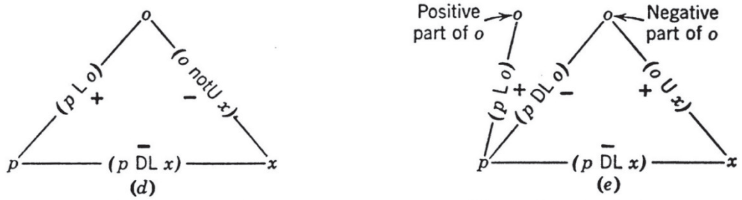
As discussed by Greenwald et al. (2002), BIT shares an underlying goal of integrating social psychology's most important constructs with two other



The given situation is unbalanced: two positive relations and one negative relation.

Change in sentiment relation resulting in a balance of three positive relations.

Change in sentiment relation resulting in a balance of two negative relations and one positive relation.



Change in unit relation resulting in a balance of two negative relations and one positive relation.

Change in unit relation through differentiation resulting in a balance of two negative relations and one positive relation.

FIGURE 8.2. Heider's representation of principles of cognitive consistency. Copyright 1958 by John Wiley & Sons, Inc. Reproduced with permission from Dr. Karl G. Heider.

well-established theoretical bodies of research on social identity: Tajfel's social identity theory (SIT; Tajfel, 1982) and Turner's self-categorization theory (SCT; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). All three theories (BIT, SIT, and SCT) assume a close relation between group membership and self-esteem, and all three make at least two similar predictions involving self-esteem, ingroup identity, and ingroup attitude. First, all three theories predict that membership in a valued group will enhance self-esteem. Second, all three theories predict that people who identify strongly with a group to which they belong should display more positive attitudes toward that group (relative to those who identify weakly with the same group).

There are some structural differences between SCT and BIT. Whereas the representational elements of the SCT are *self-categorizations*, BIT takes *associations* as its conceptual building blocks. In addition, within SCT, the self is conceived of as a hierarchical structure of self-categorizations at three levels of abstraction; within BIT, the self is understood as a nonhierarchical, associative structure.

There is an even more substantial difference between SIT and BIT, which is most apparent in how the SIT and BIT treat self-esteem in relation to how strongly the individual identifies with a novel group. This difference is best exemplified in how the two theories account for the role of self-esteem in the *minimal group* phenomenon, which is the tendency to favor one's own group relative to other groups, possibly based on arbitrary and virtually meaningless distinctions between groups (Tajfel, Billig, Bundy, & Flament, 1971). BIT treats self-esteem as an associative connection of self to positive valence, and the balance–congruity principle calls for the link between the novel self-associated group and positive valence to be strengthened by the link of self to positive valence. In contrast, SIT treats self-esteem as a motivational force that leads people to use group identities to generate positive self-regard either by viewing their ingroups positively or viewing outgroups negatively. Consequently, and in contrast with BIT's expectation that the valence attached to a novel self-associated group should be greater for those with high than for those with low self-esteem, SIT predicts the reverse—that those who have low self-esteem should develop more attraction to a novel self-associated membership group.

Perhaps the greatest difference between the SIT and SCT on the one hand, and BIT on the other, comes from the research methods used in testing the theories. The research programs of SIT and SCT were developed well before researchers recognized the distinction between implicit and explicit measures. Consequently, research on SIT and SCT has occurred mostly with explicit measures. In contrast, tests of BIT have been carried out with both implicit and explicit measures, leading to (so far) consistent results showing that the relationships predicted by BIT are evident more strongly when tested with implicit measures of association strengths than when tested with parallel self-report measures.

METHODS FOR EMPIRICAL TESTS OF BIT

As described by Greenwald et al. (2002), self-report measures are not necessarily preferred for testing BIT's predictions for two reasons. First, subjects may have no introspective access to some of the associative links of SKS (cf. Greenwald & Banaji, 1995). Second, self-report measures are susceptible to artifacts (especially impression management) that may distort assessment of associative links even when they are introspectively available. Consequently, empirical tests of the balance–congruity principle have made use of the Implicit Association Test, a recently developed alternative to self-report methods (Greenwald, McGhee, & Schwartz, 1998).

The Implicit Association Test (IAT)

The IAT (Greenwald et al., 1998) is a computerized categorization task that measures relative strengths of associations among concepts. An IAT measure of association strengths is calculated by comparing the speed with which people categorize exemplars from four categories under two instructional conditions that vary assignments of the four categories to two computer keyboard responses. The measure is based on the principle that subjects should find it easier to give the same response to items from two categories when the two categories are associated than when they are not (Greenwald et al., 1998). IAT measures provide relative, not absolute, measures of association strengths. For example, an IAT measure of self-esteem assesses strength of the *Me–positive* and *other–negative* associations relative to the strengths of *Me–negative* and *other–positive* associations.

Statistical Testing of the Balance–Congruity Principle

BIT's balance–congruity principle can be tested by using the *balanced identity design* (BID) introduced by Greenwald et al. (2002). The BID requires measurement of the strengths of the associations among all pairs of three concepts. One of these concepts is always the self, and the other two are a social category, such as a group membership, and an attribute expected to be associated with that group. The three associations can be identified as self–group (SG; corresponding to *identity*), group–attribute (GA; corresponding to *attitude* toward or *stereotype* of the group), and self–attribute (SA; corresponding to *self-esteem* or *self-concept*).

Greenwald et al. (2002) described a four-test sequence that statistically assesses whether the interrelations among the BID's three measures of association strength reflect the operation of the balance–congruity principle. With the measures of SG, GA, and SA associations, this analysis can be done using, in turn, each of the three association measures as a criterion in a hierarchical regression in which, in the first step, the criterion association's strength is predicted from the product of the strengths of the other two. In the second step, the two predictor associations are entered singly. If it can be assumed that the associations are measured on scales with rational zero points that identify a point of equality of strengths of the sets of associations contained in the measure (e.g., self–positive and self–negative in a self-esteem measure), the prediction is a significant effect of the product term on the first step, and no additional variance is predicted by the component associations on the second step (Greenwald et al., 2002, p. 11).

When the three association measures in the BID (i.e., SG, GA, and SA associations) are scored so that high scores correspond to greater association of self with the "ingroup" (relative to the "outgroup") on the SG measure, more positive evaluation of the "ingroup" (relative to the "outgroup") on the GA measure, and more positive evaluation of the self (relative to others) on the SA measure, each of the three two-step regressions provides four tests: (1) the multiple *R* should have a statistically significant and numerically positive regression coefficient at Step 1; (2) the product term's coefficient should remain numerically positive at Step 2; (3) the increase in criterion variance explained at Step 2 should not be statistically significant; and (4) neither regression coefficient associated with the individual predictors should differ from zero at Step 2. This four-test sequence evaluates a *pure multiplicative* model, which asserts that the multiplicative product of two measures is the sole predictor of a criterion measure. This method bypasses the standard regression procedure of testing significance of a product term after first entering its component variables as predictors. Explanation of the 4-test procedure is given briefly in Greenwald et al. (2002, pp. 9–11) and at greater length by Greenwald, Rudman, Nosek, and Zayas (2006).

QUANTITATIVE (META-ANALYTIC) REVIEW OF EMPIRICAL FINDINGS

Search Method

Studies were initially sought using three methods: (1) PsycINFO search (using the keywords *cognitive balance, cognitive consistency, balanced identity, IAT, Implicit Association Test, implicit attitude, implicit identity, implicit self-esteem, implicit stereotype, implicit self-concept, 3 IATs, 3 Implicit Association Tests*), (2) PubMed search (using the same keywords as in the PsycINFO search), and (3) Internet search (using Google Scholar, using the same keywords as in the preceding two searches). In addition, the PsycINFO database was used to identify studies that referenced Greenwald et al. (2002). The search produced 17 reports containing 20 independent samples (Lipsey & Wilson, 2001, p. 112). For two of the 17 reports, the information needed for the meta-analysis was no longer available (Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002; Rudman & Goodwin, 2004). Consequently, the meta-analysis reported below was conducted on 18 independent samples with a total of 1,913 subjects (see Table 8.1).[1]

Calculation of Effect Sizes

Two of the statistical tests used in testing for fit of balanced identity results to multiplicative prediction are based on magnitudes of multiple *R* coefficients. These are at regression Step 1 (expected significance of multiplicative predictor) and Step 2 (expected nonsignificance of added predictors). Mean effect sizes (*r*'s) for the results at Step 1 were computed from the standard-

TABLE 8.1. Effect Sizes and Characteristics of the 18 Independent Samples Providing Implicit Data

					Criterion association measure					
					Self–group		Group–attribute		Self–attribute	
					Effect size		Effect size		Effect size	
Citation	Subjects	*N*	Group concept	Attribute concept	*r* (Step 1)	*z* (Step 2)	*r* (Step 1)	*z* (Step 2)	*r* (Step 1)	*z* (Step 2)
Aidman & Carroll (2003)	Males, females	66	Female	Pleasant	**.55*****	**–3.63*****	**.45*****	**–4.07*****	.17	**–3.77*****
Banaji et al. (1997)	Whites, Blacks	61	White	Positive	**.58*****	**–2.80*****	**.70*****	1.44	**.27***	**–1.65***
Cvencek et al. (2011)[a]	Girls, boys	222	Boy	Math	**.21*****	–.45	**.16***	.21	**.20*****	**–1.44†**
Devos & Cruz Torres (2007)[a]	Latinos, Whites	80	White	High achievement	**.57*****	.74	**.56*****	–.14	**.27***	.41
Devos & Cruz Torres (2007)[a]	Latinos	49	Significant other	High achievement	**.52*****	–.10	**.51*****	–.10	**.46*****	–.53
Devos et al. (2007)[a]	Females	60	College education	Pleasant	**.30***	.36	**.34****	–.48	**.30***	–.41
Devos, Blanco, Muñoz, et al. (2008)[a]	Males, females	169	Woman	College education	**.34*****	.77	**.32*****	.05	**.34*****	1.54
Devos, Blanco, Rico, et al. (2008)[a]	Bilingual Latinos	128	School	Pleasant	**.21***	1.15	**.21***	–.51	**.20***	**–1.29†**
Devos et al. (2010)	Latinos, Whites	108	White	American	**.47*****	.08	**.41*****	**–2.26***	**.33*****	**–2.10***
Dunham et al. (2007)	Hispanic children, Hispanic adults	129	Hispanic	Good	.06	–.18	.05	–.83	.00	–.24
Dunham et al. (2007)[a]	Hispanic children, Hispanic adults	137	Hispanic	Good	**.23****	**–1.45†**	**.20***	**–1.51†**	**.16†**	.78
Farnham & Greenwald (1999)[a]	Females	65	Female	Positive	**.47*****	–.09	**.45*****	.55	**.43*****	.04
Lane et al. (2005)[a]	Yale students	235	Yale	Good	**.31*****	–1.11	**.37*****	–.28	**.29*****	.58

(continued)

TABLE 8.1. (*continued*)

					Criterion association measure					
					Self–group		Group–attribute		Self–attribute	
					Effect size		Effect size		Effect size	
Citation	Subjects	*N*	Group concept	Attribute concept	*r* (Step 1)	*z* (Step 2)	*r* (Step 1)	*z* (Step 2)	*r* (Step 1)	*z* (Step 2)
Lane et al. (2005)	Yale students	215	Residential college	Good	__.24***__	__–1.87*__	__.30***__	–.65	__.22***__	–.22
Mellott & Greenwald (2000)	Undergraduates	98	Old	Positive	__.38***__	–.60	__.30***__	__–2.52**__	__.40***__	__–1.91*__
Nosek et al. (2002)[a]	Females	91	Male	Math	__.41***__	–.64	__.23*__	.59	__.43***__	.32
Rudman et al. (2001)[a]	Females	68	Male	Potent	__.36***__	1.95	__.36**__	–.92	__.30*__	–.88
Spalding & Kaiser (2011)[a]	Female students in STEM disciplines	147	Female	Science	__.23***__	1.49	__.32***__	.89	__.29***__	.97
				Average *r*	.346		.340		.270	
				(95% CI)	(±.067)		(±.071)		(±.051)	
				SD	.151		.159		.111	
				p[Q]	.00012		.00002		.08	

Note. Subjects, subject groups comprising independent samples in each study; *N*, number of subjects contributing to each balanced identity analysis; STEM, Science, technology, engineering, and mathematics. Balanced identity design always includes measures of associations that link the concept of *self* with one *group concept* (e.g., female, school, Hispanic) and one *attribute concept* (e.g., valence, math). Effect sizes (*r*) 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). The weighted mean effect sizes at the first regression step (*r*), their 95% confidence intervals (CIs), and their weighted standard deviations (*SD*s), 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. $z = p$ values indicating statistical significance of increase in R^2 at Step 2, converted to their z scores; *p*[Q] = probability values for fixed-effects test of homogeneity (Hedges & Olkin, 1985). **Bold** font indicates statistically significant effect sizes ($p \leq .05$).

[a] These samples confirmed balance–congruity expectations in Step. 1 and 2 for all three regressions.

† = $.05 < p < .10$; * = $.01 < p < .05$; ** = $.005 < p < .01$; *** = $p < .005$.

ized *b* coefficient in Step 1 (which are equivalent to a signed correlation coefficient, *r*). An inverse variance weight was computed for each mean *r* as ($n - 3$), with *n* being the number of subjects in the independent sample (Hedges & Olkin, 1985).

For the regression Step 2, the magnitude of effect sizes is not as meaningful as that at the regression Step 1, primarily because of the variety of ways in which the two added predictors can produce effects. The Step 2 results were therefore examined only in terms of whether the increment was statistically significant. The statistical significance of *R* increase at Step 2 was calculated from multiple regression results with the following formula:

$$F \text{ change} = [(R_2^2 - R_1^2)/(k_2 - k_1)]/[(1 - R_2^2)/(n - k_2 - 1)]$$

where R_1 and R_2 are multiple correlation coefficients at regression Step 1 and 2, respectively; k_1 and k_2 are number of predictors at regression Step 1 and 2, respectively; and *n* is the sample size. The two-tailed probabilities associated with the *F* change values were converted to one-tailed *p* values for meta-analytic use. With this step and probit conversion of one-tailed *p* values to *z*-value effect sizes, an *F* change of 0 (i.e., zero increment in explained variance at Step 2) is appropriately represented as an effect size of zero.

Aggregate Effect Sizes and Homogeneity Tests

The weighted average effect sizes of (signed) regression coefficients in Step 1 (with 95% confidence intervals) for implicit measures, aggregated across all available independent samples ($k = 18$), were close to the conventionally moderate value of $r = .3$: $r_{SG} = .346$ (±.067), $r_{GA} = .340$ (±.071), and $r_{SA} = .270$ (±.051). All three types of effect size were (1) significantly heterogeneous when tested with fixed-effects models (see bottom three rows of Table 8.1) and (2) significantly different from zero in the positive direction by a random effects test (all *p*'s < .0001).[2] A repeated-measures analysis of variance (ANOVA) was conducted on *r*-to-*Z* transformations of effect sizes at Step 1 to test for differences of the multiplicative product term at Step 1 among the three types of associations (i.e., SG, GA, and SA). This analysis found no effect of association type ($p = .11$).

Effect sizes of (signed) regression coefficients at Step 1 for explicit measures, aggregated across all available independent samples ($k = 8$), were close to conventionally small value of $r = .1$: $r_{SG} = .141$ (±.106), $r_{GA} = .088$ (±.074), and $r_{SA} = .182$ (±.120). All three types of effect size were significantly heterogeneous when tested with fixed-effects models (see bottom three rows of Table 8.2). The weighted average regression coefficients for explicit measures were significantly different from zero in the positive direction for SG and SA measures ($p = .009$ and $p = .003$ for SG and SA measures, respectively), but not for GA measures ($p = .09$) when tested with a random effects test.

TABLE 8.2. Effect Sizes and Characteristics of the Eight Independent Samples Providing Explicit Data

					Criterion association measure					
					Self–group		Group–attribute		Self–attribute	
					Effect size		Effect size		Effect size	
Citation	Subjects	*N*	Group concept	Attribute concept	*r* (Step 1)	*z* (Step 2)	*r* (Step 1)	*z* (Step 2)	*r* (Step 1)	*z* (Step 2)
Cvencek et al. (2011)	Girls, boys	220	Boy	Math	**.16***	**–1.46†**	**.12†**	–.85	**.21****	**–2.08***
Devos & Cruz Torres (2007)	Latinos, Whites	80	White	High achievement	.16	–1.00	.13	–.28	**.21†**	.03
Devos & Cruz Torres (2007)	Latinos	49	Significant other	High achievement	.20	–.42	**.41*****	**–1.98*****	**.52*****	–.36
Devos, Blanco, Muñoz, et al. (2008)	Males, females	169	Woman	College education	.12	.11	.08	.09	.12	–.37
Devos et al. (2010)	Latinos, Whites	108	White	American	**.40*****	**–2.34***	**.22***	1.21	**.34*****	**–2.87*****
Farnham & Greenwald (1999)	Females	65	Female	Positive	–.15	**–2.44*****	–.15	**–1.64†**	–.15	**–2.44****

Mellott & Greenwald (2000)	Undergraduates	98	Old	Positive	.00	**–1.33†**	.00	–.71	**.17†**	–.69
Rudman et al. (2001)	Females	68	Male	Potent	.18	**.76**	–.11	**–1.99***	–.00	.04
				Average *r*	.141		.088		.182	
				(95% CI)	(±.106)		(±.074)		(±.120)	
				SD	.138		.131		.157	
				p[Q]	.0256		.042		.0042	

Note. Subjects, subject groups comprising independent samples in each study; *N*, number of subjects contributing to each balanced identity analysis. Balanced identity design always includes measures of associations that link the concept of *self* with one *group concept* (e.g., female, school, Hispanic) and one *attribute concept* (e.g., valence, math). Effect sizes (*r*) 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). The weighted mean effect sizes at the first regression step (*r*), their 95% confidence intervals (CIs), and their weighted standard deviations (*SD*s), 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. $z = p$ values indicating statistical significance of increase in R^2 at Step 2, converted to their *z* scores; *p*[Q] = probability values for fixed-effects test of homogeneity (Hedges & Olkin, 1985). **Bold** font indicates statistically significant effect sizes ($p \leq .05$).

† = $.05 < p < .10$; * = $.01 < p < .05$; ** = $.005 < p < .01$; *** = $< .005$.

Confirmation of the Expected Data Patterns for Balanced Identity

Multiplicative Product Term at Step 1

One expectation of a pure multiplicative model is that the data of the BID should be fit entirely by the multiplicative product term. Statistically, this translates to the expectation that the multiple R associated with the product term should be statistically significant at Step 1, with no significant increase in R from adding the individual predictors on Step 2. Data obtained with the implicit measures supported this expectation more strongly than the parallel data obtained with explicit measures. A 3 (Association type: SG, GA, and SA) × 2 (Measure type: Implicit and explicit) repeated-measures ANOVA was conducted on r-to-Z transformations of effect sizes at Step 1 using only the eight samples for which both implicit and explicit measures were available. This analysis was done to test for (1) differences of the multiplicative product term at Step 1 among the three types of associations (i.e., SG, GA, and SA) for both implicit and explicit measures, and (2) differences between implicit and explicit measures.

The results suggest that, while effect sizes at Step 1 were larger for implicit than for explicit measures, they did not vary as a function of the association type. These results indicated a main effect of measure type (i.e., implicit or explicit), $F(1, 7) = 18.04$, mean square error (MSE) = 0.95, $p = .004$. There was no main effect of criterion association type (i.e., SG, GA, or SA) ($p > .38$), nor a measure × association type interaction ($p > .10$).

Statistical Significance at Step 2

For the regression Step 2, at which no significant increase in prediction is expected from adding the individual predictors, the result is useful only in terms of whether it is statistically significant (as described earlier). To test for statistical significance at Step 2, the p values from Step 2 were first converted to z values, which were then summed. The resulting sum was averaged (i.e., divided by the n number of p values) and that average was converted to the p value. This allowed for describing the average p values, which for the three implicit measures were: $p_{SG} = .36$, $p_{GA} = .28$, and $p_{SA} = .29$, none of which is close to the statistically significant level of $p = .05$. This was also true for the average p value for the Step 2 tests with explicit measures: $p_{SG} = .16$, $p_{GA} = .22$, and $p_{SA} = .14$, none of which approximated $p = .05$.

In addition, to test for the effects of the design factors on statistical significance, the same 3 × 2 repeated-measures ANOVA was conducted on z transforms of the p values obtained at Step 2. There were no main effects for either measure ($p = .08$) or association type ($p = .31$), nor was there an interaction of measure × association type interaction ($p = .23$).

Multiple Rs at Steps 1 and 2

The expectation of the pure multiplicative model regarding the multiple *R*s at regression Steps 1 and 2 can be examined by how often the two crucial tests of Greenwald et al.'s (2002) four-test method were passed. The two statistical tests presented in this section are the ones based on magnitudes of multiple *R* coefficients and passing of both tests can be used to test for fit of balanced identity results to multiplicative prediction (the full four-test method is detailed only in part in the section below, because we did not have access to the complete four-test results for all of the studies in the meta-analysis).

Table 8.1 displays the statistical significance of effect sizes at regression Step 1 and 2 for the 18 independent samples included in this report using implicit measures. Regression analyses from the 18 samples for which implicit data were available provided 54 opportunities to confirm the theoretical expectations at both regression steps. Implicit measures confirmed the expected pattern in 41 of these 54 opportunities (76%). In contrast, analyses from the eight samples for which explicit data were available (see Table 8.2) provided 24 opportunities, with the expected pattern confirmed only five times (21%). When implicit analyses were limited only to the eight independent samples for which explicit data were available, regressions confirmed the expected theoretical pattern 20 times in the 24 opportunities (83%; see Table 8.1). These results confirm previous reports that evidence conforming to the balance–congruity principle is stronger on implicit than on corresponding explicit measures (Greenwald et al., 2002, 2006).

Passing of the Four-Test Method

The expectation of the pure multiplicative model can also be examined by how often *all four tests* of Greenwald et al.'s (2002) four-test method were passed. For the seven studies for which we had results for all four tests from each of the three regression analyses using implicit measures (Banaji, Greenwald, & Rosier, 1997; Cvencek, Meltzoff, & Greenwald, 2011; Farnham & Greenwald, 1999; Mellott & Greenwald, 2000; Nosek, Banaji, & Greenwald, 2002; Rudman, Greenwald, & McGhee, 2001; Spalding & Kaiser, 2011), results indicated that of the 21 possible opportunities to pass all four-tests, implicit data passed 15 (71%). In contrast, for the four studies for which we had results for all four tests from each of the three regression analyses using explicit measures (Cvencek et al., 2011; Farnham & Greenwald, 1999; Mellott & Greenwald, 2000; Rudman et al., 2001), of the 12 opportunities to pass all four-tests, explicit data passed only two (17%).

Zero-Order Correlations

Another expectation of the balance–congruity principle's multiplicative model is that the zero-order (i.e., bivariate) correlations between any two of the three association strength measures in the BID should have the same sign as the mean value of the remaining association's measure, when that value is measured on a scale for which zero indicates equality of contrasted association strengths (see Greenwald et al., 2002, pp. 11–12). Data obtained with implicit measures conformed closely to this expectation, whereas data obtained with explicit measures did not. Figure 8.3 displays these results. Fit with prediction for the three types of implicit measures (Panels A, B, and C) is indicated by significant positive regression slopes (average $r = .74$, Stouffer method combined $p = .00038$; individual r's $\geq .66$, individual p's $\leq .003$) that do not deviate significantly from passing through the origin. Conversely, fit for the three explicit measures (Panels D, E, and F) was quite poor, as indicated by nonsignificant regressions (average $r = .19$, Stouffer method combined $p = .69$; r's $\leq .35$; individual p's $\geq .40$). These results confirm other indications that fit with the balance–congruity principle is evident with implicit, but not explicit, measures.

CONCLUSIONS

The main goal of this meta-analysis was to test predictions of the balance–congruity principle with implicit and explicit measures. These predictions were supported more strongly in the data obtained with implicit measures, as indicated, in part, by larger average effect sizes (r's) for implicit than for explicit measures at the first regression step, at which only the product term is entered. The weighted average of these SG, GA, and SA effect sizes for the implicit measures, based on 17 independent samples, were $r_{SG} = .346$, $r_{GA} = .340$, and $r_{SA} = .270$, levels close to the conventional "moderate" value of $r = .30$ (Cohen, 1988). In contrast, weighted average effect sizes for the parallel self-report measures were considerably smaller—close to the conventional "small" value of $r = .1$: $r_{SG} = .141$, $r_{GA} = .088$, and $r_{SA} = .182$. This difference in effect size for the product term at Step 1 was statistically significant.

Also consistent with the conclusion that expectations of the balance–congruity principle were better fit by data for implicit measures, implicit measures showed: (1) substantially more frequent confirmation of the combined expectation of statistical significance of the product term as the sole predictor (at Step 1) and (2) nonsignificant increment in R when the product term's component associations were added as individual predictors (Step 2). In addition, signs of the zero-order correlations between any two of the three association strength measures in the BID corresponded to the sign of the measure of the remaining association measured on a scale with rational zero value (see Figure 8.3).

Also noteworthy is that the pattern of confirmation of balance–congruity expectations in both Step 1 and 2 for all three regressions in the same study was observed for implicit measures in 12 of the 18 samples, and was very close to that in three others (Banaji et al., 1997; Devos, Gavin, & Quin-

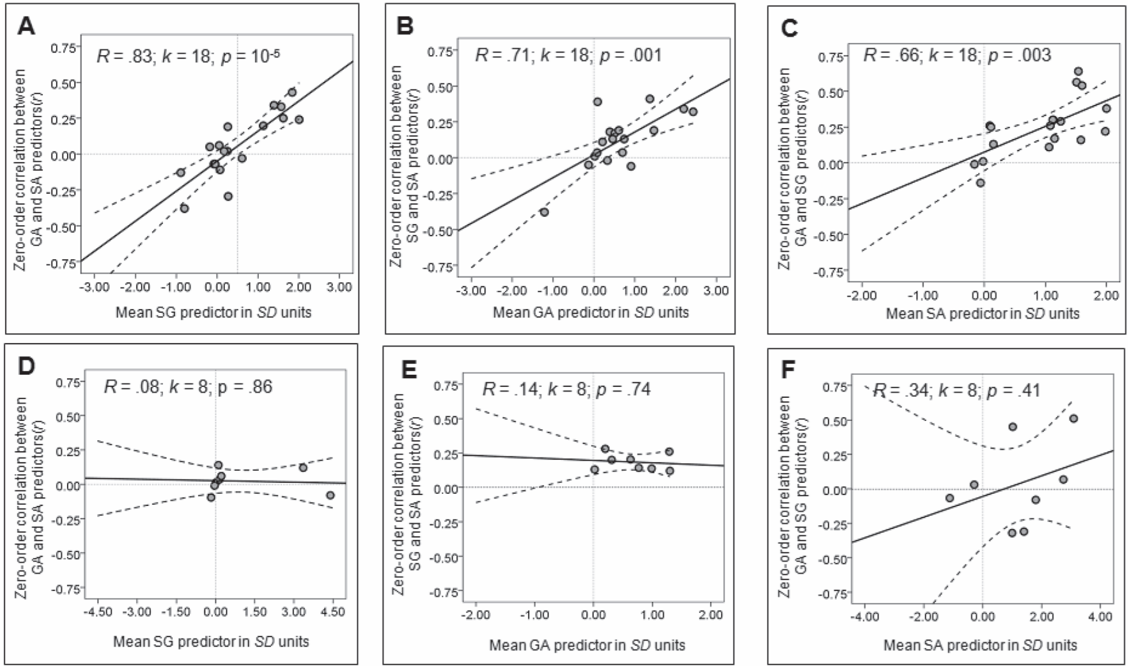


FIGURE 8.3. Zero-order correlations between pairs of association strength measures as a function of the third predictor, measured on scales with rational zero points. Dots indicate independent samples. Panels A, B, and C are for implicit measures, and Panels D, E, and F are for explicit measures. The two-letter codes indicate the two types of concepts constituting each of the three possible predictors. SG, self–group (identity) association; GA, group–attribute association; SA, self–attribute association. The regressions test the hypothesis that the zero-order correlations of any pair of association strengths should have the same sign as the mean value of the remaining predictor, when the latter is measured on a scale with a rational zero point. *r* is the zero-order correlations coefficient; *k* is the number of independent samples; *p* is the statistical significance of the correlation. The 95% confidence intervals (dashed curves) indicate that regression slopes for Panels A, B, and C do not significantly deviate from the expectation that they should pass through the origin.

tana, 2010; Lane et al., 2005). This pattern was not confirmed in full for any of the eight samples for explicit measures, was very close to confirmation in only one sample (Cvencek et al., 2011), and was not close to confirmation in any others.

Greenwald et al. (2002) attributed the relatively poor fit of explicit measure findings to predictions from the balance–congruity principle to: (1) introspective limits that may render association strengths inaccessible to measurement by self-report and (2) response factors, such as demand characteristics and evaluation apprehension, that may distort self-report measures (p. 17). The only study for which regressions involving all three explicit measures came very close to full confirmation of the pure multiplicative model was also the only study in which subjects were young children (Cvencek et al., 2011; see Table 8.2). This may indicate that response factors that may distort results with explicit measures in adult samples are less of an interfering factor in children. Because there is no other evidence for that interpretation, it should be regarded as speculation that awaits the appearance of additional relevant studies.

The primary focus of this chapter is on correlational designs and multiplicative regression patterns and, as such, we have not included a review of other, potentially related work. Notably, experimental studies that used implicit measures other than the IAT provided evidence that novel associations between the self and nonsocial categories also lead to an associative transfer of implicit self-evaluations to the objects associated with the self. For example, using an affective priming paradigm, Gawronski, Bodenhausen, and Becker (2007) showed that choosing an object leads to the formation of an association between the chosen object and the self. In addition, using an affective misattribution paradigm, Prestwich, Perugini, Hurling, and Richetin (2010) reported a positive implicit evaluation of an object when that object had previously been paired with one's self. Finally, using an evaluative conditioning paradigm, Zhang and Chan (2009) showed that pairing words related to the self (unconditioned stimulus; US) with a previously neutral stimulus (conditioned stimulus; CS) changes the valence of the CS in the direction of the evaluation of the US. Taken together, these studies suggest that the basic assumptions of BID are not restricted to a single research paradigm (i.e., the BID) and a single measure (i.e., the IAT), but may generalize to other experimental paradigms and implicit measures, thus highlighting both the theoretical contribution as well as the integrative power of BIT.

SUMMARY

This review adds to and strengthens previous indications that evidence for BIT is stronger when tested with IAT measures of association strengths than when tested with parallel self-report measures. There were four relevant results. First, the average effect sizes obtained with implicit measures were larger at the first regression step (weighted average $r = .32$) than those obtained with corresponding explicit measures (weighted average $r = .14$). Second, the implicit measures confirmed the expectation of the pure multiplicative model at regression Step 1 and 2 considerably more often than did the explicit measures. Third, the data obtained with the implicit (but not explicit) measures conformed to the expectation that the sign of the zero-order correlations between any two of the three association strength measures in the BID should correspond to the mean value of the measure of the third association. Additionally, the meta-analysis also indicated that the implicit measure findings did not vary by association type, supporting the expectation that the three association measures in a BID are effectively interchangeable in their roles in data analysis. The clear findings of this review indicate that when tested with implicit measures, BIT's balance–congruity principle effectively unifies social psychology's major cognitive (stereotype and self-concept) and affective (attitude and self-esteem) constructs.

NOTES

1. The two rows for the Lane, Mitchell, and Banaji (2005) study are for samples that differed only because of different patterns of missing data from the same larger group of subjects. These were nevertheless treated as independent samples, because the hypotheses tested were sufficiently different that it seemed inappropriate to average them into a single sample.
2. A nonsignificant Q does not always warrant a conclusion that a fixed-effects model is justified. With small numbers of effect sizes, such as the number of studies reported here, the homogeneity test may lack sufficient statistical power to reject homogeneity even when the variability among the effect sizes is considerable and due to factors other than subject-level sampling error (Lipsey & Wilson, 2001, p. 117). Consequently, random-effects estimates are reported for analyses involving weighted average effect sizes.

REFERENCES

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.

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cvencek, D., Meltzoff, A. N., Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development, 82*, 766–779.

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*, 449–471.

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.

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.

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.

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.

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.

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

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

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.

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

Greenwald, A. G., Banaji, M. R., 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. *Psychological Review, 109*, 3–25.

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

Greenwald, A. G., 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.

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

Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21*, 107–112.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17*, 482–495.

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.

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

Mellott, D. S., & Greenwald, A. G. (2000). *But I don't feel old! Implicit self-esteem, age* *identity and ageism in the elderly.* Unpublished manuscript, University of Washington, Seattle.

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.

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

Prestwich, A., Perugini, M., Hurling, R., & Richetin, J. (2010). Using the self to change implicit attitudes. *European Journal of Social Psychology, 40*, 61–71.

Rudman, L. A., & Goodwin, S. A. (2004). Gender Differences in automatic in-group bias: Why do women like women more than men like men? *Journal of Personality and Social Psychology, 87*, 494–509.

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.

Spalding, K., & Kaiser, C. R. (2011). *Stereotypes about gender and science interfere with science identification for women in STEM fields*. Manuscript in preparation.

Tajfel, H. (Ed.). (1982). *Social identity and intergroup relations*. Cambridge, UK: Cambridge University Press.

Tajfel, H., Billig, M. G., Bundy, R. F., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Psychology, 1*, 149–177.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group*. Oxford, UK: Blackwell.

Zhang, H., & Chan, D. K.-S. (2009). Self-esteem as a source of evaluative conditioning. *European Journal of Social Psychology, 39*, 1065–1074.

Guilford Publications
72 Spring Street
New York, NY 10012
212-431-9800
800-365-7006
www.guilford.com