

Why So Little Faith? A Reply to Blanton and Jaccard's (2006)

Skeptical View of Testing Pure Multiplicative Theories

Anthony G. Greenwald

University of Washington

Laurie A. Rudman

Rutgers University

Brian A. Nosek

University of Virginia

Vivian Zayas

University of Washington

Address correspondence to:

Anthony G. Greenwald  
Department of Psychology  
University of Washington  
Box 351525  
Seattle, WA 98195-1525

email: [agg@u.washington.edu](mailto:agg@u.washington.edu)

phone: (206) 543-7227

fax: (206) 685-3157

Running Head: Testing Pure Multiplicative Theories

Why So Little Faith? A Reply to Blanton and Jaccard's (2006)  
Skeptical View of Testing Pure Multiplicative Theories

Abstract (123 words; 832 characters)

H. Blanton and J. Jaccard (2006) questioned the 4-test regression method used by A. G. Greenwald et al. (2002) to test a pure multiplicative theory. We address their concerns with a combination of simulations and meta-analysis. Simulations show that (a) Blanton and Jaccard's preferred simultaneous regression method suffers severe power loss in testing multiplicative theories when predictor variables' means deviate from rational zero values, and (b) Greenwald et al.'s 4-test method has a more limited weakness when predictor means deviate from rational zero in the positive direction. Meta-analyses showed that aggregate analyses of Greenwald et al.'s five experiments confirmed a multiplicative theory regardless of which analysis method was used. However, only the 4-test method could confirm a pure multiplicative theory.

## Why So Little Faith? A Reply to Blanton and Jaccard's (2006)

### Skeptical View of Testing Pure Multiplicative Theories

Greenwald, Banaji, Rudman, Farnham, Nosek, and Mellott (2002) proposed and tested a unified account of four central constructs of social cognition research — attitude, stereotype, self-esteem, and self-concept. They used a strategy that simultaneously tested both their theory of triadic relationships among the four constructs and their assumption that measures used in testing the theory had rational zero values.<sup>1</sup>

The central hypothesis of Greenwald et al.'s (2002) theory was that interrelations among triads of the four social-cognitive constructs could be described by a *pure multiplicative* model. A pure multiplicative model asserts that the multiplicative product of two variables is the sole predictor of some effect. Because of the well known association between interpretation of multiplicative products and the scale properties of the numbers entering the multiplication, any test of this theory requires measures that have rational zero values. If the measures do not have rational zero values then theory-disconfirming results might indicate only that the measures lacked the needed rational-zero property.

The results reported by Greenwald et al. (2002) largely conformed to the expectations of their pure multiplicative theory, which meant that their results supported both their theoretical conception and their measurement assumption of rational zero values. In commenting on the Greenwald et al. article, Blanton and Jaccard (2006) made three points: (a) the theory confirmations reported by Greenwald et al. might have been spurious because Greenwald et al.'s method is prone to false confirmations in the presence of pure additive (rather than pure multiplicative) models, (b) Greenwald et al.'s

method of testing a pure multiplicative theory (described under the next heading) was inappropriate because of the inappropriateness of assuming that their measures had rational zero values, and (c) Greenwald et al. should have used the SMR analysis method for testing multiplicative theories, even though that method does not distinguish pure multiplicative theories from theories that specify additional effects.

In this reply, we (a) summarize Greenwald et al.'s (2002) method, (b) evaluate with simulations Blanton and Jaccard's (2006) concern about that method's susceptibility to false positive results (demonstrating that there is some susceptibility), (c) examine whether Greenwald et al.'s several studies had properties making them susceptible to such false positives (finding that for the most part they did not), (d) examine what Blanton and Jaccard's recommended SMR method reveals for Greenwald et al.'s data (showing that it also confirms the presence of multiplicative models), and (e) demonstrate that the SMR method has a previously unrecognized severe loss of power in detecting multiplicative effects when component factors of predictor variables have values displaced from zero on rational-zero measurement scales.

### The Four-Test Regression Method

A familiar example of a pure multiplicative model is a rectangle's area being a function purely of the multiplicative product of its length and width. Area can be predicted perfectly from the multiplicative product of length and width, with no need to use length and width as supplementary individual predictors. However, this perfect prediction will occur if (and only if) length and width are measured on scales with rational zero values, which we define more fully in the next section.

Greenwald et al. (2002) described a sequence of four tests that are based on the

example just given of a rectangle's area. Each of the four tests provides an opportunity for a pure multiplicative model to fail. We repeat here Greenwald et al.'s description (pp. 9–11) of their *4-test method*, which was built on the two-step hierarchical regression analysis represented by Equations 1 and 2, below. The present description of the 4-test method changes the symbols originally used to identify criterion and predictor variables, to correspond to Blanton and Jaccard's (2006) notation. In Equations 1 and 2,  $Y$  represents a criterion measure,  $XZ$  represents the multiplicative product that is assumed to be a sole predictor (i.e., a pure multiplicative model), and  $X$  and  $Z$  are measures of the variables that comprise the multiplicative predictor's component factors.

$$Y = b_0 + b_1(XZ) + e \quad (1)$$

$$Y = b_0 + b_1(XZ) + b_2(X) + b_3(Z) + e \quad (2)$$

Test 1. The  $R$  in Step 1 of the regression (Equation 1) should account for substantial variance in the criterion, and Step 1 should estimate a numerically positive value for  $b_1$

Test 2. The estimate of  $b_1$  should also be positive in Step 2 (Equation 2)<sup>2</sup>

Test 3. The increment in  $R$  on Step 2 should not be statistically significant

Test 4. Neither  $b_2$  nor  $b_3$  should differ significantly from zero (positively or negatively) in Step 2

The conclusion that passing all of Tests 1–4 provides support for a pure multiplicative model depends on a measurement assumption — that numeric zero values for measures  $X$  and  $Z$  correspond to rational zero values of the variables that they measure. Failure of this scaling assumption, even in the presence of a pure multiplicative relationship, might produce a significant increment in  $R$  at Step 2, along

with significant deviations of  $b_2$  and/or  $b_3$  from zero (see Aiken & West, 1991, Appendix A).

As previously noted, Blanton and Jaccard (2006) regarded Greenwald et al. (2002)'s 4-test method as improper because (as Blanton & Jaccard asserted) psychological measures could not confidently be assumed to have rational zero values (msp. 6: "Psychologists typically are reluctant to make such strong assumptions, given the arbitrary metrics that they so often rely upon"). As an alternative, they proposed use of the standard simultaneous multiple regression method (*SMR method*, hereafter) for testing multiplicative models. The SMR method uses the test of significance of the multiplicative product term in a three-term SMR analysis that includes the multiplicative predictor along with the two variables that comprise its component factors. This test is well established in statistics texts (e.g., Cohen, Cohen, Aiken & West, 2003). However, as previously noted, this test cannot distinguish a pure multiplicative theory from any other multiplicative theory.<sup>3</sup>

#### Justifying the Assumption of Rational Zero Values for Psychological Difference Measures

Greenwald et al. (2002) and Blanton and Jaccard (2006) agreed that measures with rational zero values are needed to test a pure multiplicative model using the 4-test method. They disagreed about when it is appropriate to assume that measures have rational zero values.

The phrase "rational zero" may be subject to multiple interpretations. We therefore are explicit about our definition: *For a measure described as having a rational zero value, the value of zero should indicate absence of the quantity being measured.* For example, zero length, zero weight, zero calories, zero correlation, and zero words per

minute as typing speed are interpretable as absence of length, weight, calories, correlation, and words typed in a minute. At the same time, it is much less obvious what should be meant by zero intelligence, zero achievement, or zero enthusiasm — these dimensions may lack the possibility of having rational zero values.

The context of considering multiplicative models allows an extension of the definition of rational zero: When it is appropriate to contemplate multiplication of a quantity measured on a scale having a rational zero value, the zero result of multiplication by a zero value on that scale should likewise be validly interpretable as absence of the new quantity represented by the multiplication. As an example, a zero value for a rectangle's width, measured on a rational-zero scale, implies a zero value of the product of multiplication with a length measure, regardless of whether length is measured on a scale with rational zero value.

For the Implicit Association Test (IAT) measures that were used in Greenwald et al.'s (2002) research, zero values were taken to indicate absence of difference in strengths of the sets of associations measured by the IAT's two combined tasks. In the context of Greenwald et al.'s theory, the zero product of multiplying two IAT measures (both assumed to have rational zero values) is interpretable as absence of consistency pressure on the third IAT measure in a triad of interrelated measures.

Blanton and Jaccard (2006) took the very cautious position that one should be reluctant to assume that any psychological difference measures have rational zero values. They illustrated their concern with an example of differences in ratings of height for a group of men versus a group of women. Indeed their example did not warrant the assumption of a rational zero value (Blanton and Jaccard, *msp.* 7). However, it does not follow that rational values of psychological difference measures are generally

unwarranted. For example, in a signal detection paradigm, zero difference between a perceiver's hit and false alarm rates has a rational-zero interpretation as absence of sensitivity to the signal source. Similarly, differences in latencies in time to identify presence of a target amid varying numbers of distracters can be assumed to measure search time required per distracter on a scale with a rational zero value. It is possible to generate many examples of psychological difference measures for which the assumption of rational zero values is plausible.<sup>4</sup>

### Simulations to Evaluate the 4-Test and SMR Methods

#### *Method*

All simulations reported in this article generated data from two alternative theoretical models and compared the statistical analyses of resulting data sets when analyzed by the SMR method and the 4-test method. One theoretical model for generating data sets was always the pure multiplicative model of Equation 3. The comparison theoretical model was usually one containing only an additive predictor (shown in Equation 4) — the model identified by Blanton and Jaccard (2006) as one for which the 4-test method was susceptible to false positive results.<sup>5</sup>

$$Y = a + 1.0 XZ + e \quad (3)$$

$$Y = a + 1.0 (X+Z) + e \quad (4)$$

The coefficients of 1.0 in Equations 3 and 4 are arbitrary and inconsequential to simulation outcomes. Simulations were carried out by a computer program<sup>6</sup> that controlled several parameters of interest. The program's modifiable parameters included ones known to affect covariances of variables with their sums and products (Goodman, 1960) — viz., means and standard deviations of the predictor variables (X and Z) and intercorrelation of the predictors ( $r_{XZ}$ ).



Simulations generated data for samples of  $N = 100$  cases, with estimates of Type I error rate and power being based on 3,000 iterations of a simulation for given parameters. Taking the null hypothesis to be absence of a pure multiplicative model, a Type I error was counted whenever a simulation that was not based on a multiplicative model (e.g., one generated from the additive model of Equation 4) confirmed a multiplicative model by the SMR or 4-test method. Similarly, power was indicated by the observed proportion of multiplicative-model confirmations when the simulated data were generated by the pure multiplicative model of Equation 3.

Simulations were conducted in pairs that differed only in whether data were generated from the pure multiplicative model or the pure additive model. These paired simulations differed only in the regression coefficients that generated the two types of model. Results from each simulation were tested statistically in two ways. The SMR method used the procedure advocated by Blanton and Jaccard (2006); with this method, the test that confirms a multiplicative model is the test for significance of the XZ product term in the three-term SMR analysis that predicts Y from X, Z, and XZ. The 4-test method used the procedure advocated by Greenwald et al. (2002); this test indicates confirmation of the pure multiplicative model when all four tests are passed. All tests involving possible null hypothesis rejections used a two-tailed  $\alpha = .05$  criterion. The proportion of confirmations in 3,000 runs of a pure multiplicative model was taken as the estimate of each analysis method's power. Likewise, the proportion of (spurious) multiplicative-model confirmations in 3,000 runs of a non-multiplicative (e.g., pure additive) model was taken as the estimate of each method's Type I error rate.

Because the SMR and 4-test methods generally differed in both power and Type I error rate, it was insufficient to compare them on these criteria singly. The sensitivity

( $d'$ ) index of signal detection analysis provides a well established metric for comparing the abilities of decision procedures to distinguish signal (pure multiplicative model in this case) from noise (pure additive model). The  $d'$  measure was computed from the combination of hit rate (power) and false alarm rate (Type I error rate) for each analysis method. This use of  $d'$  to measure sensitivity of test methods entails no assumptions about the relative costs of a Type I error (confirmation of a pure multiplicative model when the data were not generated by one) and a Type II error (failure to detect a pure multiplicative model that generated the data). If these two types of error do have differential costs, the criterion for confirming a multiplicative model can be adjusted for either method. For the SMR method, for example, this would be done by making  $\alpha$  smaller (than .05) to reduce both Type I errors and power (i.e., increasing Type II errors). Raising  $\alpha$  to a value higher than .05 would increase both Type I errors and power. For the 4-test method, such adjustments are possible for each of the four tests.

### *Preliminary Test*

Blanton and Jaccard (2006) reported a simulation that used a single data set produced by a pure additive model as an existence proof that the 4-test method could produce false positives. The first test using our simulation method was a large-scale replication of their simulation, using the same parameters used in their simulation. We generated 3,000 simulated experiments from each form of model (multiplicative and additive) with  $N=100$  cases in each simulated experiment using the means, SDs,  $r_{XZ}$ , and Multiple  $R$  (= .40) as described by Blanton and Jaccard for their simulation. We found for the 4-test method that its Type I error rate was .372, compared to .051 for the SMR method. This appeared to be a poor result for the 4-test method. However, the power of the 4-test method was .886, compared to only .471 for the SMR method — a

poor result for the SMR method. Using signal detection analysis, the sensitivities of the two methods were nearly identical (4-test method:  $d' = 1.53$ ; SMR  $d' = 1.56$ ).

Blanton and Jaccard (2006) described their simulation in terms of variables contained in one of the five data sets (Banaji, Greenwald, & Rosier, 1997) that had been analyzed by Greenwald et al. (2002). However, their simulation did not use estimates of simulation parameters that were available from that data set. Our second preliminary test was a repeat of the first, replacing Blanton and Jaccard's choices of means, SDs, and  $r_{XZ}$  with the estimates that were available from the Banaji et al. (1997) data set. The results were strikingly different. For the 4-test method, power = .896 and Type I error rate = .002 ( $d' = 4.14$ ), while for the SMR method, power = .966 and Type I error rate = .049 ( $d' = 3.48$ ). Especially noticeable was that the Type I error rate of the 4-test method had dropped to an extremely low value, while the power of the SMR method had risen to an extremely high value.

The wide fluctuation in power of the SMR method was entirely unexpected. Also, the great increase in sensitivity from the first to second simulation for both methods was quite surprising given that the value of Multiple  $R$  (= .40) was identical for the two simulations. These observations led us to pursue a set of further simulation studies, seeking to shed light on properties of both analysis methods.

#### *Further Comparisons of the 4-Test and SMR Methods*

Our main sets of simulations focused on effects of variation in means of the X and Z predictors on power and Type I error rate for tests of pure multiplicative theories compared to pure additive theories. This focus on predictor means was prompted partly by observing that the most obvious difference between the two preliminary tests was in the means of their predictors. Also, it was prompted by known relations among means

of variables, variances of their multiplicative products, and covariances of the variables with their multiplicative products (Bohrstedt & Goldberger, 1969; Goodman, 1960).

Values of the X and Z predictors were generated so as to have normal distributions with SD = 1.0. Values of the mean of each of the X and Z variables were varied orthogonally from -2.0 to 2.0 in steps of 0.5. The simulations generated latent values of X and Z, with correlation maintained at  $r_{XZ} = 0$ , and then generated observed values of X and Z so as to maintain a correlation of these observed values of X and Z at  $r = .75$  with their respective latent values. This latent-measured variable correlation corresponds to reliability = .56 (=  $.75^2$ ), which approximates the average observed test-retest reliability of IAT measures (Nosek, Greenwald, & Banaji, in press).

Values of Y (criterion) were generated from latent values of X and Z to simulate pure multiplicative (Equation 3) or pure additive (Equation 4) models. The latent value of Y was either the product of latent values of X and Z, or their sum. The method of adding error to Y was varied in different sets of simulations. In the first set of simulations to be reported, error variance in Y was set at the level needed to achieve a fixed correlation of latent Y with observed Y (at  $r = .6$ ) regardless of means of X and Z.

The results are shown in Figure 1, revealing that the power of the 4-test method was consistently high, in the vicinity of 85% (see Figure 1D).<sup>7</sup> By contrast, power of the SMR method varied considerably with variations in the X and Z means (see Figure 1A), such that increasing displacement from zero of means for X and/or Z was associated with reduced power. This loss of power can be understood by recognizing that the SMR test is the test of significance of a partial correlation and by considering the factors that affect the magnitude of this partial correlation.<sup>8</sup>

Type I error rates (Figure 1B and 1E) revealed a very different pattern. For the SMR

method the Type I error rate was consistently very close to the expected 5% rate for a two-tailed test of significance with  $\alpha = .05$ . For the 4-test method, Type 1 error rates varied considerably, being mostly well below 5%, but increasing to much higher levels when predictor means were positively polarized — to a maximum of 45% when means of both X and Z were at the maximum positive values (2.0 SDs) used in the simulations. Unlike the SMR method, Type I error rates for the 4-test method are not limited at the low end by the 5% alpha level of the two-tailed tests used in the simulation — they can be as low as zero.<sup>9</sup>

As explained previously, the information contained jointly in power and Type I error rates is captured by the signal detection sensitivity measure,  $d'$ . For all 81 pairs of simulations, the sensitivity of the 4-test method exceeded that of the SMR method. For the SMR method, mean  $d' = 1.55$ , compared to mean  $d' = 3.73$  for the 4-test method. Comparison of sensitivities is shown in Figure 2A, in the form of difference between  $d'$  values. The average difference in  $d'$  was 2.18,  $t(80) = 17.93$ ,  $p = 10^{-29}$ .

For the simulations shown in Figure 1, error was added to the criterion Y by maintaining a fixed correlation of  $r = .60$  between latent Y and observed Y. For the pure multiplicative model, variance of latent Y increases as means of X and Z are increasingly displaced from zero. Correspondingly, with this simulation method the simulated error component of Y increases with increasing displacements from zero of the means of X and Z. This increase in the error component of Y contributes to adverse effects on power of the SMR method while having little effect on power of the 4-test method. An alternative was to keep the error component of Y fixed through all simulations. Use of this alternative strategy was the only change made for the second major set of simulations. The error variance of Y was set at a value that produced a

correlation of  $r = .6$  between latent Y and observed Y when means of X and Z were zero. Error variance of Y was retained at that same value through all variations in means of X and Z. With this change (see Figure 2B), the SMR method achieved sensitivity superior to the 4-test method for 14 of the 81 paired-simulation comparisons. Still, the difference in sensitivity strongly favored the 4-test method, with average difference in  $d'$  of 1.28,  $t(80) = 11.15$ ,  $p = 10^{-17}$ .

The change to a fixed error for Y attenuated the power losses of the SMR method for multiplicative models when means of X and Z were displaced from zero, while having little effect on performance statistics for the 4-test method. There is no obvious way to conclude that one of the two methods of simulating the error component of Y is superior. Each has a rationale and each has drawbacks. Consider first the method of fixed correlation of latent Y with observed Y. Actual data frequently show such a correlation between means and variances, suggesting the reasonableness of this method. Nevertheless, this method allows the error component of Y to rise to rather large values when X and Z means are substantially displaced from zero. In the context of the multiplicative model, the alternate strategy of keeping the error in Y fixed allows the correlation between latent and measured Y to rise to perhaps unrealistically high values when means of X and Z are displaced from zero (e.g., the average latent–observed correlation for Y was  $r = .92$  when means of X and Z were  $\pm 2.0$  SD, compared to average latent–observed correlation for Y of  $r = .60$  when means of X and Z were 0.0). If it is reasonable to assume that criterion error variance in actual data of multiplicative models is intermediate between these two simulation strategies, then Figures 2A and 2B estimate upper and lower bounds, respectively, of sensitivity differences between the 4-test and SMR methods.

Additional simulations examined effects of (a) varying the intercorrelation of X and Z ( $r_{XZ}$  varied from  $-.40$  to  $.40$ ), (b) varying the Multiple  $R$  of the three-term regression (varied from  $.3$  to  $.7$ ), (c) varying the value of  $t$  used as criterion for passing the 4-test method's Test 2, and (d) replacing the pure additive noise model with a null model in which all predictors (X, Z, and XZ) had zero coefficients. Except for the last, all of these additional simulations were run with varied displacement from zero of means of X and Z, to determine whether variations in the other parameters might moderate the already-observed strong effects of polarization of means.

Variations of  $r_{XZ}$  had complex effects on power of the SMR method and Type I error rate of the 4-test method when means of X and Z were both non-zero. Power of the SMR method decreased with increasing magnitude of  $r_{XZ}$  when all three of  $r_{XZ}$  and means of X and Z were non-zero, with either all positive in sign or two of the three having a negative sign. In contrast, power of the SMR method increased with increasing magnitude of  $r_{XZ}$  when all three of  $r_{XZ}$  and means of X and Z were non-zero, with either all negative in sign or just one of the three having a negative sign. For the 4-test method and limited to the situation of positive means of both X and Z, negative  $r_{XZ}$  correlations reduced Type I error rate and positive  $r_{XZ}$  correlations increased Type I error rate. These effects were smaller in magnitude than the effects of polarized means shown in Figure 1.

Increased Multiple  $R$  increased power for both methods but did so more for the SMR method, which is not restricted by the 4-test method's upper limit of about 90% power.<sup>10</sup> For the 4-test method, increases in Multiple  $R$  also decreased the Type I error rate when predictor means were positively polarized. The net effect of increased Multiple  $R$  was a generally increased sensitivity for the SMR method and increased sensitivity for

the 4-test method when true means of predictors were positively polarized. Even at the highest value of Multiple  $R$  used in these simulations, power and sensitivity of the SMR method were substantially impaired with polarized means of predictors, whereas Type I error and sensitivity of the 4-test method were impaired selectively for positively polarized predictor means.

In Greenwald et al. (2002), Test 2 was passed if the regression coefficient for XZ in the regression that also included X and Z as simultaneous predictors was numerically positive (i.e., the criterion was  $t > 0$ ). Simulations varied the criterion for passing Test 2 by requiring increased positive values of  $t$ . Increased value of  $t$  used as criterion for passing Test 2 of the 4-test method had expected effects of simultaneously decreasing power and Type I error rate of the 4-test method, with virtually no effect on sensitivity.

Simulations using a null model produced an average Type I error rate for the 4-test method of .021. This Type I error rate is below .05 chiefly because Test 1 of the 4-test requires a significant positive result, which occurs with probability of .025 for the null model. Combining this Type I error rate with the average power of .825 for the two methods of simulating error in Y yields an average sensitivity of  $d' = 2.96$  for the 4-test method. In comparison, the average power for the SMR method was .610, which coupled with the theoretically expected Type I error rate of .05, yields an average sensitivity of 1.92, considerably below that of the 4-test method.<sup>11</sup>

In summary of observations from these simulations, the SMR method was consistent in Type I error rate across all combinations of parameters, but suffered very substantially in lost power and sensitivity to the extent that true values of predictor means were displaced in either direction from zero. The 4-test method was consistently high in power, but suffered an increase in Type I error rate and consequent reduced



sensitivity when predictor means were positively polarized. The 4-test method showed reduced Type I error rate and therefore had much higher sensitivity than the SMR method when predictor means were negatively polarized (see Figure 1F).

Although this summary suggests that the 4-test method is superior to the SMR method in sensitivity to presence of a pure multiplicative model under a wide variety of conditions, the two methods approached similar sensitivity — both showing impaired sensitivity — when predictor means were positively polarized (see Figure 2).

To examine properties of the two methods under a sampling of conditions that arise in research, we used the 16 data sets from the five experiments that were analyzed by Greenwald et al. (2002) as natural sources of simulation parameters. The parameters taken from the 16 analyses were (a) means and standard deviations of predictors, (b) intercorrelation of predictors ( $r_{XZ}$ ), and (c) Multiple  $R$  of the three-term regression model (see footnote 10).

The first set of these simulations used  $N=100$  cases for each of 1,000 simulated experiments for each of the 16 data sets, for each of a pure multiplicative and a pure additive generating model. The sample size of  $N=100$  cases was approximately the size of the largest of the sample sizes in Greenwald et al. (2002)'s five experiments. The resulting data can be summarized by the averages, across the 16 pairs of simulations, of power, Type I error, and sensitivity for the two analysis methods. For the SMR method, average power = .713, average Type I error = .052, and average sensitivity ( $d'$ ) = 2.45. For the 4-test method, average power = .872, average Type I error = .106, and average sensitivity ( $d'$ ) = 3.12. A paired-sample  $t$  test on the obtained 16 pairs of sensitivity values showed superiority of the 4-test method,  $t(15) = 2.89$ ,  $p = .01$ .

An additional set of simulations was done identically except for using  $N=60$ , slightly smaller than the smallest of Greenwald et al. (2002)'s sample sizes. The corresponding results for power, Type I error, and sensitivity for were .551, .053, and 1.82 for the SMR method, and .806, .136, and 2.53 for the 4-test method. The  $t$  test for comparison of sensitivities again yielded a result showing superiority of the 4-test method, with numerically the same value of  $t$  as for the test with larger sample sizes  $t(15) = 2.89$ ,  $p = .01$ .<sup>12</sup>

#### *Sensitivity of the 4-Test Method to Rational Zero Values*

Greenwald et al. (2002) observed that the 4-test method could fail when measures of the multiplicative predictor's component variables lacked rational zero values. The aim of a final set of simulations was to assess the 4-test method's sensitivity to deviation from these rational zero values, using a minor variation on the 81 pairs of simulations for which results were presented in Figure 1. After generating values of  $X$  and  $Z$  with means that varied for each of  $X$  and  $Z$  from  $-2.0$  to  $2.0$  in SD units and computing the  $XZ$  product, the true means of  $X$  and  $Z$  were subtracted from values of the respective variables for each simulated case. This effectively centered the predictors while destroying rational zero values<sup>13</sup>. As expected, the 4-test method deteriorated in power, and did so when means were displaced from rational zero values by even half a standard deviation. When measures had accurately located zero values, power in these simulations was 89%. When the zero value of each measure was misplaced by only half a standard deviation, power dropped to 29%. Type I error rates of the 4-test method were consistently near zero in these simulations. These simulations confirmed the appropriateness of regarding the 4-test method as useful only when zero values of predictor measures are quite close to rational zero values of their

underlying latent variables. These simulations results further confirmed Greenwald et al.'s (2002) expectation that the 4-test method would serve simultaneously to test their theoretical prediction of a pure multiplicative model and their measurement assumption that their measures had rational zero values.

### *Summary of Simulation Findings*

The present simulations demonstrated surprisingly low power for the SMR test of a product term when X and Z variables had means that were polarized relative to meaningful zero values. Under conditions in which the SMR method thus suffers sharply reduced power, Greenwald et al. (2002)'s 4-test method was often considerably more sensitive to presence of a pure multiplicative model (see Figures 1A, 1D, and Figure 2). The 4-test method has the added advantage of being able to confirm pure multiplicative models, something not possible for the SMR method. However, the 4-test method also had problems of reduced sensitivity, which were limited to the situation in which means of X and Z were positively polarized (see Figure 1E and Figure 2).

### *Meta-Analysis of Greenwald et al. (2002)'s Findings*

Blanton and Jaccard (2006) suggested that Greenwald et al. (2002)'s tests of pure multiplicative models were untrustworthy because the 4-test method was prone to false positives. It is indeed true that the 4-test method is prone to false positives when predictor means are positively polarized and the true model is additive (see Figure 1E), although it is not prone to false positives when the true model is a null model. Blanton and Jaccard (2006) did not provide evidence to justify their implied further conclusion that Greenwald et al.'s. (2002) published results should be treated as likely false positives resulting from true additive models.

To evaluate what should be concluded from the results reported by Greenwald et al.

(2002) we conducted a meta-analysis of the 16 analyses they reported. The 16 data sets were first analyzed using the SMR method. The meta-analysis (see Appendix for details) revealed that (a) the SMR method confirmed a multiplicative model in 9 of the 16 data sets, (b) in aggregate analyses of Greenwald et al.'s five studies, the SMR test for presence of a multiplicative model was statistically significant in two of three aggregate analyses, and (c) there was limited evidence for heterogeneity among the five studies, with one of the five studies (Mellott & Greenwald, 2000) appearing to be discrepant from the others.

As previously described, simulations using as input parameters the means and standard deviations of each data set's X and Z variables and its observed values of  $r_{XZ}$  and Multiple R were used to obtain estimates of power, Type I error rate, and sensitivity of both the 4-test and SMR methods for these 16 data sets. Only three of the 16 data sets (all from the study identified as FG in Table A1) were found to have parameters that made them prone to spurious confirmation of a pure multiplicative model when the generating model was pure additive rather than pure multiplicative.

Although the SMR method confirmed the multiplicative relationship expected from Greenwald et al. (2002)'s theory in nine of the 16 data sets, the added value of the 4-test method was its indication that most of the observed multiplicative effects were plausibly interpreted as pure multiplicative effects and, simultaneously, that the measures used in these tests were plausibly interpreted as having rational zero values.

Blanton and Jaccard (2006) raised the question of whether regression product-term effects in Greenwald et al. (2002)'s studies were *bilinear* in shape. This is indeed an expectation of the pure multiplicative theory, as was noted by Blanton and Jaccard (2006). Blanton and Jaccard further reported an analysis to test whether the Rudman,

Greenwald, and McGhee (2001) data (one of the five experiments analyzed by Greenwald et al.) conformed to this expected bilinear shape of the product-term effect. They used a method described by Kenny and Judd (1984), which requires fitting regression models of the form of Equation 5.

$$Y = a + b_1X + b_2Z + b_3Z^2 + b_4XZ + b_5XZ^2 \quad (5)$$

Significance of the  $b_5$  effect indicates a deviation of the multiplicative effect from bilinearity. Blanton and Jaccard (2006) reported that “when the above equation was tested for the Rudman et al. [data],  $b_5$  was statistically significant” (msp. 30). Because there are three variables that can be used as criterion ( $Y$ ) and, for each of these three, the remaining two variables can be switched between the roles of  $X$  and  $Z$ , it is actually possible to conduct six tests of the form shown in Equation 5 for the Rudman et al. data. In conducting all six of these tests, we found a significant  $b_5$  coefficient ( $p = .0378$ ) for just one — for the other five, the 2-tailed  $p$  values were .39, .50, .80, .81, and .97. By contrast, the linear product term was statistically significant in the appropriate tests corresponding to all six of these analyses — these are the tests shown in the “Test 2  $p$ ” column of Table A1 for the three data sets identified there as RGM. In summary, and contrary to Blanton and Jaccard’s conclusion, the Kenny and Judd (1984) method that they advocated clearly indicated lack of complication due to quadratic contributions.

### Summary and Conclusion

Blanton and Jaccard (2006) criticized the 4-test method that Greenwald et al. (2002) used as support for the conclusion of a pure multiplicative model underlying balance-like relations among implicit attitudes, stereotypes, self-concepts, and self-esteem. This article has examined both Greenwald et al.’s conclusions and their methods in light of Blanton and Jaccard’s critique. We re-examined all of Greenwald et al.’s data sets

using the simultaneous multiple regression (SMR) method preferred by Blanton and Jaccard. That method provided very substantial support for presence of a multiplicative model underlying the triadic interrelationships of variables in the majority of Greenwald et al.'s data sets. However, the SMR method has the limitation of not being able to test the pure multiplicative model that was proposed in Greenwald et al.'s theory.

In further examining Blanton and Jaccard's (2006) critique, we confirmed that Greenwald et al.'s (2002) 4-test method had a susceptibility to spurious conclusions as was suggested by Blanton and Jaccard. Our various simulation studies showed (a) that this susceptibility was much more delimited than implied by Blanton and Jaccard, and (b) was a plausible complicating factor in only one of the five experiments analyzed by Greenwald et al. An unanticipated result of our simulation studies was to show that Blanton and Jaccard's preferred SMR method has a previously unrecognized weakness in the form of sharply reduced power in testing pure multiplicative models when predictor means are displaced from rational zero values.

Blanton and Jaccard (2006) went so far as to conclude by proposing that physicists could effectively use the SMR method to test pure multiplicative models. "Had Einstein or Newton been stuck with the types of measures that psychologists typically work with, and if their experimental designs were such that they had to work with correlational data and regression models, then they would indeed have needed to include component parts of the product term in their empirical tests of their models" (msp. 21). Our simulations show that this conclusion was mistaken. With fallible measures, the SMR method suffers severe power loss in tests such as the illustration they chose — predicting rectangle area from the product of length and width. Physicists would indeed be unwise to use that method, and psychologists should likewise not judge it wise to

follow Blanton and Jaccard's suggestion to rely on the SMR method to test pure multiplicative models.

In advocating the SMR method for detecting multiplicative effects of predictor variables, Blanton and Jaccard (2006) abandoned two significant possibilities: (a) identifying pure multiplicative models and (b) assessing whether measures used in tests of multiplicative models have rational zero values. We believe that Blanton and Jaccard might reasonably have had greater faith in the possibility of psychological measures having rational zero values, might have had greater faith in the strategy of simultaneously testing theoretical and measurement assumptions, and might have recognized that the 4-test method has real advantages over the SMR method in testing pure multiplicative theories.

## References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, *68*, 181–198.
- 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.
- Biernat, M., Manis, M. & Nelson, T. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, *60*, 485–499.
- 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*, in press.
- Bohrstedt, G. W., & Goldberger, A. S. (1969). On the exact covariance of products of random variables. *Journal of the American Statistical Association*, *64*, 1439–1442.
- Bussemeyer, J. R., & Jones, L. E. (1983). Analysis of multiplicative combination rules when the causal variables are measured with error. *Psychological Bulletin*, *93*, 549–562.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Curd, M., & Cover, J. A. (1998). *Philosophy of Science: The Central Issues*. New York: Norton.
- Duhem, P. M. M. (1954). *The aim and structure of physical theory*. (trans. by P. P. Wiener). Princeton, NJ: Princeton University. (Original French publication, 1906)



- Farnham, S. D., & Greenwald, A. G. (1999, June). In-group favoritism = implicit self-esteem  $\times$  in-group identification. Paper presented at the 11th annual meetings of the American Psychological Society, Denver, CO.
- Goodman, L. A. (1960). On the exact variance of products. *Journal of the American Statistical Association*, *55*, 708–713.
- Greenwald, A.G., Banaji, M., Rudnam, L., Farnham, S., Nosek, B.A. & Mellott, D. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie*, *48*, 85-93.
- 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.
- Greenwald, A. G, Nosek, B. A., & Sriram, N. (2005). *Consequential validity of the Implicit Association Test: Comment on the article by Blanton and Jaccard*. Manuscript submitted for publication.
- Kenny, D. A. (1999). *Meta-analysis: Easy to Answer*. Unpublished manuscript, University of Connecticut.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, *96*, 201–210.
- Lakatos, I. (1970). *Criticism and the growth of knowledge*. New York: Cambridge University Press.
- Popper, K. R. (1959). *The logic of scientific discovery*. New York: Basic Books.
- Mackie, J. L. (1967). Fallacies. In Edwards, P. (Ed.), *The encyclopedia of philosophy*

(Vol. 3, Pp. 169–179). New York: Macmillan.

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376–390.

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.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior*. Psychology Press.

Quine, W. V. (1953). *From a logical point of view*. Cambridge, MA: Harvard University.

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.

## Appendix

### Meta-Analysis of the Sixteen Analyses in Greenwald et al. (2002)'s Five Studies

Here we report details of analyses done to determine what conclusions can be justified by considering as a group the 16 analyses (from five experiments) that were reported in the Greenwald et al. (2002) article. Tests based on SMR analyses use standard meta-analytic procedures, as described below. Those based on the 4-test method are in a box-score format, there not being available techniques for meta-analysis of the 4-test method. Table A1 collects results of all tests for the 16 analyses included in Greenwald et al.'s (2002) article.

We consider first the results from the SMR method, which uses the same statistic as Test 2 of the 4-test method. Whereas the 4-test method requires only a numerically positive result of this test, the SMR method requires a statistically significant result to confirm a multiplicative model. To assure independence of the effects that were meta-analytically combined in aggregated SMR results, we conducted three separate meta-analyses, each including one partial correlation from each of the five studies. The tests for significance of aggregate effect sizes were  $t$  tests against the null hypothesis of zero, with each of the five samples weighted by the square root of its sample size. The tests for homogeneity were Fisher tests for the homogeneity of correlations (cf. Kenny, 1999). These meta-analyses used, respectively, data in rows of Table A1 that had the same type of criterion measure — either a self–group association, a self–attribute association, or a group–attribute association.<sup>14</sup>

The average product term coefficient for the SMR method (the  $pr$  for Test 2 in Table A1) was statistically significant in meta-analyses for self–group associations as criteria (SG in Table A1; average  $pr = .20$ ,  $t(4) = 2.86$ ,  $p = .05$ ) and for self–attribute

associations (SA; average  $pr = .23$ ,  $t(4) = 3.61$ ,  $p = .02$ ). For group–attribute associations (GA), the average effect size was almost as large (average  $pr = .18$ ), but was not statistically significant,  $t(4) = 2.00$ ,  $p = .12$ . These results are consistent with the conclusion that the SMR method demonstrated presence of a multiplicative predictor in the majority of the 16 data sets.

In tests of the hypothesis that effect sizes were homogeneous, tests for criterion measures of types SG and SA met criteria of homogeneity,  $P^2(4) = 8.56$  ( $p = .07$ ) and  $P^2(4) = 7.35$  ( $p = .12$ ), but not for type GA,  $P^2(4) = 13.13$  ( $p = .01$ ). Because of the low power of these tests — each being based on only five effect sizes — and because of indications (see just below) that results for the Mellott and Greenwald (2000) study deviated from the other four, we are reluctant to accept the conclusion that effect sizes were homogeneous.

We next considered results for the 4-test method. For 10 of the 16 data sets, the 4-test method confirmed a pure multiplicative model by yielding a passing result for all four tests. Fewest tests were passed in the Mellott and Greenwald (2000) study (MG in Table A1), which involved young and old subjects' implicit attitudes toward the concepts *old* and *young*, their implicit identification (association of self) with *old* and *young*, and their implicit self-esteem. Mellott and Greenwald's study produced the surprising observation that elderly subjects tended to associate *self* more with *young* than with *old* on the IAT. This was an unexpected discovery that a group of subjects (elderly) did not implicitly identify with a group to which they nominally belonged. In discussing this study, Greenwald et al. (2002) had noted that its results did not fit well with the multiplicative-model expectations of their theory.

The study that conformed next least well to the expectations of the pure

multiplicative model was Banaji et al.'s (1997) study (BGR in Table A1), which involved Black and White subjects' implicit attitudes toward and implicit identification with their racial groups, in relation to their implicit self-esteem. For this study, the SMR test for all three analyses revealed a significant effect of the multiplicative predictor. Two of the three analyses nevertheless failed at Test 3 of the 4-test method, meaning that — even though a multiplicative predictor was strongly indicated — there were significant increments in explained variance when the component factors were added as predictors to the solitary product term of Test 1. This pattern indicated either that the model was not a pure multiplicative model (i.e., component factors were additionally involved) or that the measurement scales deviated from having rational zero values, or both.

To assess the plausibility of regarding some of the 10 confirmations by the 4-test method in Table A1 as Type I errors, we used the means and standard deviations of predictors and intercorrelations of predictors to generate simulations that appraised both the power of the 4-test method and its proneness to Type I error (i.e., spurious confirmation when the true model was additive, rather than multiplicative). The results for simulated sample sizes of  $N=100$ , shown in Table A1's two rightmost columns, indicated that three of the 16 analyses had high potential for Type I error. These three tests were all from the Farnham and Greenwald (1999) study (FG in Table A1). At the same time, Table A1 shows that one of these three analyses had a statistically significant product term in the SRM test, indicating presence of a multiplicative model. Five of the 16 data sets passed both the four-test and the SMR test for a multiplicative model. For these five data sets, the evidence for a pure multiplicative model was especially strong.

## Figure Captions

1. Summarized results of simulation studies with analyses by the SMR method (A, B, C) and the 4-test method (D, E, F), for 9 values of true means of each of Predictors X and Z, varied orthogonally. Results are averaged over 3,000 iterations of each simulation, with N=100 cases per iteration. Means of predictors are on scales with rational zero values and are in standard deviation units. The method of simulating error in the criterion, Y, was a fixed correlation between its latent and observed values. See text for further description.
2. Summarized results from simulation studies indicating sensitivity differences between the SMR and 4-test methods for the simulations shown in Figure 1 (Panel A). Panel B summarizes results for a parallel set of simulations in which the method of simulating error in the criterion, Y, was change to maintaining a fixed magnitude of error variance through all variations of mean values of the X and Z predictors. See text for further description.

Figure 1

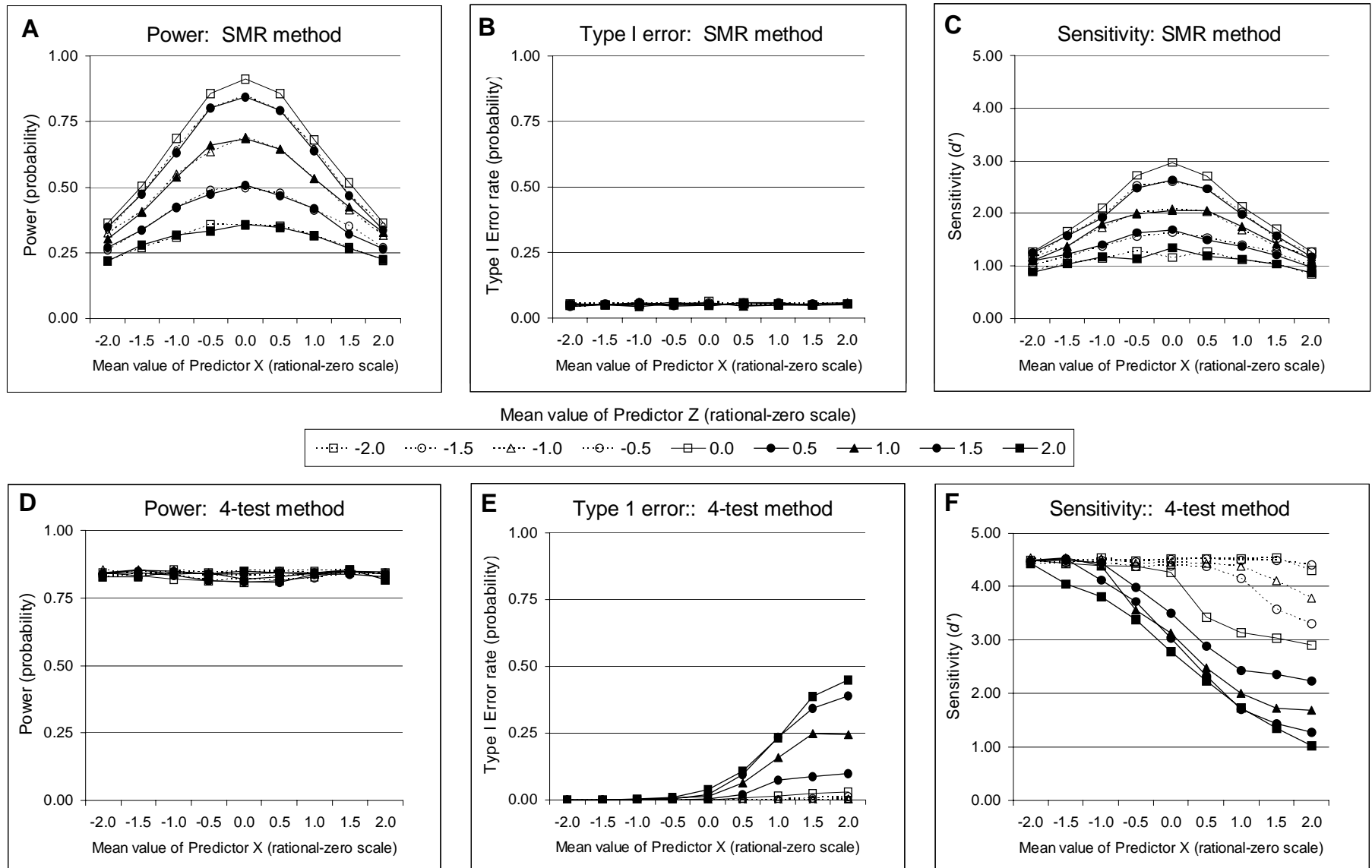


Figure 2

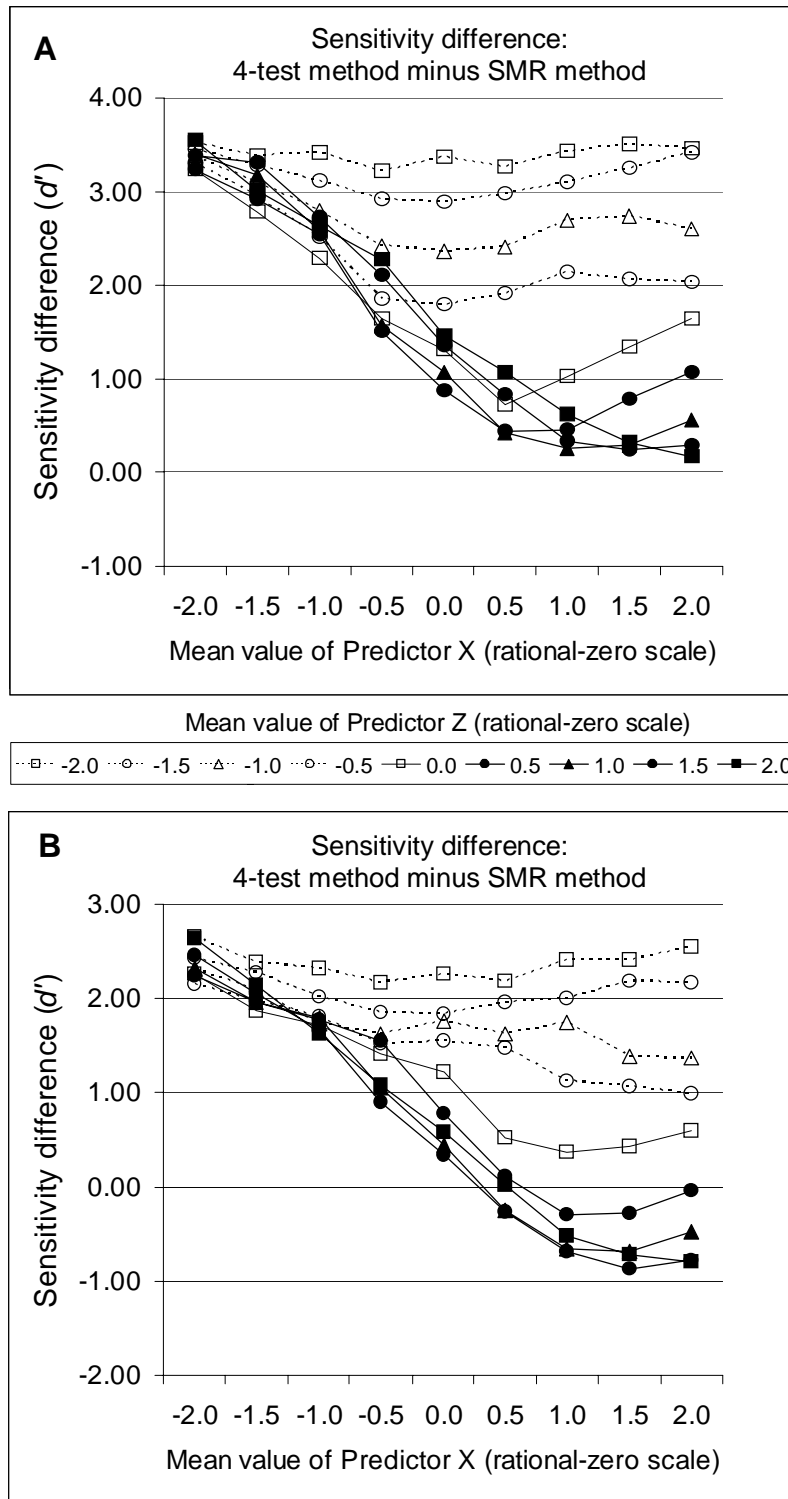




Table A1. Summary of Results for Greenwald et al. (2002)'s Four-Test Procedure Applied to 16 Analyses from Their Five Experiments

Study	Measure type <sup>a</sup>	N	Criterion (Y)			Test 1 beta	Test 2 <i>pr</i>	Test 3 <i>R</i> <sup>2</sup> gain	Test 4a <i>pr</i>	Test 4b <i>pr</i>	No. of tests passed	Simulated <sup>b</sup>	
			measure	<i>d</i>	<i>r</i> <sub>XZ</sub>							Type 1 error	Power
FG	SG	65	self-female	1.39	.340	.472*	.269*	.018	-.146	-.066	4	.471	.854
FG	SA	65	self-esteem	2.00	.380	.428*	.216	.014	-.110	-.017	4	.425	.874
FG	GA	65	female-positive	2.43	.320	.445*	.100	.008	.022	.091	4	.389	.874
BGR	SG	61	self-white	0.25	.020	.578*	.250*	.122*	-.236	.196	3	.009	.908
BGR	SA	61	self-esteem	1.54	.640	.267*	.394*	.093*	-.292*	.042	2	.124	.899
BGR	GA	61	white-positive	0.91	-.060	.700*	.357*	.001	.031	.022	4	.000	.900
MG	SG	98	self-old	-0.80	-.380	.375*	.019	.027	-.175	.103	4	.001	.886
MG	SA	98	self-esteem	1.25	.290	.403*	.031	.059*	-.140	-.258*	2	.000	.897
MG	GA	98	old-positive	-1.21	-.380	.298*	-.112	.092*	.193	-.305*	1	.001	.879
RGM	SG	95	self-female	0.04	-.030	.363*	.390*	.045	-.227*	-.057	3	.108	.899
RGM	SA	95	self-warm	0.43	.110	.296*	.343*	.046	-.219*	-.030	3	.100	.840
RGM	GA	95	female-warm	1.33	.130	.355*	.336*	.001	-.008	-.004	4	.024	.867
NBG	SG	91	self-male	-0.05	-.070	.407*	.094	.023	.129	-.115	4	.017	.903
NBG	SA	91	self-math	-0.06	-.140	.433*	.173	.005	.082	-.008	4	.020	.899
NBG	GA	91	male-math	1.37	.410	.228*	.216*	.015	-.102	-.023	4	.007	.666
NBG	AA	91	math-positive	-1.19	.410	.449*	.302*	.013	-.058	.112	4	.001	.905

Note. These data are from five experiments that were described in Greenwald et al. (2002): FG = Farnham and Greenwald (1999); BGR = Banaji, Greenwald, and Rosier (1997); MG = Mellott & Greenwald (2000); RGM = Rudman, Greenwald, & McGhee (2002); NBG = Nosek, Banaji, & Greenwald (2002). *d* = mean of indicated criterion measure in SD units, *pr* = partial correlation. Tests 1 thru 4 are the respective tests of Greenwald et al. (2002)'s 4-test procedure. Test 4 involves two effects, identified as 4a and 4b.

<sup>a</sup>IAT measures of associations involved three types of concepts: self (S); groups (G — e.g., male, old, Black); and attributes (A — e.g., valence, math). The 2-letter codes indicate the two types of concept included in the criterion IAT measure in each analysis. The two predictor variables for each analysis are the measures indicated as criterion measures in the other two rows. For the last row of the table, the two predictors were NBG's GA and SG measures.

<sup>b</sup>The simulated Type I error and power figures are based on 1,000 repetitions of simulations with samples of N=100 cases.

\**p* < .05.

### Acknowledgment Note

This research was supported by National Institute of Mental Health Grants MH-41328, MH-01533, and MH-57672. The authors thank Robert Abbott, Richard Gonzalez, and David Kenny for helpful suggestions. Jerome Busemeyer and Gary McClelland provided advice on the simulation methods and Betsy Becker provided advice on the meta-analyses reported in the Appendix. These helpful colleagues are of course not responsible for any errors that may remain in the analyses we report. Correspondence concerning this article should be addressed to Anthony G. Greenwald, Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195-1525. E-mail: [agg@u.washington.edu](mailto:agg@u.washington.edu)

## Footnotes

1. This strategy was in accordance with the widely held view (cf. Curd & Cover, 1998) that, necessarily, empirical tests of a theory will simultaneously appraise the theory and any auxiliary measurement assumptions made in testing it (cf. Duhem, 1906/1954; Lakatos, 1970; Popper, 1959; Quine, 1953).
2. The  $b_1$  coefficients the Tests 1 and 2 should be positive if the predictors (X and Z) have been scored so that the sign of their multiplicative product (XZ) is consistent with the theoretical expectation of a positive relationship of that product with the criterion measure, Y. An appropriate basis for deeming that  $b_1$  explains “substantial variance” in Test 1 is a statistically significant result, whereas Test 2 requires only a numerically positive (i.e., greater than zero) estimate of  $b_1$ .
3. As a relevant historical observation, we note that the initially submitted version of the Greenwald et al. (2002) article used the same SMR method advocated by Blanton and Jaccard (2006). However, as a result of the review process, Greenwald et al. were persuaded to use the 4-test method and adopted it. We can report that they remain persuaded of the method’s value and take full responsibility for their published conclusions using that method.
4. Further consideration of when a zero value of a difference measure can be taken as a rational zero value appears in Greenwald, Nosek, and Sriram (2005).
5. The simulations we shall report did not consider ability of either the SMR method or the 4-test method to distinguish models containing both additive and multiplicative predictors from ones with only multiplicative predictors. However, the simulation methods we developed (see footnote 6) can perform such tests.
6. All simulations reported in this article can be reproduced with Excel spreadsheets

that are downloadable at

<http://faculty.washington.edu/agg/MeanXZ.Simulations.Model2&3.13Jul05.zip>. The spreadsheets provide instructions enabling their use both to reproduce this article's simulations and to generate a wide variety of other simulations.

7. Power of the 4-test method is limited to a maximum of approximately 90% because passing each of Tests 3 and 4 involves accepting a null hypothesis that has a Type I error rate of 5%. At least one of these two tests will fail approximately 10% of the time.

8. In most of the present simulations of pure multiplicative theories, each of X and Z is composed of a true-score (or latent) component and an error component, such that the XZ product also has true-score and error components. The true-score component of the XZ product then generates the observed value of the criterion Y, adding an independent error component. With nonzero means of X and Z, the XZ product includes variance shared with X and Z (multicollinearity). With no measurement error in X and Z, the regression analysis effectively partials the variance contributions of X and Z from the XZ product, leaving a residual that fully captures the multiplicative relationship that generates Y. However, measurement error reduces the regression analysis's effectiveness in partialing variance associated with true-score components of X and Z from the true-score component of the XZ product. This reduces the portion of the XZ variance that is associated with Y independently of associations of X and Z with Y, reducing the partial correlation corresponding to the SMR method's test of the multiplicative product term. Increased displacement of the X and Z means from zero exacerbates this problem because of the associated increases in correlations of X and Z with XZ. It is beyond the scope of this article to describe these relationships formally, although such description can follow the lines of previous work by Goodman (1960),

Borhnstedt and Goldberger (1969), Busemeyer and Jones (1983), and McClelland and Judd (1993).

9. An end-point correction is needed to compute signal detection sensitivity measures, either when power reaches 1.0 or Type I error rate drops to 0.0. The correction that was used replaces these extreme values of 0.0 with  $(.25 \div k)$  and 1.0 with  $1.0 - (.25 \div k)$ , where  $k$  is the number of simulations used to estimate the proportion. For the simulations in this article,  $k = 3,000$ , such that 0.0 and 1.0 were replaced with .000083 and .999917, respectively. The logic of this end-point correction is to treat an observed value of  $1 \div k$  as if it were the midpoint of a range between  $0.5 \div k$  and  $1.5 \div k$ . Extending that reasoning to interpret  $0 \div k$  as the observed value for a true-value range extending from  $0 \div k$  to  $0.5 \div k$ , it is replaced by  $0.25 \div k$ . The logic of replacing 1.0 with  $1.0 - (.25 \div k)$  is the same. This end-point correction method was first reported by Banaji and Greenwald (1995, Footnote 5).

10. In order to use the Multiple  $R$  as a simulation parameter, it was necessary to return to the Y-error strategy of the preliminary simulations. This strategy treats X and Z as measured without error and incorporates error only into the criterion, Y, at the level required to produce the target value of  $R$ .

11. The theoretical value of .05 was used in place of the observed value of .054, which was presumably a chance deviation from theoretical expectation.

12. Details of the data and results of the 16 pairs of simulations are given in Appendix Table A1. The 16 pairs of simulations were repeated using the two alternative methods that permitted modeling measurement error in X and Z. (In these simulations,  $R$  could not be used as a parameter.) With fixed error variance in Y the 4-test and SMR methods were similar in sensitivity to multiplicative models (average sensitivities over

the 16 data sets were, respectively,  $d' = 3.20$  and  $d' = 3.21$ ). With fixed correlation ( $r = .6$ ) between latent  $Y$  and observed  $Y$ , the 4-test method was generally more sensitive than the SMR method (sensitivities, respectively, of  $d' = 3.17$  and  $d' = 2.30$ ).

13. The pre-subtraction measures  $X$  and  $Z$  had rational zero values because the  $XZ$  product values generated from them had a value of zero if either or both of the original measures had a value of zero. The linear (subtraction) transformations destroyed this rational-zero property. These simulations used the method of fixing the Multiple  $R$  (at  $R = .40$ ), treating  $X$  and  $Z$  as measured without error.

14. We thank Betsy Becker for recommending the meta-analytic strategy of conducting three separate tests, each including five independent effect sizes.