How to Test Questions about Similarity in Personality and Social Psychology Research:

Description and Empirical Demonstration of Response Surface Analysis

Maxwell Barranti

Erika N. Carlson

Stéphane Côté

University of Toronto

Author Note

Correspondence concerning this article should be addressed to Maxwell Barranti, Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road, Mississauga, Ontario L5L 1C6. Email: max.barranti@mail.utoronto.ca

Abstract

Social and personality psychologists are often interested in the extent to which similarity, agreement, or matching matters. The current paper describes response surface analysis (RSA), an approach designed to answer questions about how (mis)matching predictors relate to outcomes while avoiding many of the statistical limitations of alternative, often-used approaches. We explain how RSA provides compressive and often more valid answers to questions about (mis)matching predictors than traditional approaches provide, outline steps on how to use RSA (including modifiable syntax), and demonstrate how to interpret RSA output with an example. To bolster our argument that RSA overcomes many limitations of traditional approaches (i.e., incomplete or misleading inferences), we compare results from four popular approaches (i.e., difference scores, residuals, moderated regression, and the Truth and Bias Model) to those obtained from RSA. We discuss specific applications of RSA to social and personality psychology research.

Keywords: similarity, agreement, accuracy, polynomial regression, response surface analysis

How to Test Questions about Similarity in Personality and Social Psychology Research:

Description and Empirical Demonstration of Response Surface Analysis

Psychologists, practitioners, and the general public are often interested in questions about whether (mis)matches matter. Does similarity foster attraction or do opposites attract (Luo, & Klohnen, 2005; Selfhout, Denissen, Branje, & Meeus, 2009)? Are positive illusions adaptive or are realistic self-perceptions the hallmark of mental health (Church et al., 2014, Duftner et al., 2012)? Do employees perform better when their values match the values espoused by their organizations (Edwards & Parry, 1993)?

Table 1 outlines example questions in psychology concerning whether (mis)matching perspectives are associated with more (or less) favorable outcomes. These questions are at the heart of theoretical issues in the field (e.g., is self-knowledge adaptive?) and have important practical implications (e.g., should people learn more about themselves?), but they present formidable analytical challenges (Cronbach & Furby, 1970; Edwards, 1994). Indeed, these analytical difficulties have left many questions about the importance of (mis)matched predictors in psychology unanswered. To advance knowledge in these areas, we describe polynomial regression and response surface analysis (RSA; Edwards, 1994; Edwards & Parry, 1993; Nestler, Grimm, & Schönbrodt, 2015), a comprehensive analytical tool specifically designed to answer questions about whether (mis)matches matter.

Table 1

Common Questions in Social and Personality Psychology that RSA Answers

Index of (mis)matching	Example questions
Interpersonal Attraction and Relationships	
Similarity	Do people who are more similar to one another have higher quality relationships than do people who are less similar? Do similar individuals like each other more or less than do dissimilar individuals?
Assumed similarity	Do romantic partners who think they are more similar to one another have higher quality relationships than do partners who think they are less similar?
Preferences	Are people more likely to pursue a relationship with a potential partner whose actual qualities match their ideal mate preference compared to a potential mate whose actual qualities differ from their ideal mate preference?
Equity	Do romantic partners who contribute equally to their relationship report higher quality relationships than do partners who report inequality?
Interpersonal Perception	
Personality Accuracy	Are individuals who can accurately evaluate other people's personalities more adjusted than individuals who form less accurate judgments of others?
Meta-Accuracy	Are individuals who know how other people perceive them more adjusted than individuals who are less aware of how others view them?
Empathic Accuracy	Are people who accurately identify others' emotions more successful than people who misread others' emotions?
Self-Processes	
Social comparison	Do people who tend to make upward comparisons report more or less psychological adjustment than people who tend to make downward comparisons? Do these individuals perform better or worse in the domain of comparison as a result?
Self-enhancement	Are individuals whose self-perceptions match what other people think about them better adjusted than individuals whose self-perceptions are less closely matched with what other people think about them? Is self-enhancement associated with well-being?
Self-verification	Does the match between a target's self-perception and the impression a judge forms of that target predict the target's willingness to pursue a

relationship with the judge?

Self digests	Are people whose actual self matches their ideal self happier than people whose actual self differs from their ideal self? Are people whose actual self matches their ought self happier than people whose actual self differs from their ought self?

Person-Environment Fit

Employee abilities and workplace demands	Are employees whose abilities that match the demands of their jobs more productive and satisfied than employees whose abilities differ from the demands of their jobs?
Employee and organizational values	Are employees whose values match those of their employers more productive and satisfied than employees whose values differ from those of their employers?
Cultural values	Is well-being related to a match between personal and cultural values?
Flow	Does a match between challenge and skill relate to in the moment satisfaction?

Self-Knowledge

Social reality	Do people who overestimate their social value (e.g., status, popularity) get ahead or do they face negative consequences compared to people who have an accurate sense of their social value?
Abilities	Are individuals who know their levels of cognitive and emotional abilities more successful than people who over- or under-estimate their levels of abilities?
Emotional forecasting	Is the ability to accurately predict future emotional states linked to higher psychological adjustment, or are people who are overly optimistic or pessimistic about their future states more psychologically adjusted?
Leadership	Do employees trust leaders whose self-perceptions of their strengths match their actual strengths more than leaders whose self-perceptions of their strengths are incorrect?

Consistency

Authenticity	Are people happier when they behave in ways that match their self-perceptions, compared to when they behave in ways that differ from their self-perceptions?
Attitudes	Are people whose implicit and explicit attitudes match more adjusted than people whose implicit and explicit attitudes differ?

	Do people who change their attitudes versus hold onto their original attitudes engage in more attitude-congruent behavior (e.g., voting)?
Behavioral concordance	Are people enjoyed more by others when they behave in ways that are consistent over time?
Change	Is personality change good or bad, and does the direction of personality change qualify whether personality change is good or bad?
Body Image	
Weight	Are people whose ideal weight match their actual weight happier compared to when there is a discrepancy?
Interpersonal Dynamics	
Warmth	Do people enjoy interactions more with an interaction partner when they match in warmth versus mismatch?
Dominance	Do people enjoy interactions more or less when there is a mismatch in dominance (i.e., when one partner is dominant and the other is submissive)?

Why should researchers learn about RSA? RSA provides comprehensive answers to core questions in psychology, such as those listed in Table 1, and is far superior to frequently used alternative approaches that often provide incomplete or even erroneous conclusions because of their statistical limitations. To foster a better appreciation and understanding of why and how to use RSA, we explain the merits of the approach, provide instructions on how to use RSA, and interpret results using real data. We leverage free R software and provide syntax that researchers can adapt to their own research questions. We also compare results from four popular approaches with those from RSA to demonstrate how these alternatives produce incomplete (at best) or misleading inferences (at worst). In sum, the current paper aims to encourage the use of RSA by providing an intuitive guide on why and how to adopt this approach.

Merits of RSA

RSA has at least two major conceptual strengths. First, RSA assesses whether (mis)matches matter by modeling how all possible combinations of two predictors are associated with an outcome, and does so in three-dimensional space (Edwards, 1994; Edwards & Parry, 1993; Nestler et al., 2015; Shanock, Baran, Gentry, Pattison, & Heggestad, 2010). This has important consequences for how much information RSA provides and for the validity of the results. With respect to validity, RSA models (mis)matching without using mathematical operations that conceal or distort information, such as the subtraction of one predictor from the other (i.e., difference scores; Edwards, 2002). Further, matches are operationalized in an intuitive way, specifically as the exact match between predictors. Using the example of Jordan and Taylor, the pair is matched if Jordan's level of an attribute is the same as Taylor's level, such as when both are a 6 on a 1 to 7 scale. Plotting response surfaces in three-dimensional space provides a thorough visualization and facilitates researchers' understanding of their data. In sum, RSA models matches in an intuitive, statistically valid, and comprehensive way.

Second, RSA answers more nuanced questions than traditional approaches. Like many traditional approaches, RSA tests if matching attributes are associated with more (or less) favorable outcomes than mismatching attributes (e.g., if self-knowledge is more or less adaptive than self-deception). However, theories about the consequences of (mis)matches can—and likely often should—be more complex. RSA is designed to address these complexities. In particular, rather than stopping at the general finding that matches are overall better than mismatches (e.g., self-knowledge is better than self-deception, or similarity is better than dissimilarity), a researcher can use RSA to discover if matched attributes at one level of the predictors have different outcomes than matched

attributes at another level. For instance, RSA would detect—but alternative approaches would fail to show—that Jordan and Taylor are less likely to split up if they both have high levels of agreeableness than if they both have low levels of agreeableness. Examples like this, where matches at some levels are not better than mismatches, are easy to imagine but are missed by approaches that fail to differentiate between matches at different levels of a predictor.

In addition, a researcher can use RSA to test whether one type of mismatch (e.g., an overestimate) is worse than another (e.g., an underestimate). For example, if researchers find that greater discrepancies in intelligence between partners predicts lower quality relationships, researchers would also want to know if some types of discrepancies are worse than others. Is Jordan less satisfied when Jordan is more intelligent than Taylor or less intelligent? Or, is self-enhancement better or worse than self-effacement? Thus, rather than limiting hypotheses to the basic question of whether a match is better or worse than a mismatch, RSA answers richer questions about how (mis)matches matter. Indeed, past research using RSA has revealed that (mis)matches are often not the same (Barranti, Carlson, & Furr, 2016; Bleidorn et al., 2016; Edwards & Rothbard, 1999).

Steps for Conducting RSA

The entire process that we outline below can generally be achieved in one step using the RSA package in R (Schönbrodt, 2016). However, RSA conceptually involves two steps: a) running a polynomial regression model and b) using effects from this model to generate a response surface and test for if and how mis(matches) matter (Box & Draper 1987; Edwards & Parry 1993). Thus, the interpretation of results of RSA focuses on the response surface rather than the polynomial regression effects.

To use RSA, data must meet the assumptions of multiple regression (Shanock et al., 2010). Additionally, the two predictors must be commensurate, representing the same content domain and measured on the same interval or ratio scale (Edwards, 1994; 2002). A researcher could use RSA to explore if there are costs associated with (mis)matching self- and peer-perceptions of intelligence on the same Likert-type scale, but could not explore costs associated with (mis)matches between self-perceptions on a Likert-type scale and measures of actual intelligence on a different scale (e.g., Wonderlic). The outcome can be measured on a different scale.

Establish the Existence of Both Matches and Mismatches

Researchers should verify that the data include both matched and mismatched observations because the results are not reliable in the absence of one or the other. The RSA package automatically generates this output (i.e., the percent of observations where X is greater than, equal to, or less than Y) based on whether the predictors are within half a z-score unit.

Center Predictors

Centering both predictors on the scale midpoint ensures that the interpretation of the results is consistent with theories of how (mis)matches relate to outcomes (i.e., as the exact match between predictors). Predictors should be unstandardized. If predictors are standardized, a one unit change in one predictor may not have the same substantive meaning as a one unit change in the other predictor (Edwards & Parry, 1993), precluding inferences about how (mis)matching relates to outcomes. Researchers should exercise extreme caution if they do not center predictors on the scale midpoint because it fundamentally changes the interpretation of a match—often to something convoluted,

unintuitive, and inconsistent with theory. For example, mean centering predictors

operationalizes a match as each predictor deviating from their respective mean by the

same amount, which substantially complicates interpretation.

Conduct Polynomial Regression

Regress the outcome on the main effects of X and Y, their squared terms (X^2 and

Y^2), and the interaction term (X*Y). If the polynomial regression model is significant and

the inclusion of the squared terms and interaction increased R^2, the next step is to

examine the three-dimensional response surface and the tests of its shape.

Generate the Response Surface

The RSA package automatically generates the response surface. A hypothetical

example is shown in Figure 1. As shown, the X and Y-axes range from negative to

positive values, and zero reflects the scale midpoint. Thus, positive values (e.g., +2)

represent the points above the midpoint, and negative values (e.g., -2) represent points

below the midpoint. The Z-axis depicts the outcome on its own scale of measurement.

This hypothetical response surface displays the expected values of the outcome at all

possible combinations of the two predictors. For example, it indicates the expected Z-

value when X and Y are both high (the back corner where both are +2) or low (the front

corner where both are -2), when X is high while Y is low (right corner), when Y is high

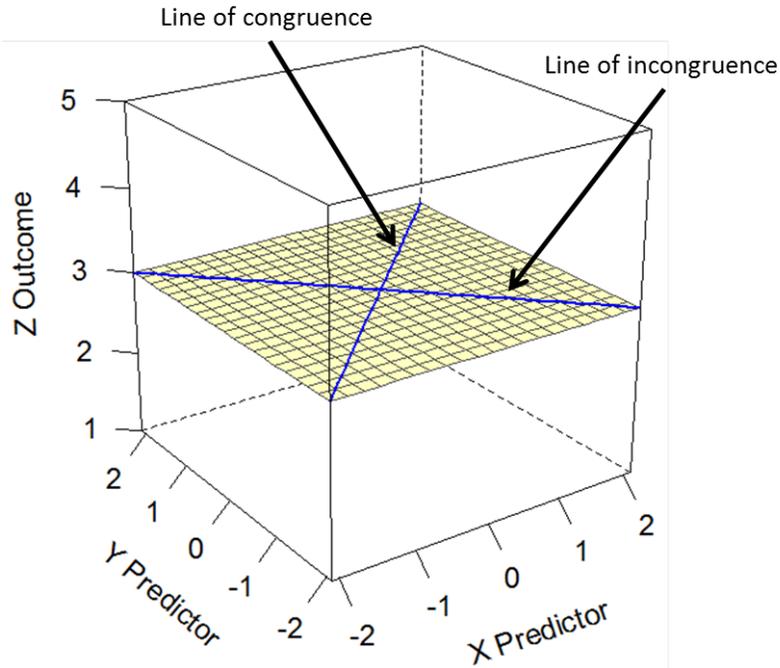while X is low (left corner), and everything in between.

Figure 1. Response Surface with Labeled Features. Predictors are centered on the

midpoint of the scale. X and Y values of 0 reflect the midpoint of the scale. The ***line of***

congruence reflects cases where values of X and Y perfectly match, at all levels of the

scale. The ***line of incongruence*** represents cases where values of X are the opposites of

values of Y.

Figure 1 also shows the two lines that test hypotheses about (mis)matched

predictors. The ***line of congruence*** reflects cases where values of X and Y perfectly

match at all levels of the scale. Using a similarity example, this line indicates points

where Jordan and Taylor both report being very low (-2) or both report being fairly high

(+1). The ***line of incongruence*** represents cases where values of X are the opposites of

Y. This line would indicate all points where, if Jordan reports being high (+2), Taylor

reports being low (-2), or if Jordan reports being fairly high (+1), Taylor reports being

fairly low (-1).

Interpret Tests of the Response Surface's Shape

RSA automatically provides statistical tests for four coefficients (a_1-a_4) that answer unique questions about how (mis)matches matter. Table 2 outlines each of the four questions these coefficients answer and illustrates response surfaces for possible answers to these questions (more details on the statistical tests of the coefficients appear in supplemental materials). Rather than discuss coefficients in numerical order, we explain them in terms of the conceptual questions they test. We first discuss each of the coefficients in isolation, describing how each coefficient should be interpreted when it is significant but all other coefficients are not. We then provide example interpretation for when more than one coefficient is significant.

**Are matches associated with higher or lower outcomes than mismatches?**

The test of the curvature of the line of incongruence, the a_4 coefficient, is the critical test for whether the mismatching of predictors matters overall. It indicates if the outcome increases or decreases more sharply as predictors diverge. Thus, a_4 could reveal if, for example, self-knowledge predicts greater adjustment than self-deception, or if similarity predicts more liking than dissimilarity. The bottom right panel of Table 2 shows examples of a_4 effects. As shown, it essentially tests if outcomes are higher (or lower) in the middle of the line (where X and Y are matched) compared to the ends of the line (where X and Y differ more). A positive a_4 indicates a convex (upward) curve, suggesting the outcome increases more sharply as the two predictors diverge. A negative a_4 indicates a concave (downward) surface, suggesting that the outcome decreases more sharply as the two predictors diverge.

Does the type of discrepancy matter?

RSA also reveals if the direction of mismatch matters by testing the slope of the line of incongruence, the a_3 coefficient. In the context of self-knowledge, a_3 would reveal if people are less liked when they self-enhance versus self-efface. As shown in the bottom left panel of Table 2, a positive a_3 indicates that the outcome is higher when X is greater than Y than the other way around. This would suggest that people are more liked when their self-views (X) exceed actual ratings (Y) than when their actual ratings are higher than their self-views. A negative a_3 indicates that the outcome is higher when Y exceeds X. In our example, this would suggest people are more adjusted when their actual ratings are higher than their self-views.

Table 2.

Four RSA Coefficients and the Questions They Answer

Line of congruence: How do matches matter?			
Slope of the line of congruence: a_1		**Curvature of the line of congruence: a_2**	
Do matches at high values have different outcomes than matches at low values?		Do matches at extreme values have different outcomes than matches at less extreme values?	
Positive a_1	Negative a_1	Positive a_2	Negative a_2
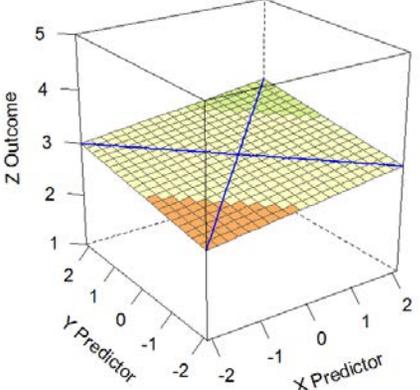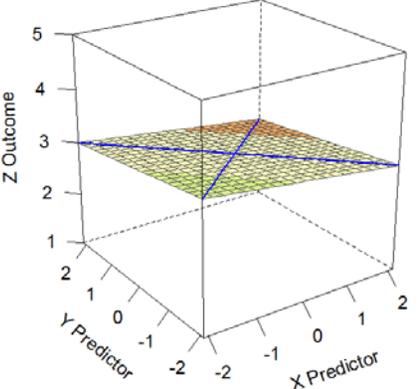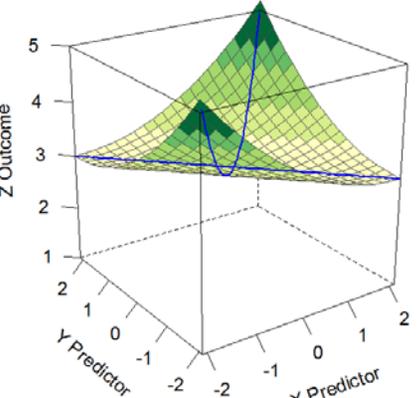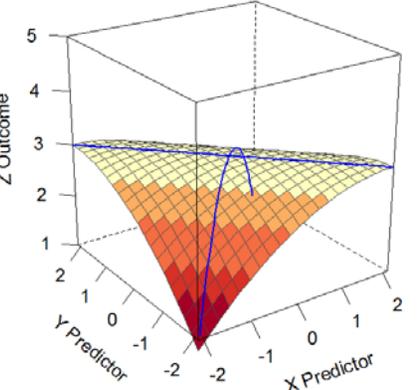
The outcome is higher when X and Y match at higher levels than at lower levels.	*The outcome is higher when X and Y match at lower levels than at higher levels.*	*The outcome is higher when X and Y match at more extreme levels than at midrange levels.*	*The outcome is higher when X and Y match at midrange levels than at more extreme levels.*

Line of incongruence: How do mismatches matter?			
Slope along the line of incongruence: a_3		**Curvature of the line of incongruence: a_4**	
Is one mismatch $(X > Y)$ better or worse than the other $(X < Y)$?		Are matches better or worse than mismatches?	
Positive a_3	Negative a_3	Positive a_4	Negative a_4
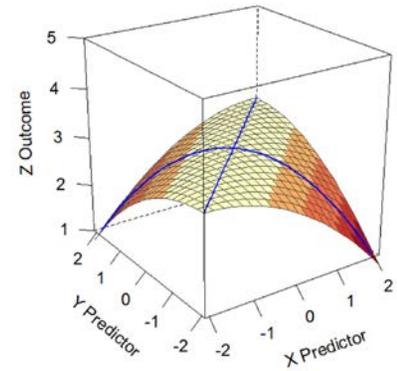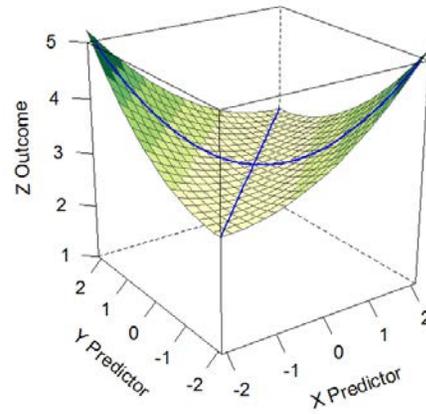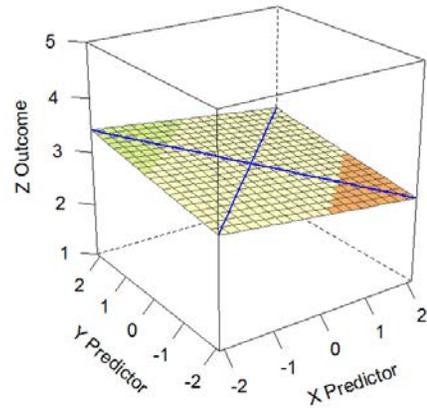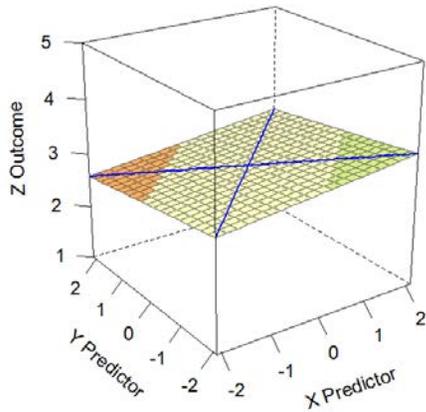
The outcome is higher when X is higher than Y than when Y is higher than X. *The outcome is higher when Y is higher than X than when X is higher than Y.* *The outcome is higher the more X and Y deviate from one another.* *The outcome is higher the more X and Y match one another.*

Note. Coefficients are based on polynomial regression's unstandardized coefficients: $a_1 = b_1 + b_2$; $a_2 = b_3 + b_4 + b_5$; $a_3 = b_1 - b_2$; $a_4 = b_3 - b_4 + b_5$. Please see the supplemental materials for more modeling details and graphing syntax.

Are some matches better or worse than other matches?

The test of the slope of the line of congruence, the a_1 coefficient, reveals if the effect of a perfect match is different at higher or lower levels of the scale. Using our self-knowledge example, a_1 indicates if self-knowledge for high levels is more or less adaptive than self-knowledge for low levels of the attribute. As shown in the upper left panel of Table 2, a positive a_1 indicates that matches at higher levels are associated with higher outcomes than matches at lower levels. A negative a_1 coefficient indicates that matches at higher levels are associated with lower outcomes than matches at lower levels.

Do matches at extremes have different effects than matches at mid-levels?

The test of the curvature of the line of congruence, the a_2 coefficient, indicates if matches at extreme ends of the scale predict higher or lower standing on the outcome than matches at midrange levels. More specifically, the a_2 indicates if the outcome increases or decreases more sharply as predictors match at increasingly high and low levels. A positive a_2 indicates a convex (upward) curve or that matches which deviate from the scale midpoint predict higher outcomes than matches at mid-levels of the scale. Using a self-knowledge example, a positive a_2 might be observed if it is especially important for individuals to leverage their high levels of ability and also to be aware of any low levels of ability. A negative a_2 suggests a concave (downward) surface, suggesting that self-knowledge has diminishing returns at increasingly higher and lower ends of the scale.

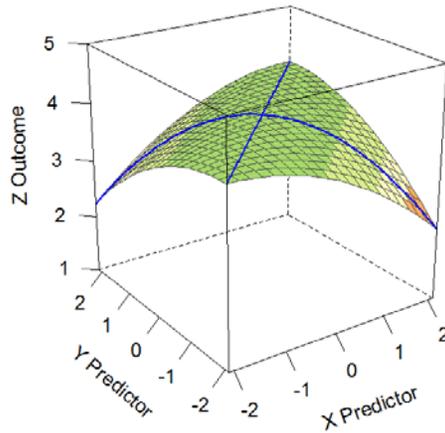**Interpreting combinations of coefficients.**

Each RSA coefficient yields important information, but researchers are also interested in the combination and size of these effects. Indeed, focusing on one

coefficient and ignoring the others could lead researchers astray, because the outcome is

often determined by a combination of effects. Figure 2 helps demonstrate the importance

of considering the combination of effects by illustrating an a_4 coefficient with increasing

size levels of an a_3 coefficient. In our analysis of real data below we provide another

example of how to interpret complex response surfaces.

Interpreting Combinations of Coefficients along the Line of Incongruence

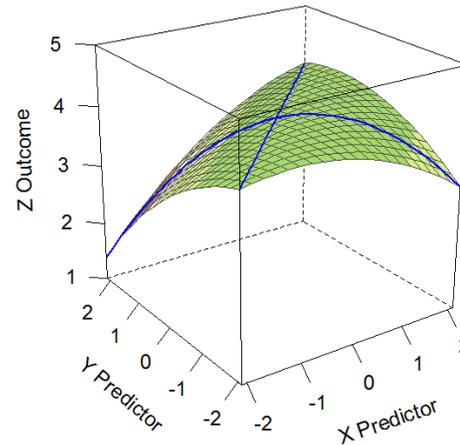A) Stronger curvilinear than linear effect: B) Equal curvilinear and linear effects: C) Stronger linear than curvilinear effect:

Moderate curvilinear effect ($a_4 = -.40$)   Moderate curvilinear effect ($a_4 = -.40$)   Moderate curvilinear effect ($a_4 = -.40$)
No linear effect ($a_3 = 0.0$)   Moderate linear effect ($a_3 = .40$)   Strong linear effect ($a_3 = .80$)

The outcome is higher the more X and Y match one another.

Matches tend to be better than mismatches, but underestimates are worse than overestimates.
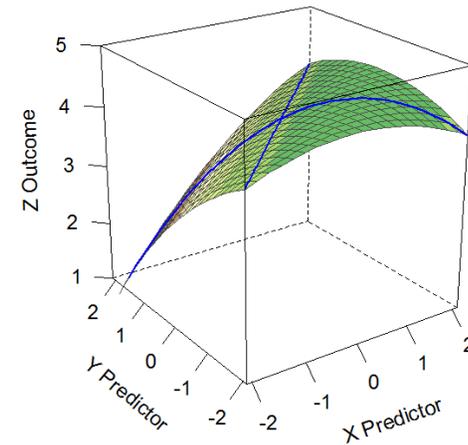
Matches tend to be better than mismatches overall, however overestimates also tend to be associated with the same favorable outcomes as matches, while underestimates are associated with unfavorable outcomes.

Figure 2. Interpreting Coefficient Combinations. From left to right, response surfaces reflect constant $a_1 = 0.00$, $a_2 = 0.00$, and $a_4 = -0.40$ coefficients, with a_3 coefficients of increasing magnitude: A) $a_3 = 0.00$; B) $a_3 = 0.40$; C) $a_3 = 0.80$.

Example of RSA Analysis

For our demonstration, we focus on how assumed similarity on the personality trait of conscientiousness (i.e., how much Jordan's conscientiousness matches Jordan's belief about Taylor's conscientiousness) is associated with relationship quality among romantic couples. Past work suggests that assumed similarity predicts higher quality (Montoya, Horton, & Kirchner, 2008), but to our knowledge this question has not been answered with RSA. We first show how this question can be addressed with RSA, and then explain how traditional methods provide incomplete or erroneous conclusions about if and how assumed similarity matters.

Our data are a subsample of the St. Louis Personality and Aging Network study, specifically participants who nominated their romantic partner as an informant ($N = 322$; age $M = 62.22$, $SD = 2.72$; 60% male; 81.4% Caucasian, 17.7% African American, .3 % Latino, .6% Middle Eastern (see Oltmanns, Rodrigues, Weinstein, & Gleason, 2014 for study details). Partners knew each other for about 30 years ($M = 32$, $SD = 12$). A power analysis revealed that this sample provided .99 power to detect a medium ($f^2=.15$) sized change in R-square going from a two main effects model to a polynomial model (i.e., adding the interaction and two quadratic terms), or 0.54 power to detect a small ($f^2=.02$) sized change (Faul, Erdfelder, Buchner, & Lang, 2009).

Participants (Jordan) described their own and their partners' (Taylor) personality on Five Factor Model traits (Costa & McCrae, 2009). Our analyses focus on conscientiousness (self-report: $M = 2.86$; $SD = 0.57$, $\alpha =.68$; impression: $M = 2.83$; $SD =0.76$, $\alpha =.86$), but we report results for other traits in the supplemental section. We focus on the perception of quality from the partners making the assumed similarity judgments

(i.e., Jordan). Conscientiousness was measured using a 5-point Likert-type scale ranging from 0 to 4. We centered the scores by subtracting the scale midpoint (i.e., 2). Perception of quality was measured by the 4-item version of the Dyadic Adjustment Scale (Sabourin, Valois, & Lussier, 2005; $M = 16.21$, $SD = 3.17$; $\alpha = .83$).

Figure 3 shows how all combinations of Jordan's self-perception (X) and Jordan's impression of Taylor's (Y) conscientiousness relate to Jordan's satisfaction (Z). Did assumed similarity predict higher relationship satisfaction? People were overall more satisfied when they thought they were more similar to their partner than when they thought they were more dissimilar, an association revealed by a negative curvature of the line of incongruence ($a_4 = -1.64$; 95% CI [-2.64, -0.60]).

Dissimilarity was associated with lower quality than similarity, but were all mismatches equally detrimental? No. People were particularly unsatisfied when they perceived themselves to be more conscientiousness than their partner, an association revealed by the negative slope of the line of incongruence ($a_3 = -1.45$, 95% CI [-2.48, -0.43]). Thus, there were costs to feeling dissimilar to one's partner in terms of conscientiousness, and these costs were particularly large when people believed they were more (rather than less) conscientious than their partner.

Assumed similarity predicted higher quality than assumed dissimilarity, but was assumed similarity equally beneficial at all levels of conscientiousness? People were more satisfied when they thought their partner was similar to them at high versus low levels of conscientiousness, an association revealed by the positive slope of the line of congruence ($a_1 = 1.68$, 95% CI [0.08, 3.28]). There was no curvilinear association along the line of congruence ($a_2 = -.01$, 95% CI [-0.94, 0.92]), suggesting that Jordan's

satisfaction did not increase more sharply as Jordan assumed similarity at extremely high
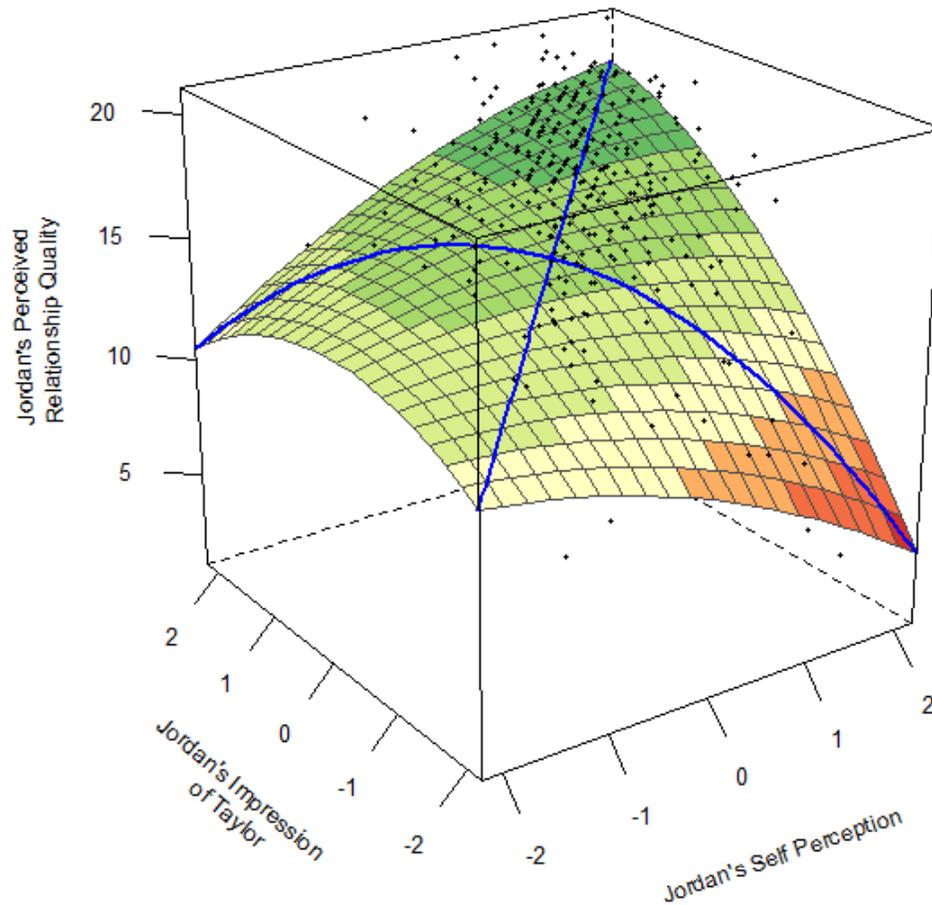
versus low levels of conscientiousness.

Figure 3. Response Surface for Assumed Similarity of Conscientiousness. The

polynomial coefficients were as follows: b_0 = 15.15, 95% CI [14.33, 15.97]; Jordan's

self-perception b_1 = .11, 95% CI [-.94, 1.17]; Jordan's impression of Taylor b_2 = 1.57,

95% CI [.74, 2.40]; Jordan's self-perception squared b_3 = -.27, 95% CI [-.91, .36];

Jordan's self-perception and impression interaction b_4 = .81, 95% CI [.12, 1.51]; Jordan's

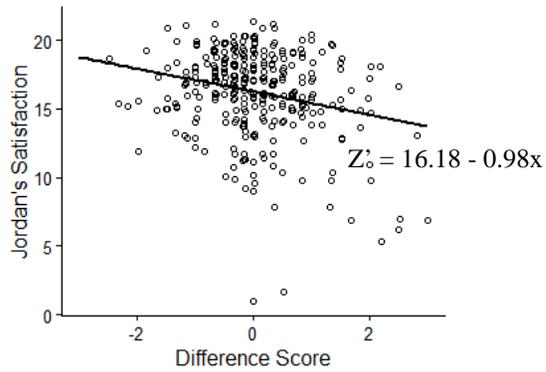impression of Taylor squared b_5 = -.55 [-0.92, -0.18].

How Do Results of RSA Compare to Alternative Approaches?

Questions about whether (mis)matches matter have been researched extensively

using approaches other than RSA. To demonstrate their limitations, we reanalyzed our

data using four common approaches for testing hypotheses about matching attributes: a)

difference scores (including absolute difference scores), b) residual scores, c) moderated

regression, and d) the Truth and Bias Model (West & Kenny, 2011). The results are

shown in Figure 4, but please see the supplemental section for detailed information about

these approaches. As we shall see, none of the alternative approaches provides the

information revealed by RSA, and some alternatives even provide erroneous conclusions.

A) Difference Scores   B) Residual Scores

Difference Score Approach   Absolute Difference Score Approach

Difference Score Approach: $Z'_o = 16.18 - 0.98x$

Absolute Difference Score Approach: $Z' = 17.05 - 1.34x$

Residual Scores: $Z' = 16.13 + 0.28x$

Conclusion: Satisfaction is not related to assumed similarity, $b = 0.28$, 95% CI [-0.35, 0.92]

Conclusion: Satisfaction tended to be higher when people assumed they were less conscientious than when they assumed they were more conscientious, with a match falling in between. $b = -0.98$, 95% CI [-1.35, -0.61].
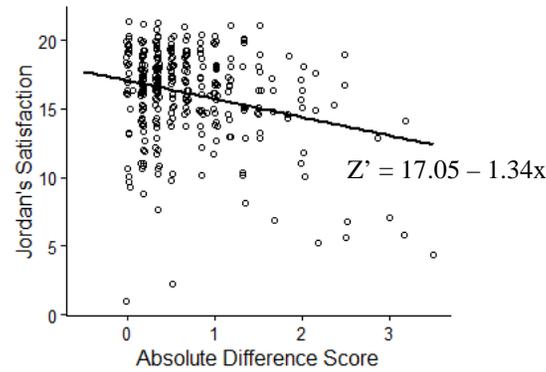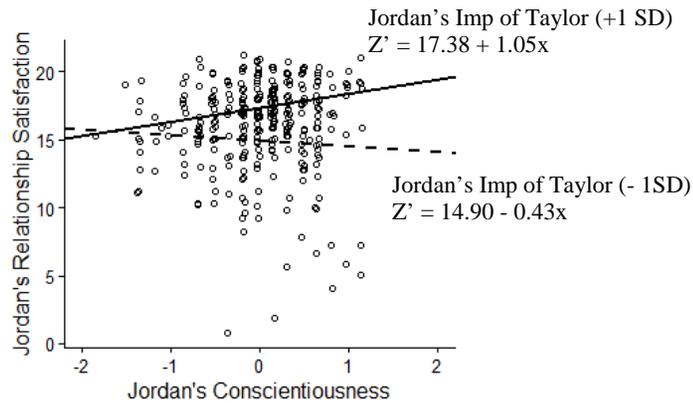
Conclusion: Satisfaction tended to be higher when people assumed they were more similar to their partner in terms of conscientiousness than when they thought they were more dissimilar (in any direction), $b = -1.34$, 95% CI [-1.88, -0.79].

C) Moderated Regression

D) Truth and Bias Model

Conclusion: The effect of Jordan's conscientiousness on relationship satisfaction depended on the level of Jordan's impression of Taylor's conscientiousness, $b_{interaction} = 1.00$, 95% CI [0.30, 1.69]. When Jordan viewed Taylor as high (+1 SD), relationship satisfaction tended to be higher as Jordan's conscientiousness was higher, $b = 1.05$, 95% CI [0.28, 1.82]. When Jordan viewed Taylor as low (-1SD), there was no significant relationship between relationship satisfaction Jordan's conscientiousness, $b = -0.43$, 95% CI [-1.19, 0.35].
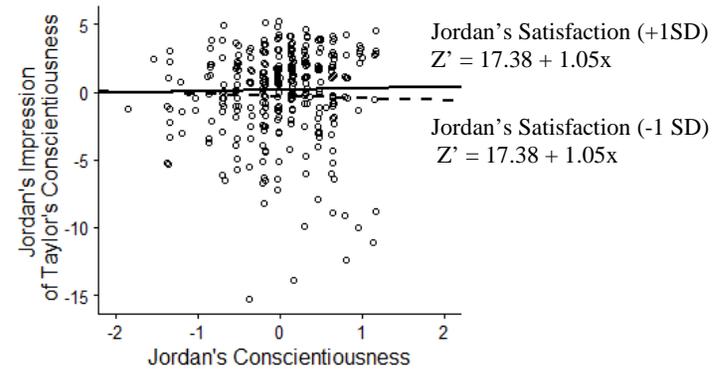
Conclusion: The relationship between Jordan's impression of Taylor's conscientiousness and Jordan's conscientiousness depended on Jordan's relationship quality, $b_{interaction} = 0.04$, 95% CI [0.00, 0.08]. However, the relationship between Jordan's conscientiousness and Jordan's impression of Taylors conscientiousness was not significant at high or low levels of relationship quality, $b_{+1SD} = 0.11$, 95% CI [-0.09, 0.30]; $b_{-1SD} = -0.17$, 95% CI [-0.33, 0.00]. There was no significant mean level difference, b = -0.05, 95% CI [-0.12, 0.02]. However, the main effect of the moderator was significant and indicated that the mean level difference between impressions depended on the level of relationship quality, $b = 0.08$, 95% CI [0.06, 0.80].

Figure 4. *Traditional Approaches to Testing if Assumed Similarity Relates to Relationship Satisfaction.* For panel A (Difference Scores), to calculate difference scores we subtracted Jordan's impression of Taylor's conscientiousness from Jordan's conscientiousness. For Absolute Difference Scores, we took the absolute value of the difference scores. For panel B (Residual Scores), we regressed Jordan's conscientiousness on Jordan's impression of Taylor's concientiousness and saved the residuals. For panel C (Moderated Regression), we mean-centered Jordan's conscientiousnes and Jordan's impression of Taylor's concientiousness, and then regressed Jordan's satisfaction on both centered variables and the interaction between them. The plot reflects the simple slopes of Jordan's conscientiousness at +1 and -1 standard deviation (SD) of Jordan's impression of Taylor's concientiousness. For panel D (Truth and Bias Model), Jordan's conscientiousness and Jordan's impression of Taylor's conscientiousness were centered on the mean of Jordan's conscientiousness, and Jordan's relationship satisfaction was mean-centered. We regressed Jordan's impression of Taylor's conscientiousness on Jordan's conscientiousness and Jordan's satisfaction. The plot reflects the simple slopes of Jordan's conscientiousness at +1 and -1 SD of Jordan's satisfaction.

Difference Scores

For both the difference score and absolute difference score approaches, a score of zero reflects a perfect match. For difference scores, positive and negative scores reflect mismatches. For absolute difference scores, positive scores reflect mismatches. The results of these analyses are shown in Figure 4. If we only examined difference scores, we would conclude that relationship quality is highest when people perceive their partner as more conscientious than themselves, moderate when people perceive their partner is equally conscientious, and lowest when people perceive their partner as less conscientious than themselves. This conclusion is erroneous—RSA revealed that when people perceive their partners as more conscientious than themselves, their relationship quality is lower than when people perceive a match (but not as low as when people perceive that they are more conscientious than their partner).

Further, if we only examined absolute difference scores, we would conclude that assumed similarity is positively related to satisfaction, compared to assumed dissimilarity. This conclusion is correct. However, we would also conclude that relationship quality is the same when individuals think they are more conscientious than their partners and when individuals think they are less conscientious than their partners. This conclusion is also erroneous—RSA revealed that the former is associated with significantly lower quality than the latter.

Combining the conclusions from difference score and absolute difference score results will still provide limited conclusions as compared to RSA. With both versions of the difference score approach, matches at all levels (high-high, moderate-moderate, or low-low) are zero. Thus, difference scores and absolute difference scores cannot reveal if

assumed similarity has the same effect at all levels of conscientiousness (as RSA does with an a_1 coefficient), nor can they test nonlinear effects of matching (as RSA does with an a_2 coefficient). Therefore both versions of the different score approach miss the finding that assumed similarity at high levels of conscientiousness is associated with higher quality than assumed similarity at low levels. RSA detected this pattern.

Residual Scores

As is sometimes done in the literature, we computed residual scores by regressing one perspective (X; Jordan's self-perception) onto the other (Y; Jordan's impression of Taylor) and saved the residuals. The magnitude and direction of residuals indicate the degree to which what was predicted by one perspective tended to be above (or below) what was actually observed by the other (i.e., if Jordan was more or less conscientious than what would be predicted by Jordan's impression of Taylor). A residual of zero reflects perfect assumed similarity.

Results in Figures 4 suggested that there was no relationship between assumed similarity and satisfaction. Using this approach, researchers would infer that assumed similarity is unrelated to satisfaction. This conclusion is erroneous—RSA revealed that overall assumed similarity is associated with higher satisfaction than dissimilarity, plus one direction of dissimilarity is worse than the other. Further, like difference scores, the residual score approach masks effects of different types of matches (i.e., RSA's a_1 and a_2 coefficients) by assigning all matches a score of zero. Thus the residual score approach conceals the fact that assumed similarity at high levels of conscientiousness is associated with higher satisfaction than assumed similarity at low levels.

Moderated Regression

The moderated regression approach reveals if the link between one predictor (Jordan's self-perception) and the outcome (satisfaction) depends on the level of the other predictor (Jordan's impression of Taylor). For example, this approach reveals if Jordan is more satisfied when both Jordan's self-perception and Jordan's impression of Taylor are both high or both low. Results in Figure 4 suggested that assumed similarity was associated with satisfaction (i.e., the interaction was significant). Simple slope tests revealed that people were particularly satisfied when they perceived that both they and their partners were highly conscientious.

Results somewhat mirrored RSA effects, but unlike RSA, moderated regression does not provide a direct answer to the question posed in similarity research: whether matches are generally associated with higher quality than mismatches (which is revealed by the a_4 in RSA). The focus in moderated regression is on specific comparisons of arbitrary levels of each predictor (typically, + or – one SD from the mean) rather than overall comparisons of matches versus mismatches. Further, moderated regression typically does not formally test if high-high matches are associated with more satisfaction than low-low matches (a_1) or if high-low matches are associated with different levels of satisfaction than low-high matches (a_3) because these data points are on different regression lines (Shanock et al., 2010). Moderated regression also cannot test the possible curvilinear nature of matching (a_2). Researchers thus miss nuanced ways in which matching predicts outcomes.

Truth and Bias Model

Researchers have adapted the Truth and Bias Model (West & Kenny, 2011) to test hypotheses about matching. This approach essentially makes the outcome of RSA a

moderator and incorporates a centering procedure that formally tests directional or mean-level biases in predictors. We regressed one predictor (e.g., Jordan's impression of Taylor) on the other predictor (e.g., Jordan's conscientiousness), the outcome (e.g., Jordan's satisfaction), and their interaction. This model tests if the association between predictors depends on the 'outcome,' but it also formally tests if mean-level differences depend on the 'outcome' (see the supplemental for modeling details).

Results in Figure 4 suggested that the degree to which people tended to assume they were similar to their partner depended on their satisfaction (i.e., the interaction was significant), but simple slopes were not significant. The main effect of satisfaction was a significant positive predictor, suggesting that when Jordan's conscientiousness was higher than Jordan's impression of Taylor, Jordan was less satisfied. While these effects approximate some of the effects provided by RSA, they have similar limitations to moderated regression outlined above. There is no direct test of whether matches are associated with higher quality than mismatches. There are also no direct comparisons of different types of mismatches (e.g., Jordan thinks Taylor is more versus less conscientious than him). Finally, there are no direct comparisons between high-high and low-low matches and no test of potential curvilinearity among matches.

Summary of Comparison between Alternative Approaches and RSA

Our demonstration suggests that alternative, often-used approaches to testing questions about similarity do not provide as much information as RSA provides about if and when (mis)matches matter. While some approaches approximated one or two of the effects provided by RSA (e.g., overall assumed similarity was beneficial), none provided all of the information RSA provided, and some approaches led to erroneous conclusions.

The results that RSA identified—but that no other approach can— such as how matches matter (a_1 and a_2) are important for both theoretical and practical reasons. Researchers may develop inadequate theories if they assume that all matches are associated with the same outcomes, when in fact some types of matches have better outcomes than others. Further, theories may be inadequate if all mismatches are assumed to be equally detrimental, when in fact some mismatches are associated with worse outcomes than others. On a practical note from our data, therapists or laypeople informed by difference score results would believe that individuals who assume their partner is similar to them will be the most satisfied, when in fact, this is only true if both are high, not low, on conscientiousness.

Additional Features of RSA

RSA can be adapted to fit a variety of research designs. For example, the RSA package includes modifications for binary outcomes (e.g., voting behavior). Many research questions involve designs that introduce dependencies in the data (e.g., modeling similarity and both partners' satisfaction). To adapt RSA to answer questions that involve multilevel modeling, researchers center predictors on the scale midpoint, conduct polynomial regression in multilevel modeling, and use the unstandardized coefficients, standard errors, covariances, and degrees of freedom from the multilevel model to generate and test the response surface (e.g., Barranti et al., 2016; Muise, Stanton, Kim, & Impett, 2016). RSA can also be adapted to include control variables. Please see the supplemental materials for additional syntax for these more complex specifications.

Researchers might have questions that go beyond tests of the slope and curvature of the lines of congruence and incongruence. A researcher might find that self-knowledge

predicts better adjustment than self-deception (i.e., a negative a_4) and wonder if the a_4 effect apply at all levels of the attribute. Visually inspecting the graph could provide a researcher with some ideas as to whether this is true, but tests of simple slopes are needed to make formal conclusions about these effects (Edwards & Parry, 1993). To aid in answering these more complex questions about boundary conditions, we provide some additional information in the supplemental section (see Ilmarinen, Lönnqvist, and Paunonen (2016) for an example).

Finally, to provide a rule of thumb for sample sizes, we calculated 80% power to detect a change in R-squared when going from two main effects to the full polynomial model (five predictors). The rationale is that, if adding the interaction and squared terms does not increase the predictive power of the model, it would be inappropriate to probe for matching effects that are derived from the interaction and squared terms. As such, researchers should aim for 550 observations to detect a small ($f^2=.02$), 77 to detect a medium ($f^2=.15$), and 36 to detect a large effect ($f^2=.35$; Faul et al., 2009).

Conclusion and Implications

At the heart of some of the most important questions in social and personality psychology is if and when (mis)matches matter. Our goal was to present a flexible and statistically rigorous tool that can advance unresolved questions about (mis)matching perspectives, especially in light of the statistical challenges of such questions. Yet, it might also be useful to re-visit seemingly resolved questions in the literature using RSA. An important implication of our comparative examples is that the conclusions researchers make about (mis)matches depend on which approach they use. Traditional analytical approaches mask effects (e.g., whether effects of all matches are the same), and some

approaches lead to incorrect inferences due to severe problems with statistical validity (Cronbach, 1955, 1958; Cronbach & Furby, 1970; Edwards, 1994). By masking or distorting findings, traditionally used approaches ultimately undermine the validity of inferences, which has serious theoretical and practical consequences (Edwards, 1994; Edwards & Parry, 1993). Findings from entire literatures might need to be re-analyzed using RSA to better understand if and how (mis)matches matter. While we hope the current demonstration provides the necessary background for researchers to apply this tool to their own work, we also hope researchers use this tool to re-explore seemingly resolved issues in the field.

<div align="center">**References**</div>

Barranti, M., Carlson, E. N., & Furr, R. M. (2016). Disagreement about moral character is linked to interpersonal costs. *Social Psychological and Personality Science*, 7, 806-817. doi:1948550616662127.

Bleidorn, W., Schönbrodt, F., Gebauer, J. E., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2016). To live among like-minded others: Exploring the links between person-city personality fit and self-esteem. *Psychological Science*, *27*, 419-427. http://doi.org/10.1177/0956797615627133

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces* (Vol. 424). New York: Wiley.

Costa, P., & McCrae, R. (2009). The five-factor model and the NEO inventories. Oxford handbook of personality assessment [e-book]. New York, NY: Oxford University Press.

Cronbach, L. (1955). Processes affecting scores on" understanding of others" and"
   assumed similarity.". *Psychological Bulletin, 52*, 177-193.

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception
   scores. In Tagiuri R., Petrullo L. (Eds.) Person perception and interpersonal
   behavior (pp. 353-379). Stanford, CA: Stanford University Press.

Cronbach. L. J., & Furby, L. (1970). How should we measure "change"—or should we?
   Psychological Bulletin, *74*, 68-80.

Church, a T., Katigbak, M. S., Mazuera Arias, R., Rincon, B. C., Vargas-Flores, J. D. J.,
   Ibáñez-Reyes, J., … Ortiz, F. a. (2014). A four-culture study of self-enhancement
   and adjustment using the social relations model: Do alternative conceptualizations
   and indices make a difference? *Journal of Personality and Social Psychology*,
   106, 997–1014. doi:10.1037/a0036491

Dufner, M., Denissen, J. J. a, van Zalk, M., Matthes, B., Meeus, W. H. J., van Aken, M. a
   G., & Sedikides, C. (2012). Positive intelligence illusions: On the relation
   between intellectual self-enhancement and psychological adjustment. *Journal of
   Personality*, *80*, 537–572. doi:10.1111/j.1467-6494.2011.00742.x

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal
   of Management*, *20*, 683 – 689. doi:10.1177/014920639402000311

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*,
   265-287.

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis
   and response surface methodology. In F. Drasgow & N. W. Schmitt (Eds.),

Advances in measurement and data analysis (pp. 350–400). San Francisco, CA: Jossey-Bass.

Edwards, J.R., & Parry, M.E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*, 1577–1613.

Edwards, J. R., & Rothbard, N. P. (1999). Work and family stress and well-being: An examination of person-environment fit in the work and family domains. *Organizational Behavior and Human Decision Processes, 77*, 85-129.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149-1160.

Ilmarinen, V. J., Lönnqvist, J. E., & Paunonen, S. (2016). Similarity-attraction effects in friendship formation: Honest platoon-mates prefer each other but dishonest do not. *Personality and Individual Differences, 92*, 153-158.

Luo, S., & Klohnen, E. C. (2005). Assortative mating and marital quality in newlyweds: A couple-centered approach. *Journal of Personality and Social Psychology, 88*, 304–326.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships, 25*, 889-922.

Muise, A., Stanton, S. C., Kim, J. J., & Impett, E. A. (2016). Not in the mood? Men under-(not over-) perceive their partner's sexual desire in established intimate relationships. *Journal of Personality and Social Psychology, 110*, 725-742.

Nestler, S., Grimm, K. J., & Schönbrodt, F. D. (2015). The social consequences and

    mechanisms of personality: How to analyse longitudinal data from individual,

    dyadic, round-robin and network designs. *European Journal of Personality, 29*,

    272-295.

Oltmanns, T. F., Rodrigues, M. M., Weinstein, Y., & Gleason, M. E. (2014). Prevalence

    of personality disorders at midlife in a community sample: Disorders and

    symptoms reflected in interview, self, and informant reports. *Journal of*

    Psychopathology and Behavioral Assessment, 36, 177-188.

Sabourin, S., Valois, P., & Lussier, Y. (2005). Development and validation of a brief

    version of the dyadic adjustment scale with a nonparametric item analysis model.

    Psychological Assessment, 17, 15–27. http://dx.doi .org/10.1037/1040-

    3590.17.1.15

Shanock, L. R., Baran, B. E., Gentry, W. A., Pattison, S. C., & Heggestad, E. D. (2010).

    Polynomial regression with response surface analysis: A powerful approach for

    examining moderation and overcoming limitations of difference scores. *Journal*

    of Business and Psychology, *25*, 543–554. http://doi.org/10.1007/s10869-

    010-9183-4

Schönbrodt, F. D. (2016). RSA: Response surface analysis (R package Version 0.9.10)

    [Computer software]. Retrieved from http://cran.r-

    project.org/web/packages/RSA

Selfhout, M., Denissen, J., Branje, S., & Meeus, W. (2009). In the eye of the beholder:

    perceived, actual, and peer-rated similarity in personality, communication, and

friendship intensity during the acquaintanceship process. *Journal of Personality and Social Psychology*, *96*, 1152–1165. doi:10.1037/a0014468

West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review, 118*, 357-378.