Testing similarity effects with dyadic response surface analysis

Felix D. Schönbrodt
Ludwig-Maximilians-University, Munich

Sarah Humberg
Münster University

Steffen Nestler
Leipzig University

Dyadic similarity effect hypotheses state that the (dis)similarity between dyad members (e.g., the similarity on a personality dimension) is related to a dyadic outcome variable (e.g., the relationship satisfaction of both partners). Typically, these hypotheses have been investigated by using difference scores or other profile similarity indices as predictors of the outcome variables. These approaches, however, have been vigorously criticized for their conceptual and statistical shortcomings. Here, we introduce a statistical method that is based on polynomial regression and addresses most of these shortcomings: Dyadic response surface analysis (DRSA). This model is tailored for similarity effect hypotheses and fully accounts for the dyadic nature of relationship data. Furthermore, we provide a tutorial with an illustrative example and reproducible R and Mplus scripts that should assist substantive researchers in precisely formulating, testing, and interpreting their dyadic similarity effect hypotheses.

Keywords: congruence, similarity, dyadic data, response surface analysis, polynomial regression

A number of interesting psychological research questions in dyadic contexts refer to the effects of the dyad members' similarity on some outcome. For example, how is the similarity between the husband's and wife's personality associated with other variables such as the relationship satisfaction of each of them (e.g., Dyrenforth, Kashy, Donnellan, & Lucas, 2010)? Similarly, what is the correspondence between a supervisor's and a subordinate's personality and is this correspondence related to more positive performance outcomes (e.g., Strauss, Barrick, & Connerley, 2001)?

We will refer to hypotheses of this kind as dyadic *similarity effect hypotheses*; they state that the similarity (or congruence, or discrepancy) between two predictor variables are related to two outcome variables stemming from the same dyad members. Hence, we do not focus on the question whether or

how similar dyad members are in absolute terms, but rather on the effect of different levels of similarity (on a continuous scale) on a third variable. To test such dyadic similarity effect hypotheses, researchers need to assess dyadic data that is interdependent. For the relationship satisfaction example, such data could include the husband's and the wife's personality assessments for each couple, as well as each of their relationship satisfaction scores, which will most likely be highly related.

Most previous research on dyadic similarity effects computed discrepancy scores or profile similarity indices to measure the correspondence between the scores provided by two dyad members. The resulting similarity scores were then used as predictor variables in statistical models such as multiple regression, analysis of variance, or multilevel modeling. However, the usage of the similarity scores is statistically highly problematic due to untested constraints (e.g., Edwards, 2001). We therefore introduce dyadic response surface analysis (DRSA) as an alternative mean to examine dyadic similarity effects (see Nestler, Grimm, & Schönbrodt, 2015 for briefly mentioning this method, and Weidmann, Schönbrodt, Ledermann, & Grob, 2017, for an existing application of DRSA).

DRSA offers a number of advantages compared to similarity-score based approaches. First and foremost, as described below, DRSA circumvents most of the problems associated with the analysis of similarity indices. Second, DRSA allows a researcher to test whether the similarity effects are the same for both partners. This is less straight-

forward when a researcher uses a similarity-score based approach. Finally, DRSA can be employed when the outcome variables were measured on a continuous or on a categorical (e.g., binary) scale.

This article is organized as follows: We will first briefly discuss some previous approaches that have been used to examine dyadic similarity effects and point out their problems and limitations. We will then shortly introduce classical RSA as well as the Actor-Partner-Interdependence Model (APIM; Kenny, Kashy, & Cook, 2006), and then integrate these two methods to introduce DRSA. Finally, we will use an empirical example to illustrate the suggested method and give practical advice and code examples on how to compute the model in the *R* environment for statistical computing (R Core Team, 2014).

Throughout this article, we focus on the case of distinguishable dyads. That is, we consider dyads in which the dyad members can be differentiated from one another on the basis of a theoretically meaningful variable. We use the effect of couple personality similarity between male and female romantic partners on relationship outcomes as our exemplary research question. Gender thus serves as the distinguishing variable. This example can be easily transferred to all other kinds of distinguishable dyads, such as non-identical twins, mother-child dyads, or mentors and mentees. DRSA is also applicable for interchangeable dyads (i.e., no distinguishing variable exist) such as identical twins or homosexual dyads. We give instructions, references, and *R* code examples that demonstrate how to adapt the model to indistinguishable dyads.

Previous Approaches to Investigate Effects of Couple Similarity

Researchers used a variety of methods to examine the influence of couple similarity on third variables (see Kenny et al., 2006, for an overview). The bottom line of most of these approaches is to (1) compute a similarity score for each dyad and (2) relate the resulting similarity scores to the two interdependent outcome variables. A number of different indices have been used in the literature to quantify the correspondence between the dyad members' values, most of which can be divided into two types: discrepancy scores and profile correlation indices (see also Dyrenforth et al., 2010).

Discrepancy scores represent the difference between the two dyad members' values. A discrepancy score for a dyad is typically calculated by computing the absolute or squared difference between both members' scores on the respective variable (e.g., the difference of their agreeableness scores). Such discrepancy scores provide an index of how dissimilar the dyad members are, with smaller values of the score reflecting more similarity between the two dyad members. Although these discrepancy scores have a strong intuitive meaning, they are problematic when ones use them

in subsequent statistical models (Blanton, Jaccard, Gonzales, & Christie, 2006; Edwards, 2001; Johns, 1981; Peter, Churchill, & Brown, 1993). The most serious problem is that they impose implicit constraints on the data which are typically not tested. For example, usage of discrepancy scores presumes that (dis)similarity on a higher level (e.g., 9 - 8 = 1) is psychologically equivalent to (dis)similarity on a lower level (e.g., 2 - 1 = 1) and that these discrepancy scores result in the same psychological consequences. In the worst case, such untested constraints may lead researchers to accept their similarity hypothesis although there are in fact no similarity effects present in the data. This is equally true for related approaches that predicted the outcome variables from discrepancy scores, while statistically controlling for linear main effects of the predictor variables, for dyad averages of the predictors, or for interaction effects of dyad averages and the discrepancy scores. Such approaches loosen some, but not all implicit constraints, and are therefore still biased towards falsely accepting the similarity hypothesis (see also Edwards, 2001).

The second type of similarity measures, profile similarity indices, focus on the correspondence in the rank-ordering of different assessed variables across dyad members (e.g., the couples' similarity in their Big Five profiles). This similarity is determined by computing (a) the Pearson product-moment correlation or (b) the intraclass-correlation between the scores of the two dyad members. By capturing the similarity in the dyads members' idiosyncratic ordering of measures relative to each other, higher values in profile similarity indices indicate greater correspondence. Although frequently used, application of profile correlations in this context is seen as problematic for several reasons. First, the psychological meaning of profile correlation coefficients is less clear than the interpretation of discrepancy scores, because profile correlations do not only capture profile similarity but, for example, also level effects (see Kenny et al., 2006). Furthermore, when both dyad members tend to respond in a standard or normative fashion, the amount of similarity is over-estimated (Cronbach, 1955; Cronbach & Gleser, 1953; Kenny & Acitelli, 1994). Finally, and unrelated to whether one controls for level effects or not, profile similarity indices also impose equality constraints when one uses them in subsequent statistical models as they also involve sums of squared differences (Edwards & Parry, 1993).

Researchers usually combined one of the above described approaches with some strategy to take the nested structure of couple data into account, for example by accounting for within-dyad dependencies in separate predictions of female and male outcome variables, or by applying a multilevel framework. While such strategies are surely necessary when working with dyadic data, they cannot prevent from the limitations that are inherent in every approach that relates discrepancy or profile similarity scores to outcome variables.

Implicit, untested constraints occur in these models, independent of whether and how one accounts for the dyadic data structure.

To summarize, typical previous approaches contained untested constraints that can lead researchers to falsely accepting a similarity hypothesis, or to missing important effects which cannot be detected due to the constraints. Before closing we would like to note that the amount of nonindependence in a dyadic sample is sometimes interpreted as the dissimilarity (or conversely the similarity) of the scores from the dyads (see Kenny et al., 2006). Here, we are not interested in this type of similarity but rather in whether the within-dyad correspondence has an effect on an outcome variable. We are thus interested in similarity effects (sometimes called dyadic index effects, see Kenny et al., 2006) and which mathematical model should ideally be used for their investigation.

Dyadic RSA: A Marriage of RSA and APIM

Dyadic response surface analysis is based on a combination of response surface analysis (RSA) and the Actor-Partner-Interdependence Model (APIM). In the following, we first describe RSA and then go on with the APIM. Thereafter, the DRSA is introduced.

Response Surface Analysis

Similarity effect hypotheses have not only been posited for dyadic outcomes (e.g., the effect of couples' similarity in agreeableness on *both* partners' relationship satisfaction) but also for individual or between-dyad outcomes. Such a similarity effect hypothesis could be "the more similar a woman's openness X is to her husband's openness Y, the higher is a couple's amount of time Z they spend together". RSA is the appropriate statistical tool to investigate such similarity effects on a single outcome (e.g., see Edwards, 2002; Edwards & Shipp, 2007; Humberg, Nestler, & Back, in press; Schönbrodt, 2015).

The basic element of an RSA is the estimation of a second-order polynomial regression model:

$$Z = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + e \quad (1)$$

where the outcome variable Z is regressed on the predictors X and Y, their respective squared terms X^2 and Y^2, and their interaction XY. Note that only including the multiplicative interaction term is not sufficient to model a similarity effect (see Edwards, 2001, p. 269).

The key idea behind RSA is a visualization of the estimated regression equation (1) in a three-dimensional plot. An example response surface is shown in Figure 1: The two predictor variables X and Y are located on the two axes that span the bottom of the coordinate cube. For each combination of X and Y, the respective model-predicted Z value is

reflected on the vertical axis, thus indicated by the height of the surface. For example, given the (X, Y) coordinate $(0, -1)$ (e.g., a dyad in which the female has an openness score of 0 and the male has an openness score of -1), which is depicted as a circle on the bottom of the cube, the surface in Figure 1 predicts the outcome value $Z = 3.75$, which is symbolized as a star on the regression surface.
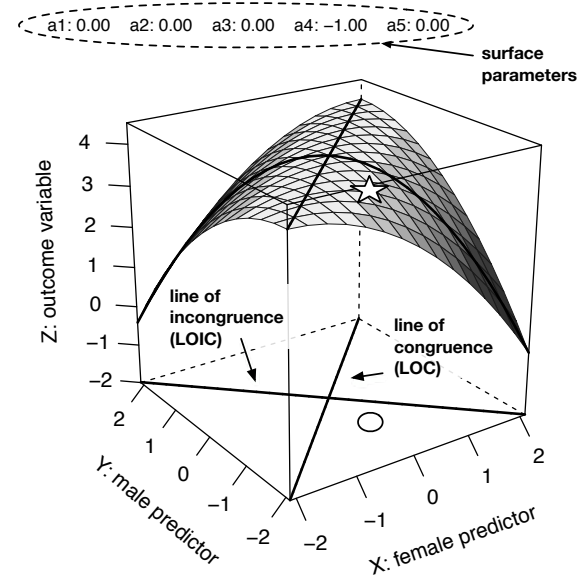

Figure 1. **A prototypical similarity effect: Response surface of the estimated regression equation** $Z = 4 + 0X + 0Y - 0.25X^2 + 0.5XY - 0.25Y^2$**. The circle on the cube's bottom is the position of a dyad's predictor variables (female:0, male: -1), the star is the corresponding predicted variable of the outcome variable on the response surface.** Figure available at https://osf.io/ftsrd/, under a CC-BY4.0 license.

RSA provides tools for a graphical and statistical interpretation of the regression surface, for example with regard to the (non-)existence of similarity effects.

The line of congruence. For similarity effect hypotheses, an important feature of response surface plots is the *line of congruence* (LOC, Edwards & Shipp, 2007). Perfect congruence of two variables is not reflected in a single combination of matching X and Y values, but in all combinations for which X equals Y. Hence, partners can be similar on a low level (e.g., both are low on openness), on a mid level, or on a high level (both are very open). All possible combinations of perfect similarity are located on the line of congruence on the bottom of the coordinate cube (see Figure 1), which diagonally connects the front corner (the congruent low/low combination) to the back corner (the congruent high/high combination) of the cube.

The surface above the LOC is depicted as a respective black line on the surface, which reflects how the predicted

outcome behaves for varying values of $X = Y$. Statistically, the surface above the LOC is described by the formula $Z = a_1 X + a_2 X^2$ (omitting the intercept). The newly introduced parameters a_1 and a_2 are derived from the estimated regression coefficients of equation (1), where $a_1 = b_1 + b_2$, and $a_2 = b_3 + b_4 + b_5$ (Shanock, Baran, Gentry, Pattison, & Heggestad, 2010). The parameter a_2 thus indicates whether the LOC is of a linear (if $a_2 = 0$) or of a curvilinear shape (if $a_2 \neq 0$), and a_1 describes the slope of the LOC above the point $(0, 0)$. In Figures 1, 2A, and 2B, a_2 equals zero, meaning that the LOC is a straight line. In this case, a_1 is simply the slope of the linear line. Because $a_1 = 0$ in Figure 1, the predicted outcome value above the LOC is constant. For an example with a linear and non-constant predicted outcome over the LOC, see Figure 2A, which depicts a *rising ridge surface*: The outcome is higher when X is identical to Y. Additionally to this basic similarity pattern there is an effect of the *level* of similarity: Congruent combinations on a high predictor level lead to higher outcome values than congruent combinations on a low predictor level. Figures 2D, 2E, and 2F, by contrast, depict surfaces with a curvilinear LOC. In Figure 2D, for example, the outcome is highest for specific combination of female and male openness (namely $X = 0$ and $Y = 0$), and it is lower for all other combinations of the two predictors.

The line of incongruence. Equally important for similarity hypotheses is the *line of incongruence* (LOIC, Edwards & Shipp, 2007), which is perpendicular to the LOC and specifically meaningful with regard to similarity effects. It consists of all combinations of the predictors for which $Y = -X$ holds. In Figure 1, this line is depicted as a black line on the bottom of the cube and on the surface, respectively, which connects the left corner and the right corner of the cube. It runs from one incongruent extreme (low X/high Y) to the other (high X/low Y). Analogously to the LOC, the LOIC is mathematically determined by a quadratic equation $Z = a_3 X + a_4 X^2$ (omitting the intercept), where $a_3 = b_1 - b_2$ and $a_4 = b_3 - b_4 + b_5$. Again, the coefficient of the quadratic term, a_4, indicates whether the LOIC on the surface is linear or curvilinear. In the case of Figure 1, a_4 is negative ($a_4 = -1$), indicating that the LOIC has an inverted U-shape: It has a highest point and symmetrically bends downwards at both sides of this point. In such a setting (i.e., when $a_4 \neq 0$) the coefficient a_3 reflects the position of the highest or lowest point of the LOIC. Here, $a_3 = 0$, which means that the highest point of the LOIC is positioned exactly above $(0, 0)$: The LOIC is highest for $X = Y = 0$ and bends downwards for values of X and Y that more and more deviate from one another, either in the direction of $X > Y$ (towards the right corner of the cube) or in the direction of $X < Y$ (towards the left corner). Figure 2B depicts an exemplary surface with an inverted U-shape on the LOIC (i.e., $a_4 < 0$) for which the ridge is shifted away from the LOC ($a_3 \neq 0$): Here the out-

come is maximal when Y is larger than X by a certain amount ("optimal margin model", e.g. Baumeister, 1989).

Commensurability and measurement invariance. The LOC refers to the numerical congruence of the predictor scales, and moving along the LOIC corresponds to increasing numerical differences in the one or the other direction. This numerical comparison only makes sense if both variables have been measured on the same scale. Just as it is a meaningless question whether 10 kilogram are larger than 8 meter, it does not make sense to do numerical comparisons on psychological constructs which have not been measured on the same scale.

When two measures are on the same scale they are *commensurable* (Edwards & Shipp, 2007). Commensurability consist of two aspects, namely (a) nominal equivalence, which requires that both scales measure the same latent construct, and (b) scale equivalence, which requires that both scales have the same metric. These aspects are typically addressed by using the same measurement scales for both partners. This, however, is only a necessary but not a sufficient condition for commensurability. Even if the same operationalizations are used in subgroups (such as males and females), the relevant latent construct is not necessarily equivalent. For example, subgroups might have a different understanding of the items, as has been extensively discussed in the literature on measurement invariance and differential item functioning (for an overview, see Meredith & Teresi, 2006). Meaningful numerical group comparisons require *strict measurement invariance*, which is established by an invariance of factor structure, factor loadings, item means, and residual variances. Only under these conditions comparative statement such as 'larger', 'smaller', or 'equal' are meaningful. Consequently, for the investigation of similarity hypotheses in dyads, researchers should ensure strict measurement equivalence across genders. Comparisons of *incommensurable* measures require special treatment and have to be interpreted with extreme caution (Schönbrodt, 2015).

A prototypical similarity pattern. In sum, the example surface in Figure 1 reflects a prototypical similarity effect of a female and male personality dimension on a single outcome variable. A couple's amount of time spent together is highest when the predictor combinations of the partners are identical. Furthermore, all levels of similarity are equally beneficial, and increasing dissimilarity leads to decreasing time spent together. Both directions of dissimilarity are equally detrimental. The statistical counterparts of this verbal description are $a_2 = 0$ (the surface is not curved above the LOC), $a_1 = 0$ (the predicted outcome above the LOC is constant, i.e., the ridge is flat; all levels of congruence are equally beneficial), $a_4 < 0$ (surface above the LOIC is curvilinear with an inverted U-shape), and $a_3 = 0$ (highest point of LOIC is above the LOC).

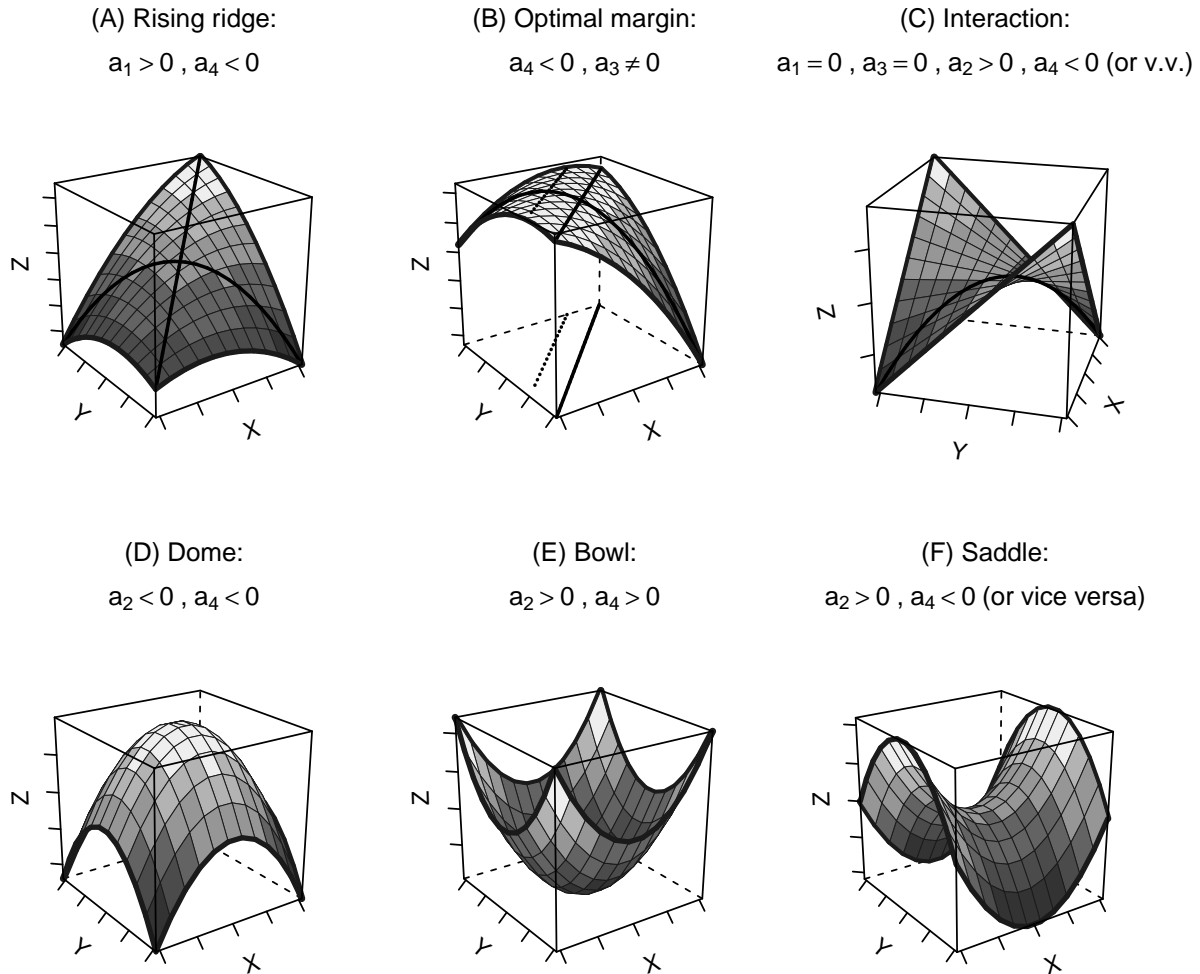Finally, a similarity effect requires that the ridge line of

(A) Rising ridge:

$a_1 > 0$, $a_4 < 0$

(B) Optimal margin:

$a_4 < 0$, $a_3 \neq 0$

(C) Interaction:

$a_1 = 0$, $a_3 = 0$, $a_2 > 0$, $a_4 < 0$ (or v.v.)

(D) Dome:

$a_2 < 0$, $a_4 < 0$

(E) Bowl:

$a_2 > 0$, $a_4 > 0$

(F) Saddle:

$a_2 > 0$, $a_4 < 0$ (or vice versa)

Figure 2. Multiple examples of RSA configurations. Figure available at https://osf.io/ftsrd/, under a CC-BY4.0 license.

the surface is positioned exactly on the LOC, as in Figure 1. This property is reflected in an additional condition $a_5 = b_3 - b_5 = 0$ (ridge line equals the LOC; see Appendix A for details on this condition). The *R* package *RSA* (Schönbrodt & Humberg, 2018), which we will use later in the tutorial section, provides point estimates, confidence intervals, and significance tests for all surface parameters.

Note that Figure 1 provides only one possible shape of a regression surface. Figure 2 shows some further surfaces, corresponding to different estimates of the coefficients in Equation 1. Response surfaces can have some typical shapes. For example, the surface in Figure 2A is linear above the LOC ($a_2 = 0$), but it differs from Figure 1 in that the surface rises above the LOC. This is indicated by a positive parameter $a_1 > 0$. One of the many advantages of RSA is that it enables to detect such rising ridge effects. Other typical shapes include a bowl (e.g., Figure 2E), a dome (e.g., Figure 2D), or

a saddle (e.g., Figure 2F).

In sum, RSA is an appropriate tool to test hypotheses which posit that the similarity of two predictor variables should explain variance in a single outcome variable. RSA is gaining momentum as a tool in the personality literature. For example, recent publications investigated the effects of personality similarity on romantic attraction at zero acquaintance (Olderbak, Malter, Wolf, Jones, & Figueredo, 2017), relationship intensity in a network analysis (Ilmarinen, Vainikainen, Verkasalo, & Lönnqvist, 2017), effects of parent–offspring personality similarity on externalizing problems (Franken, Laceulle, Van Aken, & Ormel, 2017), or person–group dissimilarity in personality on peer victimization (Boele, Sijtsema, Klimstra, Denissen, & Meeus, 2017) and self-esteem (Bleidorn et al., 2016). Detailed overviews of further potential applications of RSA can be found in Table 1 in Barranti, Carlson, and Côté (2017) and in Table 1 in

Humberg et al. (in press).

However, in the dyadic setting, we aim at predicting *two* interdependent outcome variables (e.g., female and male relationship satisfaction) from dyad similarity. We therefore need to extend RSA to the dyadic case. The core idea which we will use to achieve such an extension comes from the Actor-Partner-Interdependence Model (APIM). The APIM allows to model linear main effects of two predictors on two interdependent outcome variables. After a short introduction to the APIM, we will combine its reasoning with RSA, resulting in a tool to model similarity effects on the two dyadic outcome variables.

The Actor-Partner-Interdependence Model

Early treatments of dyadic data sometimes circumvented the interdependence of dyadic data by computing separate models for the dyad members (e.g., for women and for men). However, a more suitable approach is to simultaneously estimate the two models in a joint model, as it allows, for example, to directly test for gender differences (Kenny et al., 2006). The model contains two dyad members (e.g., women and men) and two variables (e.g., a personality trait such as agreeableness and an outcome such as relationship satisfaction). By convention, the focal person is called 'actor' and the other person in the dyad is called 'partner'. The basic idea of the APIM is that an actor's dependent variable is not only a function of his or her own characteristics, but also a function of the characteristics of the partner. In that way the APIM captures one key aspect of social relationships: the interdependency of partners.

From a statistical perspective, the APIM is a multivariate regression model (cf. Kenny et al., 2006; Nestler et al., 2015) in which the outcome variable of the actors and of the partners (e.g., relationship satisfaction of women and men) is regressed on the predictor variable of both partners (e.g., women's and men's personality traits):

$$Z_f = b_{0f} + b_{1f}X + b_{2f}Y + e_f$$
$$Z_m = b_{0m} + b_{1m}X + b_{2m}Y + e_m \qquad (2)$$
$$e_f \sim\sim e_m$$

where Z_f and Z_m denote the outcome variable of women ('f') and men ('m'), X refers to the female predictor, and Y to the male predictor.[1] The coefficients b_{1f} and b_{2m} are called *actor effects*: They describe the influence of a dyad member's predictor (e.g., agreeableness) on her or his own outcome variable (e.g., relationship satisfaction). Actor effects thus indicate intrapersonal effects. The coefficients b_{2f} and b_{1m} are the *partner effects*. They measure the influence that the dyad member's predictor has on the other member's outcome. They thus reflect interpersonal effects. In the example, the partner effect describes the effect of the man's (or the

woman's) agreeableness on the woman's (or the man's) relationship satisfaction. Finally, the residual scores of the two outcome variables are allowed to be correlated ($e_f \sim\sim e_m$) to account for dependence in the dyadic data that is not explained by the modeled predictor variables.

The goal of employing the APIM is to estimate the magnitude of the actor and partner effects. In case of distinguishable dyads, researchers can use structural equation modeling (SEM) software for this purpose (Kenny et al., 2006), as the standard APIM is mathematically equivalent to a cross-lag path model with two time-points (Nestler et al., 2015). Alternatively, the APIM for distinguishable dyads can also be estimated using a double-intercept multilevel model (Kenny et al., 2006). In the case of indistinguishable dyads, again both multilevel modeling software or SEM software can be used (see Kenny et al., 2006; Olsen & Kenny, 2006). One important advantage of the SEM approach is that such software allows defining constraints on the model parameters. Such constraints can be used to examine APIM patterns, such as the actor-only pattern (see below). This is important insofar as each APIM pattern suggests a very different process occurring in the dyadic relationship (see Kenny & Ledermann, 2010). Therefore, we focus on the SEM framework to estimate the dyadic RSA model which we propose below.

Dyadic Response Surface Analysis

We shortly reflected on two statistical approaches here: RSA enables to test similarity effects of two predictors on a single outcome variable. APIM allows to test simple main effects of two predictors on two interdependent (dyadic) outcome variables. It is now straightforward to combine these two approaches to obtain a method to test similarity effects on two interdependent outcomes (Nestler et al., 2015).

Formally, the DRSA is defined by two polynomial regressions, where the same pair of predictor variables (and their higher terms) predict the male and the female outcome variable. To be consistent with previous literature on RSA, we use X, Y, and Z as variable names. X denotes the female predictor variable, Y the male predictor variable, Z_f and Z_m the female and male outcome variables, respectively. The error terms e_f and e_m are correlated to account for the non-independence of the dyadic data structure:

$$Z_f = b_{0f} + b_{1f}X + b_{2f}Y + b_{3f}X^2 + b_{4f}XY + b_{5f}Y^2 + e_f$$
$$Z_m = b_{0m} + b_{1m}X + b_{2m}Y + b_{3m}X^2 + b_{4m}XY + b_{5m}Y^2 + e_m$$
$$e_f \sim\sim e_m$$

$$(3)$$

---

[1]Here we deviate from the typical APIM notation that uses Y_f and Y_m for the dependent variables and X_f and X_m for the predictor variables. We changed the notation to achieve continuity to the dyadic RSA extension explained below.

Within this model, one distinguishes actor effects and partner effects, depending on whether a regression path connects a predictor and an outcome variable belonging to one person (actor effects) or whether it connects a predictor of one person to the outcome variable of the other persons (partner effects; see also Figure 3). Hence, b_{1f}, b_{3f}, b_{2m}, and b_{5m} are actor effects, and b_{2f}, b_{5f}, b_{1m}, and b_{3m} are partner effects. Furthermore, the terms b_{4f} and b_{4m} are statistical actor-partner-interactions, as they combine information both from the actor and from the partner[2].

As two outcome variables are modeled, also two different response surfaces can be plotted, using b_{0f} to b_{5f} for the female surface and b_{0m} to b_{5m} for the male surface. Likewise, the surface parameters for LOC and LOIC can be computed both for the female and the male surface. The female surface parameters are defined as:

$$a_{1f} = b_{1f} + b_{2f}$$
$$a_{2f} = b_{3f} + b_{4f} + b_{5f}$$
$$a_{3f} = b_{1f} - b_{2f} \quad (4)$$
$$a_{4f} = b_{3f} - b_{4f} + b_{5f}$$
$$a_{5f} = b_{3f} - b_{5f}$$

The male surface parameters are similarly defined, only using the coefficients with the 'm' subscript.

Centering. When higher terms are present in a regression model, such as the interaction term or the squared terms in a polynomial regression, it is advisable to center the predictor variables such that they obtain a meaningful zero point. This enhances the interpretability of the linear main effects (b_{1f}, b_{1m}, b_{2f}, and b_{2m}) and of the surface parameters that reflect the slope of the LOC or the LOIC above $(0,0)$ (a_{1f}, a_{1m}, a_{3f}, and a_{3m}).

In DRSA, it is furthermore essential to center both partners' predictor variables so that they have a common zero point (Kenny & Cook, 1999; Kenny et al., 2006). This can for example be achieved by subtracting the midpoint of the response scale (e.g., 4 on a 1–7 scale ranging from *Strongly Disagree [1] – Undecided [4] – Strongly Agree [7]*) from X and Y. This is often appropriate when the midpoint of the scale is semantically meaningful. If the response scale has no meaningful midpoint, it is also possible to center X and Y to the mean of the variables computed across both genders. If one wants to additionally standardize the two predictor variables, this should be also done by use of the common standard deviation across both genders.

This kind of common centering across genders is necessary for two reasons. First, commensurability would be lost when females and males were centered or standardized within their groups. With incommensurable scales, the LOC and the LOIC are no longer interpretable. Second, a common scale is necessary to make comparisons of the effect sizes between genders. Note that, although it might seem like a valid

alternative at first sight, one cannot center X and Y within dyads, because this would force a couple's predictors onto the same absolute value and make them perfectly collinear.

Constraining the Model. In a full DRSA model, 10 directed path coefficients are estimated (along with the (co)variances), which can be quite a large number depending on the sample size. Both for parsimony and for testing substantive theoretical predictions, constraints can be applied that simplify the full model.

Constraining path coefficients. Path coefficients **within a surface** can be constrained based on multiple principles. From a dyadic data analysis point of view, one can test whether some path coefficients do (not) differ from zero, and whether some path coefficients are equal for both genders. Kenny and Cook (1999) distinguish several patterns of actor effects and partner effects which reflect meaningful psychological phenomena:

actor-oriented: actor effects $\neq 0$, partner effects $= 0$

partner-oriented: actor effects $= 0$, partner effects $\neq 0$

couple-oriented: actor effects $=$ partner effects

social comparison: actor effects $= -$partner effects

Actor-oriented patterns assume that a person is only affected by his or her own predictor score, but not by the partner's predictor score. Partner-oriented patterns assume that a person's outcome is only affected by the partner's score, but not by his or her own score. A couple-oriented pattern implies that the sum (or the average) of the dyad's predictors determines the outcome. This means, that the own and the partner's standing are equally important for the outcome, and that own sufficiencies can be compensated by the partner. A social comparison pattern, in contrast, implies that the outcome increases when the *difference* between the own and the partner's score increases. **Note that these four constraints are applied *within* the male, respectively the female, surface, and do not equate parameters *between* the male and the female surface. Hence, it is technically possible that the female surface follows, for example, an actor-oriented pattern (i.e., b_{1f} or $b_{3f} \neq 0$, $b_{2f} = b_{4f} = b_{5f} = 0$) while the male surface follows a partner-oriented pattern (i.e., b_{1m} or $b_{3m} \neq 0$, $b_{2m} = b_{4m} = b_{5m} = 0$), or any other combination of patterns. Additional between-role constraints can be added to test whether both roles follow the same within-role constraints (e.g., whether both roles follow an couple-oriented pattern with the same parameters).**

These classical APIM constraints only refer to the linear main effects. In the extended polynomial regressions of the

_____

[2]Note that the multiplicative interaction is only one of several possible operationalizations of actor-partner-interactions (Kenny & Cook, 1999).

Female response surface
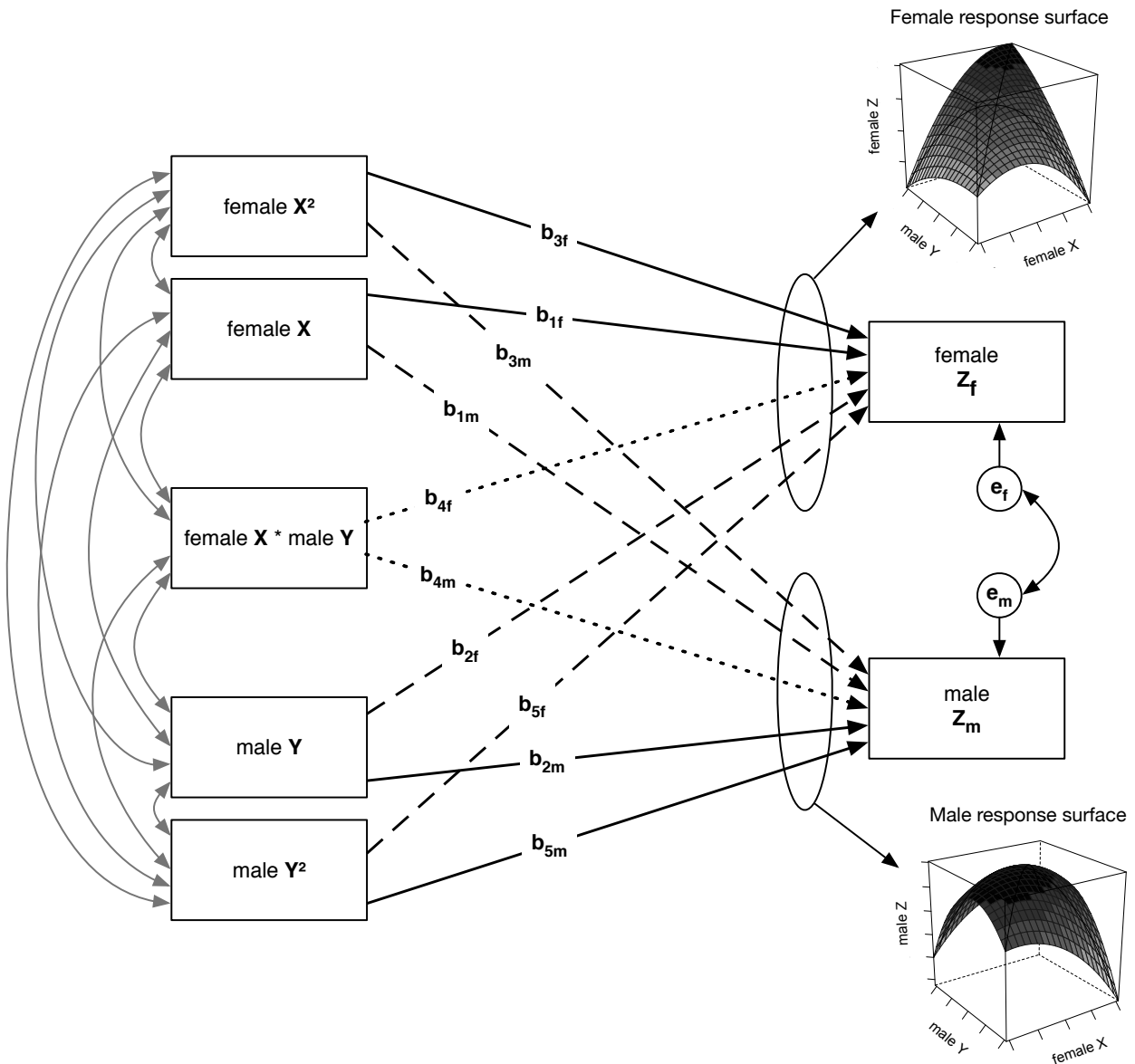

Male response surface

Figure 3. The dyadic RSA path model. Black solid paths are actor effects, dashed paths are partner effects, and the dotted paths are statistical partner interactions. Intercepts are not displayed. Figure available at https://osf.io/ftsrd/, under a CC-BY4.0 license.

DRSA, further constraints have to be defined for the higher order terms. For example, for a pure actor-oriented pattern, additionally all partner effects and the interaction terms should be zero, whereas the higher order actor effects could be unequal to zero.

Additionally to the APIM constraints, one can also apply constraints typically used in RSA. For example, the prototypical similarity effects surface displayed in Figure 1 implies several constraints on the regression coefficients, as it assumes that the surface has an inverted U-shape with the ridge being flat and lying exactly above the LOC (i.e., the outcome is maximized at perfect congruence and the level of congruence is irrelevant). In the non-dyadic case, these assumptions about the surface are achieved with the following constraints: (1) $b_1 = 0$, (2) $b_2 = 0$, (3) $b_3 = b_5$, and (4) $b_4 = -2b_5$. This model is also called 'squared difference model', as a simple squared difference as predictor implies these constraints (Edwards, 2002; Schönbrodt, 2015). If theory implies that

the outcome for both genders in a DRSA should follow a squared difference model, analogous constraints have to be applied: (1) $b_{1f} = b_{1m} = 0$, (2) $b_{2f} = b_{2m} = 0$, (3) $b_{3f} = b_{5f}$, (4) $b_{3m} = b_{5m}$, (5) $b_{4f} = -2b_{5f}$, and (6) $b_{4m} = -2b_{5m}$. If one furthermore assumes that the response surface is the same for both genders, the free coefficients have to be equated between genders: (1) $b_{1f} = b_{2f} = b_{1m} = b_{2m} = 0$, (2) $b_{3f} = b_{5f} = b_{3m} = b_{5m}$, and (3) $b_{4f} = b_{4m} = -2b_{5f}$. Additional theoretically meaningful models, such as the rising ridge model, can be tested by applying respective other constraints to the regression coefficients (Schönbrodt, 2015).

Nested and non-nested model comparison. By constraining parameters, model complexity can be reduced compared to estimating every single coefficient in the model. A simpler model is said to be nested under a more complex model if the simpler model can be obtained from the complex model by adding parameter constraints. For example, the actor-oriented pattern is nested within the unconstrained APIM, as it can be defined by constraining all partner effects to be zero.

More complex models always yield a better fit to the data in terms of the amount of variance they can explain, but they contain the risk of overfitting to the data at hand. Complex models therefore threaten the goals of parsimony and generalizability to other data. Nested models can be compared via χ^2 likelihood ratio (LR) tests. If a simpler model is not significantly worse, indicated by a non-significant *p*-value in the LR test, the simpler model can be retained.

Not all models, however, are nested. For example, the partner-oriented pattern is not nested under the actor-oriented pattern (or vice versa) – there is no way to add constraints to the actor-oriented pattern to get to the partner-oriented pattern (in contrast, one would have to *release* a constraint to allow non-zero partner effects). Consequently, the performance of actor-oriented and partner-oriented patterns cannot directly be compared to each other via LR tests. For the comparison of non-nested models, one can instead use information criteria. One of the most commonly used is the Akaike Information Criterion (AIC; for an overview see Burnham & Anderson, 2002). Models with a relatively lower AIC value show a better fit to the data (the absolute value of AIC is irrelevant), and ΔAIC values ≥ 7 typically are seen as an indicator to prefer the model with the lower AIC. ΔAICs values ≤ 2, in contrast, indicate that both models are essentially equally good and the data at hand does not contain enough evidence to decide between both models (Burnham, Anderson, & Huyvaert, 2011; Symonds & Moussalli, 2011). ΔAIC values between 2 and 7 can be seen as some support for the better model, that is not decisive, however. Other information criteria beyond AIC are also available, such as BIC (Burnham & Anderson, 2004), or variations of the AIC (e.g., AICc or CAIC; for an overview and comparison, see Dziak, Coffman, Lanza, & Li, 2012, Garamszegi, 2011).

Why apply constraints? There are several reasons for testing constraints on the full model. First, in an exploratory analysis, the patterns of APIM constraints and of the RSA constraints can be used to give post-hoc descriptions to the results of an analysis (e.g., "The data follows a partner-oriented pattern"). Second, constrained models can be understood as confirmatory hypotheses. To test such a hypothesis, one would test the theoretically predicted model with constraints against an unconstrained model. If the constrained model is not significantly worse than the unconstrained model, it is retained as a more parsimonious model and the theory is corroborated. Finally, as a pragmatic reason to apply constraints, they also increase power. If men's and women's effects are not significantly different from each other, the respective paths can be set equal. These constraints effectively increase the sample size and the power to detect smaller effects. However, as always is the case with a non-rejected null hypothesis, a non-significant difference of paths on its own is no direct evidence of 'no difference' (e.g., Goodman, 2008), but should rather be interpreted as 'not enough power to detect a potential difference', depending on the sample size and the smallest effect size of interest.

Dyadic RSA for indistinguishable dyads. So far we focused only on noninterchangeable dyads, but the procedure can also be adapted to dyads that are naturally interchangeable, following the approach of Olsen and Kenny (2006). Furthermore, even if dyads have a potentially distinguishing variable, such as gender, their means, variances, and covariances could be statistically indistinguishable. The omnibus SEM test for distinguishability tests for these equalities (see Gonzalez & Griffin, 1999, and Kenny et al., 2006, pp. 129-131).

If dyads are naturally interchangeable, several constraints must be applied to the SEM (Olsen & Kenny, 2006): Equal (linear and squared) actor effects, equal (linear and squared) partner effects, equal predictor means, equal predictor variances, equal outcome intercepts, and equal residual variances. The accompanying *R* code contains an example how to set up such a model. In the remainder of the paper, however, we focus on distinguishable dyads.

Power analysis / Sample size planning. Planning for appropriate sample sizes is a necessary step to draw strong inferences from a data set. Doing a power analysis for models with multiple predictors and correlated outcomes, however, is not trivial, as it requires to define a priori several (co)variance components and multiple effect sizes which themselves depend on the reliability of the measured variables and their higher terms, many of which might be hard to estimate. An in-depth treatment of this topic is beyond the scope of this paper, but we want to outline some rules of thumb. A general recommendation for higher order terms provided by Aiken and West (1991) is to assess 2-3 times as many participants as one would need to detect linear effects

of the predictors. Another possibility for reducing complexity would be to collapse multiple parameter estimates into a single index of effect size, namely the increase in explained variance ΔR^2 of 5%. This could relate, for example, to the increase in explained variance that a congruence effect (i.e., a negative a_4) explains beyond the two main effects. For example, to detect a ΔR^2 of 5% with a statistical power of 90% ($\alpha = 5\%$) in a non-dyadic RSA, around 200 participants are needed if a single parameter, such as a_4 in a squared difference model, explains the additional variance. The impact of the dyadic nature of the data depends on the intra-class-correlation of the outcome variable, where the effective sample size is somewhere between the number of individuals and the number of dyads. If actor and partner effects can be constrained to be equal across roles, this effectively increases the sample size. Beyond these simple approximations, we suggest that the ideal way to determine a reasonable sample size is a simulation study (Nestler et al., 2015) which incorporates as much prior knowledge about plausible effect sizes as possible. We provide an example of such a power simulation in the code examples, but note that this requires to fully specify an assumed population model with all (co)variances and path coefficients.

Example and R Code Walkthrough

In the following sections, we will consider practical questions on how to conduct a DRSA in *R* and demonstrate the usage with a concrete example. For a clearer demonstration, we will use simulated data. In the tradition of Raymond Cattell, we decided to give an artificial name to our hypothetical personality scale ("bavaria"), in order not to confuse readers about the purported implications of this simulated data set. The full code for reproducing the results and the figures can be downloaded from https://osf.io/ftsrd/.

Data Preparation

Before computing the model, several steps of data preparation should be done.

Data format. The SEM framework requires data to be stored in the wide format, so that each row contains data from one dyad and each measured variable has a separate column for each dyad member. This data format clarifies that the unit of analysis is the dyad, and that degrees of freedom usually refer to the number of dyads. Table 1 shows an example of the required data structure.

Missing values. Missing values are a potential source for bias and distorted standard errors. In the case of RSA, a single missing value in, X, for example, propagates further to missing values in X^2 and in XY. An in depth treatment of missing values is beyond the scope of this paper, but we mention that the 'default' option of listwise deletion usually is not recommended. An often recommended technique is to use a maximum likelihood estimation based on all available

Table 1

The First Rows of a Demo Data Set in Wide Format.

pairID	X	Y	Z_f	Z_m
1	8	8	7	2
2	7	8	5	4
3	6	7	5	2
4	5	2	5	5
5	5	8	5	6
6	5	4	8	6

Note. X = female predictor variable, Y = male predictor variable, Z_f = female outcome variable, Z_m = male outcome variable.

data ("full information maximum likelihood", FIML). This estimation assumes that missing values are at least *missing at random* ("MAR"; Schafer & Graham, 2002), which means that the probabilities of missingness do not depend on the missing data. The question whether missing values are MAR or not, can typically not be tested but has to be assumed. However, Collins, Schafer, and Kam (2001) demonstrated that violations of the MAR assumption often only have minor impact on the estimation and standard errors. The *lavaan* package which we recommend here is capable of a FIML estimation and we will apply this treatment of missing values in our example below.

Amount of predictor discrepancy. A potential challenge with couple data is that partner correlations of predictor variables are often high, leading to a bivariate range restriction where the incongruent quadrants are much less populated than the congruent quadrants. This leads both to a diminished power to detect discrepancy effects (McClelland & Judd, 1993), but also to a heightened susceptibility to extreme values, as few discrepant values can drive an apparent discrepancy effect. Shanock et al. (2010) suggest to compute a standardized discrepancy score of the predictor variables to determine the percentage of couples that can be considered discrepant. For this purpose, predictor scores are standardized within gender, and any absolute difference larger than 0.5 *z*-points between male and female scores is considered a "discrepant couple". The `RSA` function from the *RSA* package automatically reports the percentage of discrepant units according to these criteria. If only very few couples are discrepant, it might be futile to assess the impact of dissimilarity on an outcome.

Outliers. Regression results can be spurious if they are driven by a small number of outliers, and the squared terms of the polynomial regression even exaggerate the impact of outliers. Therefore, we recommend to screen the data set for multivariate outliers. Typical outlier detection is based on indices such as Cook's distance or leverage points (e.g., Bollen & Jackman, 1985). The `RSA` function from the *RSA* pack-

age automatically checks for multivariate outliers according to these criteria (see also the example R script, where the data set is screened for outliers). A sensitivity analysis that explores the effect of excluding some of the most discrepant outliers can show whether an apparent similarity effect is robust, or whether it is only driven by a small number of extreme values.

Higher terms computation. As a final step, the higher terms for the polynomial regression have to be added to the data set. For that purpose, the centered predictors (referred to as *X.c* and *Y.c*) have to be squared (referred to as *X.c*2 and *Y.c*2) and their multiplicative interaction has to be computed (referred to as *XY*)

Computing the Model in an SEM Framework

We compute the DRSA models with a path modeling approach using the *lavaan* package (Rosseel, 2012) for the R statistical environment (R Core Team, 2014). Note that the same model can also be estimated in a multilevel model, when the classical APIM model for distinguishable dyads (Kenny et al., 2006) is extended by the multiplicative actor-partner interaction and the squared terms for the actor and the partner effect.[3]

In the *lavaan* syntax, the model is defined as follows:

```
dRSA.full.model <- "
Z_f ~ b1f*X.c + b2f*Y.c + b3f*X.c2 + b4f*XY
    + b5f*Y.c2
Z_m ~ b1m*X.c + b2m*Y.c + b3m*X.c2 + b4m*XY
    + b5m*Y.c2
Z_m ~~ Z_f
"
```

The '~' operator defines regression paths, the '~~' operator defines residual correlations. This model can be estimated using the `sem` function. The standard maximum likelihood estimator of SEM packages assumes a multivariate normal distribution of the endogenous variables. As the estimation of the standard errors and *p*-values can be biased when this assumption is violated, we recommend to use a robust estimator (`estimator="MLR"`) for the model development stages, which yields less biased standard errors (`se="robust"`) with non-normal data. Furthermore, we request a FIML treatment of missing data:

```
s.full <- sem(dRSA.full.model, data=df,
    meanstructure=TRUE, estimator="MLR",
    se="robust", missing="fiml")
```

For the final model, we recommend to compute bootstrapped standard errors and *p*-values, with at least 5000 replications. This requires switching to the ML estimator:

```
s.full.boot <- sem(dRSA.full.model,
    data=df, meanstructure=TRUE,
    estimator="ML", missing="fiml",
    se="boot", bootstrap=10000)
```

With the `summary(s.full.boot)` command the parameter estimates can be printed (see Table 2). The unconstrained DRSA model is a saturated model with zero degrees of freedom. Hence, no tests of model fit are possible.

Plotting the Model

The joint impact from the multitude of model parameters in Table 2 on both partners' outcome variables can be difficult to interpret. Furthermore, we caution against the attempt to interpret single coefficients in isolation. For a more intuitive interpretation of the joint impact of all predictors we recommend to visualize the final models using the *RSA* package (Schönbrodt & Humberg, 2018; see the online material for the code for the plots).[4] Figure 4 shows a side-by-side comparison of the female and the male surface. The visualization of both partners' response surfaces illustrates that the (dis)similarity between partners can have different effects on the female and the male outcome.

In the current hypothetical example, females show an optimal margin pattern, where the highest female relationship satisfaction is expected when the female *bavaria* is somewhat higher than the male *bavaria*. This general pattern, however, includes a rising ridge, which means that optimal predictor combinations on a higher level are better than optimal combinations on lower levels. The male predicted relationship satisfaction, by contrast, is highest when the male partner is high in *bavaria* while the female is low, medium when both are on similar levels, and lowest in the incongruent corner of high female/low male *bavaria*.

Generally, extreme caution is necessary when the regression surface is extrapolated to regions without actual data points. As this extrapolation rests on very unlikely assumptions (Montgomery, Peck, & Vining, 2012), only regions of the surface within the range of the original data should be interpreted. To facilitate a valid interpretation, we strongly recommend to show the raw data imposed on the surface plot (see also Wilkinson & Task Force on Statistical Inference, 1999; Tufte, 2001). In Figure 4, a bagplot around the raw data points is shown to give a visual aid for the 'interpretable region'. The bagplot is a bivariate extension of a boxplot (Rousseeuw, Ruts, & Tukey, 1999), which describes the position of the inner 50% of points (within the inner polygon, called *bag*) and the outer 50% of points (within the outer

---

[3]In the associated OSF project (https://osf.io/ftsrd/), we also provide Mplus scripts that do the same job. Furthermore, we provide example code how the same model can be estimated in a multilevel modeling framework.

[4]Surface plots can be also created (with limited graphical options and limited visual refinement) using Excel sheets, see for example http://www.springer.com/cda/content/document/cda_downloaddocument/Excel+spreadsheet+for+response+surface+analysis.xls?SGWID=0-0-45-940137-p35536793.

Table 2

Regression Coefficients b_1 to b_5 and Derived Model Parameters for the DRSA Model, With Bootstrapped Standard Errors and p-Values

|	Estimate	95% CI	Standard error	p value
b_{1f}	0.298	[0.21; 0.39]	0.047	0.000
b_{2f}	−0.078	[-0.17; 0.00]	0.043	0.074
b_{3f}	−0.044	[-0.08; -0.01]	0.018	0.016
b_{4f}	0.162	[0.11; 0.21]	0.026	0.000
b_{5f}	−0.088	[-0.11; -0.06]	0.015	0.000
b_{1m}	−0.321	[-0.50; -0.14]	0.092	0.000
b_{2m}	0.481	[0.30; 0.64]	0.088	0.000
b_{3m}	0.020	[-0.05; 0.10]	0.036	0.573
b_{4m}	0.042	[-0.05; 0.14]	0.050	0.395
b_{5m}	−0.006	[-0.06; 0.06]	0.032	0.841
a_{1f}	0.221	[0.09; 0.34]	0.063	0.000
a_{2f}	0.029	[-0.03; 0.10]	0.033	0.380
a_{3f}	0.376	[0.26; 0.51]	0.065	0.000
a_{4f}	−0.295	[-0.37; -0.23]	0.037	0.000
a_{5f}	0.044	[-0.00; 0.09]	0.023	0.051
a_{1m}	0.160	[-0.10; 0.40]	0.129	0.214
a_{2m}	0.056	[-0.07; 0.21]	0.071	0.430
a_{3m}	−0.802	[-1.03; -0.55]	0.126	0.000
a_{4m}	−0.028	[-0.16; 0.10]	0.068	0.676
a_{5m}	0.027	[-0.07; 0.12]	0.047	0.569
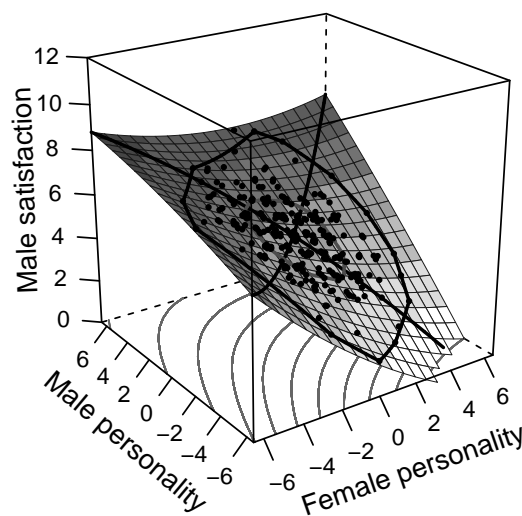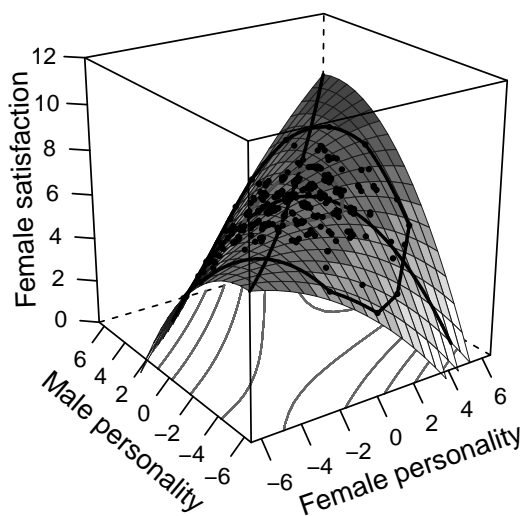
Figure 4. Female and male response surface from the DRSA. Raw data points are plotted at their predicted value (i.e., they are projected onto the response surface). Figure available at https://osf.io/ftsrd/, under a CC-BY4.0 license.

polygon, called *fence*).[5]

---

[5]In the gray scale plots the inner bag is hard to detect; the dark polygon is the fence that marks the bivariate range of the data.

Applying Constraints

Furthermore, we wanted to check whether the model can be constrained in a way that allows to use an identical surface as an adequate model for both genders' outcomes. Therefore we applied the following constraints:

```
# actor effect equality constraints
b1f == b2m
b3f == b5m

# partner effect equality constraints
b2f == b1m
b5f == b3m

# interaction effect equality constraint
b4f == b4m
```

These constraints force the surface to be identical in shape for both genders, except for a different elevation (i.e., the intercept is free to vary between genders). A LR model comparison shows that the constrained model is significantly worse than the full model, $\chi^2(5) = 34.2$, $p < .001$. The same conclusion can be drawn from a AIC difference of 24, also favoring the full model. Hence, we reject the constrained model and conclude that the joint impact of the personality predictors is different for each gender.

Conclusion

This paper presents a general framework for testing and interpreting similarity effect hypotheses in dyads. Based on previous work on polynomial regression and response surface analysis, we extended the model to dyadic designs. The DRSA model allows a direct test of similarity effect hypotheses and avoids the pitfalls of previous approaches. A particular strength of the approach is the fact that new classes of hypotheses beyond basic similarity effect hypotheses can be tested which would be hard or impossible to test with conventional methods. For example, previous approaches using difference scores implicitly assume (and cannot test) that similarity effects are identical on all levels of similarity. For many domains of research questions, however, it is reasonable to assume that a high/high combination of the predictors is more beneficial than a low/low combination. Rising ridge models outlined here and covered in detail elsewhere (Schönbrodt, 2015) are in particular suitable to model and test these interesting hypotheses. DRSA is also able to model binary outcome variables in a generalized linear model. In this case, a probit link function is used to model the probability for a positive outcome (see the code in the online material for an example on modeling and plotting binary outcomes).

The DRSA approach has many advantages compared to previous approaches, but users should be aware of some limitations. First, a large number of model parameters is estimated. From a hypothesis testing point of view, this also involves a multitude of significance tests. Depending on the type of hypothesis tested, corrections such as the Bonferroni correction (Bonferroni, 1935) or more powerful alternatives, such as the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995), might be necessary to ensure the nominal error rate. This also emphasizes the necessity of ensuring an adequate power when planning a study. Second, the manifest path models assume that variables have been measured without error. Latent models of all variables are possible; however, there are currently several ways and no general consensus on how to model and estimate latent interactions (e.g., Harring, Weiss, & Hsu, 2012). Finally, the approach that was introduced here can be used to investigate how the similarity on a single dimension relates to an outcome; it does not cover similarity across multiple predictor dimensions.[6]

To conclude, we hope that the conceptual DRSA framework and the provided R scripts assist substantive researchers in precisely formulating, testing, and interpreting similarity effect hypotheses in their dyadic research.

References

Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage Publications, Inc.

Barranti, M., Carlson, E. N., & Côté, S. (2017). How to test questions about similarity in personality and social psychology research: Description and empirical demonstration of response surface analysis. *Social Psychological and Personality Science*, *8*, 465–475. doi:10.1177/1948550617698204

Baumeister, R. F. (1989). The optimal margin of illusion. *Journal of Social and Clinical Psychology*, *8*, 176–189.

Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*, 289–300. JSTOR: 2346101

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, *42*, 192–212.

---

[6]See Edwards (1994) for a generalization of profile similarity indices to polynomial regression. This approach, however, requires to expand the linear model with a second-order polynomial for each predictor dimension. For example, reformulating a sum-of-squares profile discrepancy score of the Big-5 personality traits as an unconstrained model requires 5*5 = 25 terms instead of a single discrepancy score.

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. doi:10.1177/0956797615627133

Boele, S., Sijtsema, J. J., Klimstra, T. A., Denissen, J. J., & Meeus, W. H. (2017). Person–Group dissimilarity in personality and peer victimization. *European Journal of Personality*, *31*, 220–233. doi:10.1002/per.2105

Bollen, K. A. & Jackman, R. W. (1985). Regression diagnostics: An expository treatment of outliers and influential cases. *Sociological Methods & Research*, *13*, 510–542. doi:10.1177/0049124185013004004

Bonferroni, C. E. (1935). *Il calcolo delle assicurazioni su gruppi di teste*. Tipografia del Senato.

Burnham, K. P. & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer Science & Business Media.

Burnham, K. P. & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, *33*, 261–304.

Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, *65*, 23–35. doi:10.1007/s00265-010-1029-6

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, *6*, 330–351. pmid: 11778676

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

Cronbach, L. J. & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin*, *50*, 456–473.

Dyrenforth, P. S., Kashy, D. A., Donnellan, M. B., & Lucas, R. E. (2010). Predicting relationship and life satisfaction from personality in nationally representative samples from three countries: The relative importance of actor, partner, and similarity effects. *Journal of Personality and Social Psychology*, *99*, 690–702.

Dziak, J. J., Coffman, D. L., Lanza, S. T., & Li, R. (2012). *Sensitivity and specificity of information criteria*. Retrieved from https://methodology.psu.edu/media/techreports/12-119.pdf

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: 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. & Shipp, A. J. (2007). The relationship between person-environment fit and outcomes: An integrative theoretical framework. In C. Ostroff & T. A. Judge (Eds.), *Perspectives on organizational fit.* (pp. 209–258). The organizational frontiers series. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and a proposed alternative. *Organizational Behavior and Human Decision Processes*, *58*, 51–100. Retrieved February 3, 2017, from http://econpapers.repec.org/article/eeejobhdp/v_3a58_3ay_3a1994_3ai_3a1_3ap_3a51-100.htm

Edwards, J. R. (2007). Polynomial regression and response surface methodology. In C. Ostroff & T. A. Judge (Eds.), *Perspectives on organizational fit* (pp. 361–372). San Francisco: Jossey-Bass.

Franken, A., Laceulle, O. M., Van Aken, M. A., & Ormel, J. (2017). Using response surface analysis to interpret the impact of Parent–Offspring personality similarity on adolescent externalizing problems. *European Journal of Personality*, *31*, 104–117. doi:10.1002/per.2088

Garamszegi, L. G. (2011). Information-theoretic approaches to statistical analysis in behavioural ecology: An introduction. *Behavioral Ecology and Sociobiology*, *65*, 1–11. doi:10.1007/s00265-010-1028-7

Gonzalez, R. & Griffin, D. (1999). The correlational analysis of dyad-level data in the distinguishable case. *Personal Relationships*, *6*, 449–469. doi:10.1111/j.1475-6811.1999.tb00203.x

Goodman, S. (2008). A dirty dozen: Twelve p-value misconceptions. In *Seminars in Hematology* (Vol. 45, pp. 135–140). Elsevier. Retrieved February 8, 2017, from http://www.sciencedirect.com/science/article/pii/S0037196308000620

Harring, J. R., Weiss, B. A., & Hsu, J. C. (2012). A comparison of methods for estimating quadratic effects in nonlinear structural equation models. *Psychological Methods*, *17*, 193–214. doi:10.1037/a0027539

Humberg, S., Nestler, S., & Back, M. D. (in press). Response surface analysis in personality and social psychology: Checklist and clarifications for the case of congruence hypotheses. *Social Psychological and Personality Science*. Retrieved from osf.io/mhf2y

Ilmarinen, V.-J., Vainikainen, M.-P., Verkasalo, M. J., & Lönnqvist, J.-E. (2017). Homophilous friendship assortment based on personality traits and cognitive ability in middle childhood: The moderating effect of peer

network size. *European Journal of Personality*, *31*, 208–219. doi:10.1002/per.2095

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance*, *27*, 443–463.

Kenny, D. A. & Acitelli, L. K. (1994). Measuring similarity in couples. *Journal of Family Psychology*, *8*, 417–431.

Kenny, D. A. & Cook, W. (1999). Partner effects in relationship research: Conceptual issues, analytic difficulties, and illustrations. *Personal Relationships*, *6*, 433–448.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.

Kenny, D. A. & Ledermann, T. (2010). Detecting, measuring, and testing dyadic patterns in the actor–partner interdependence model. *Journal of Family Psychology*, *24*, 359–366.

McClelland, G. H. & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390. doi:10.1037/0033-2909.114.2.376

Meredith, W. & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*, S69–S77.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis*. John Wiley & Sons.

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.

Olderbak, S. G., Malter, F., Wolf, P. S. A., Jones, D. N., & Figueredo, A. J. (2017). Predicting romantic interest at zero acquaintance: Evidence of sex differences in trait perception but not in predictors of interest. *European Journal of Personality*, *31*, 42–62. doi:10.1002/per.2087

Olsen, J. A. & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, *11*, 127–141. doi:10.1037/1082-989X.11.2.127

Peter, J. P., Churchill, G. A., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, *19*, 655–662.

R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from http://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, *53*, 382–387.

Schafer, J. L. & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*, 147–177. pmid:12090408

Schönbrodt, F. D. (2015). *Testing fit patterns with polynomial regression models*. Ludwig-Maximilians-Universität München. Munich, Germany. Retrieved March 23, 2016, from https://osf.io/ndggf

Schönbrodt, F. D. & Humberg, S. (2018). *RSA: An R package for response surface analysis (version 0.9.12)*. Retrieved from https://cran.r-project.org/package=RSA

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. doi:10.1007/s10869-010-9183-4

Strauss, J. P., Barrick, M. R., & Connerley, M. L. (2001). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, *74*, 637–657.

Symonds, M. R. E. & Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology*, *65*, 13–21. doi:10.1007/s00265-010-1037-6

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd edition). Connecticut: Graphics Press.

Weidmann, R., Schönbrodt, F. D., Ledermann, T., & Grob, A. (2017). Concurrent and longitudinal dyadic polynomial regression analyses of big five traits and relationship satisfaction: Does similarity matter? *Journal of Research in Personality*. doi:10.1016/j.jrp.2017.04.003

Wilkinson, L. & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, *54*, 594–604. doi:10.1037/0003-066X.54.8.594

Appendix A

To determine whether a response surface reflects a similarity effect, one needs to test the four conditions on a_1 to a_4 (i.e., $a_1 = a_2 = a_3 = 0$ and $a_4 < 0$), and in addition, one needs to find out whether the ridge line ("first principal axis", in mathematical terms) of the surface is positioned at the LOC. The position of the ridge line can in general be determined by considering its projection onto the X-Y plane (e.g., the dotted line on the bottom of the coordinate cube in Figure 2B). When one understands the X-Y plane as a coordinate system in itself, one can express the position of the ridge as a linear equation that relates Y to X. Its formula is (see Edwards, 2007)

$$Y = p_{10} + p_{11}X, \tag{5}$$

where its coefficients p_{10} and p_{11} can be computed from the regression estimates b_1 to b_5 via the formulas

$$p_{11} = \frac{(b_5 - b_3) + \sqrt{(b_3 - b_5)^2 + b_4^2}}{b_4}$$

$$p_{10} = \left(\frac{b_1 b_4 - 2b_2 b_3}{4b_3 b_5 - b_4^2}\right) - p_{11}\left(\frac{b_2 b_4 - 2b_1 b_5}{4b_3 b_5 - b_4^2}\right).$$

Because the LOC is described by the formula $Y = X = 0 + 1 * X$, one can determine whether the ridge line equals the LOC by testing whether p_{10} equals zero and p_{11} equals one. However, in case that the conditions on a_1 to a_4 for a similarity effect are already satisfied (i.e., when $a_1 = a_2 = a_3 = 0$ and $a_4 < 0$), there is a much easier way to test whether the ridge equals the LOC: In this case, it suffices to test whether $a_5 = b_3 - b_5 = 0$, because this implies $p_{10} = 0$ and $p_{11} = 1$ in such a situation. We will now provide the mathematical proof that this is indeed the case.

Consider a situation in which the above introduced conditions on a_1 to a_4 hold, and in which the additional condition

$a_5 = b_3 - b_5 = 0$ holds. In particular, $a_3 = b_1 - b_2 = 0$ implies that $b_1 = b_2$, and $a_5 = b_3 - b_5 = 0$ implies $b_3 = b_5$. We can use these observations to compute the coefficients of the ridge line:

$$p_{11} = \frac{(b_5 - b_3) + \sqrt{(b_3 - b_5)^2 + b_4^2}}{b_4}$$

$$= \frac{0 + \sqrt{0^2 + b_4^2}}{b_4} = 1,$$

and

$$p_{10} = \left(\frac{b_1 b_4 - 2b_2 b_3}{4b_3 b_5 - b_4^2}\right) - 1 * \left(\frac{b_2 b_4 - 2b_1 b_5}{4b_3 b_5 - b_4^2}\right)$$

$$= \frac{b_1 b_4 - 2b_1 b_3}{4b_3 b_3 - b_4^2} - \frac{b_1 b_4 - 2b_1 b_3}{4b_3 b_3 - b_4^2} = 0.$$

That is, the position of the ridge line is $Y = p_{10} + p_{11}X = 0 + 1X = X$, which equals the position of the LOC.