Published in final edited form as:

Pers Soc Psychol Rev. 2016 May ; 20(2): 79–99. doi:10.1177/1088868315581119.

The Correlates of Similarity Estimates are Often Misleadingly Positive: The Nature and Scope of the Problem, and Some Solutions

Dustin Wood and **R. Michael Furr**

Wake Forest University

Abstract

Research on similarity constructs (e.g., dyadic similarity, personality stability; judgment agreement and accuracy) frequently find them to be associated with positive outcomes. However, a methodological pitfall associated with common 'overall similarity' indices, which we term the *normative-desirability confound (NDC),* will regularly result in similarity constructs apparently having more positive effects than they do in reality. In essence, when an individual is estimated to be similar to another person by common indices, this will strongly indicate that the individual has desirable characteristics. Consequently, the correlates of overall similarity indices can often be interpreted as indicating the beneficial effects of having desirable characteristics, without needing to attribute any additional salutary effect to similarity. We show that this confound is present in overall similarity estimates for a wide range of constructs (e.g., personality traits, attitudes, emotions, behaviors, values), how it can be accounted for, and discuss larger implications for our understanding of similarity constructs.

Personality and social psychologists are frequently interested in documenting the causes and consequences of different types of "similarities." For instance, we may be interested in whether individuals who are similar to their spouse or friends have better relationships with these people, what types of people show more consistent personalities across time or situations, or what qualities make some individuals more easily judged by others. These and other types of *similarity constructs* – referred to by terms such as *similarity, accuracy, agreement, stability, consistency, fit* – are all defined as the level of congruence between two sets of variables.

Research suggests that similarity in its various forms regularly has very positive correlates: individuals who are more similar to their spouse, who vary less across situations or time, and who are perceived by others the way they perceive themselves are generally found to be more satisfied with their relationships, and more well-adjusted generally (Carlson & Furr, 2013; Caspi & Herbener, 1990; Colvin, 1993; Donahue, Robins, Roberts, & John, 1993;

Correspondence concerning this article should be addressed to Dustin Wood, Department of Psychology, Wake Forest University, 438 Greene Hall, Winston-Salem, NC, 27109. dwood@wfu.edu.

Dustin Wood, Department of Psychology, Wake Forest University. R. Michael Furr, Department of Psychology, Wake Forest University.

Donnellan, Conger, & Burzette, 2007; Furr, Dougherty, Marsh, & Mathias, 2007; Gerstorf, Windsor, Hoppmann, & Butterworth, 2013; Gonzaga, Campos, & Bradbury, 2007; Gonzaga, Carter, & Buckwalter, 2010; Klimstra, Luyckx, Hale, Goossens, & Meeus, 2010; Luo & Klohnen, 2005; Roberts, Caspi, & Moffitt, 2001; Suh, 2002). Research also finds that the types of similarity that are beneficial may also be wide-ranging, including similarity in values, emotions, personality traits, behaviors, and attitudes (C. Anderson, Keltner, & John, 2003; Barni, Knafo, Ben-Arieh, & Haj-Yahia, 2014; Boer et al., 2011; Furr & Funder, 2004; Townsend, Kim, & Mesquita, 2013). These findings have been pointed to as evidence for the long-standing theoretical position that similarities of various types facilitate positive outcomes more generally (Byrne, 1971; Izard, 1960; Roberts et al., 2001).

Unfortunately, the story is not as simple as it might appear. There are a number of different statistics that can be used to index "similarity," but a consistent finding is that the correlates of similarity regularly differ dramatically as a function of which one is used. Such discrepancies raise important concerns about the theoretical conclusions drawn from this work, raise questions about the factors that drive apparent "similarity" effects, and highlight the need for clarity in determining the optimal ways of indexing similarity.

The purpose of this paper is threefold. First, we detail the *normative-desirability confound* (or *NDC*) as the major source of differences in in the correlates of similarity across methodologies. The result of this confound is that many apparent positive correlates of profile similarity say less about the correlates of "similarity" than of "being a desirable person." Second, we delineate the conditions under which this confound affects similarity indices. As we show, the NDC affects many common indices of similarity, and many types of similarity of interest to researchers (e.g., similarity in emotions, behaviors, values). Third, we provide suggestions regarding the examination of similarity, including ways to account for this confound.

The Correlates of Similarity Greatly Depend on How it is Indexed

How might we index the similarity between two individuals? A fairly intuitive and common method is a *profile correlation*, which estimates the correspondence between two profiles of scores on a common set of attributes or items. Frequently, these are called "*q*-correlations" (rather than the usual "*r*-correlations") to denote that items or attributes serve as the unit of analysis rather than individuals (Block, 1961; Cattell, 1952; Stephenson, 1953); we will use this convention here.[1] Using this technique, we can estimate *dyadic similarity* as a profile correlation between personality ratings of two people; *stability* (across time or situations) as a profile correlation between personality ratings from one person at two points in time; and *judgment agreement* as a profile correlation between an individual's self-ratings and ratings of the individual by one or more observers.[2]

[1]Other overall similarity indices such as the double-entry (or intraclass) profile correlation (McCrae & Terracciano, 2005), and Cohen's r_C (J. Cohen, 1969) are minor variants of the profile correlations we will discuss here, and are subject to the same issues (Furr, 2010).

[2]The correspondence between self and observer ratings is often referred to as *judgment accuracy* or *judgeability* (Colvin, 1993; Human & Biesanz, 2011). However, self-ratings and observer ratings of a given individual will often correspond due to both sharing the same biases rather than demonstrating accuracy (Leising, Erbs, & Fritz, 2010; Leising et al., 2013). Consequently, we will use the somewhat broader term *judgment agreement*.

As an example, we explore the level of personality similarity between Hermione and Ron, two adolescents who we suspect could make a good romantic match. We assess their personalities on a standard personality inventory; here: the Inventory of Individual Differences on the Lexicon (IIDL; Wood, Nye, & Saucier, 2010). As shown in Figure 1A, if we correlate their profiles, we see that Hermione and Ron have fairly similar personality profiles (q = .44); both are *kind* and *truthful*, and not *undependable* or *angry*, among other things. This is an index of their 'overall similarity' (Furr, 2008).

A long-recognized complication with this method is that peoples' profiles tend to be positively correlated, even if paired randomly (Kenny, Kashy, & Cook, 2006). This arises because people tend to endorse certain items more highly than others. We see this in Figure 1B, which shows the association between Hermione's profile and the *normative profile* – the average profile across individuals. The correlation between an individual's profile and the normative profile reflects the individual's *normativeness*: the extent to which he or she is like the average person (Furr, 2008).[3] As can be seen, both Hermione and Ron have highly normative profiles ($qs \geq .72$); that is, they have many of the characteristics that most people have. Consequently, our earlier inference of their similarity may need to be qualified: Hermione and Ron are similar in that they are relatively *kind*, *honest*, and *dependable*… just like almost everyone else.

The more important question may be whether Hermione and Ron are similar *in the characteristics that distinguish them from other people*. This can be estimated by first subtracting the normative profile from the two profiles (Furr, 2008; Kenny & Acitelli, 1994). For instance, if Hermione reports a 6 on a 0-to-10 measure of how *affectionate* she is, and if individuals on average report a score of 7, then Hermione's distinctive score will be 6 – 7 = −1. Thus, while Hermione describes herself as fairly affectionate, she describes herself as slightly less affectionate than the average person. Doing this for each item creates a *distinctive profile,* with scores of 0 indicating the person is 'average' in their level of a characteristic. Correlating two distinctive profiles reveals how similar they are in their distinctive attributes. As seen in Figure 1C, the *distinctive profile correlation* between Hermione and Ron is negative (q = −.46), indicating they are fairly dissimilar in their distinctive attributes. For instance, Hermione is more *controlling* and *organized* than the average person, whereas Ron is less.

One might assume that the correlates of similarity will be the same for overall and distinctive profile correlations. However, this is not the case. Because different items become the most extreme elements of a profile after removing the normative profile, the correlation between overall and distinctive profile similarity measures are frequently far below r = 1, and their correlates are regularly quite different from one another (Biesanz & Human, 2010; Klimstra et al., 2010; Sherman, Nave, & Funder, 2013; Wortman, Wood, Furr, Fanciullo, & Harms, 2014).

[3]Particularly within the area of person perception, the normative profile is sometimes referred to as the *stereotype*, and the profile correlation linking ratings to the stereotype profile as *stereotype accuracy* (Cronbach, 1955; Kenny & Acitelli, 1994; Kenny & Albright, 1987).

The phenomenon that frames this paper is *how* the correlates of similarity typically change when removing the normative profile. Across a range of literatures, distinctive profile correlations regularly have less positive or desirable associations with other variables than do overall profile correlations. For instance, investigators measuring dyadic similarity by calculating the overall profile correlation between personality profiles (e.g., a husband and wife's response profile) generally find that similarity is associated with well-being and relationship satisfaction (Gonzaga et al., 2010; Luo & Klohnen, 2005). However if the normative profile is removed, these associations are largely attenuated (e.g., Humbad, Donnellan, Iacono, McGue, & Burt, 2013). Likewise, in studies of stability across time or contexts, personality stability was associated with better adjustment (Campbell, Assanand, & Di Paula, 2003; Donahue et al., 1993; Roberts et al., 2001; Suh, 2002), but again these effects are attenuated if the normative profile is removed (Baird, Le, & Lucas, 2006; Klimstra et al., 2010; Sherman et al., 2013).

The fact that the correlates of overall and distinctive similarity differ so dramatically raises a question of which should be used. Distinctive similarity is appealing because it accounts for normative similarity – the fact that randomly-paired individuals usually have profiles correlated above zero. Distinctive similarity indices also clarify how similar people are on their most distinguishing characteristics, which are likely particularly important parts of how people represent one another (Karniol, 2003). These qualities might be responsible for the increasing use of distinctive indices of similarity (Baird et al., 2006; Bleidorn, Kandler, Riemann, Angleitner, & Spinath, 2012; Furr, 2008; Klimstra et al., 2010; Rogers & Biesanz, 2015). However, many investigators continue to use similarity indices that do not remove the normative profile (Boer et al., 2011; Decuyper, De Bolle, & De Fruyt, 2012; Gonzaga et al., 2010; Terracciano, McCrae, & Costa, 2010; Townsend et al., 2013), and this seems due to a sense that distinctive indices eliminate something that is not an artifact.[4] In our example, the concern is that we are removing meaningful ways in which Hermione and Ron *really are* similar to one another.

Resolving the question of which index to use requires deeper understanding of why this decision so drastically alters the correlates of similarity. Unfortunately, the core reason for these dramatic differences has remained a bit of a puzzle, and potential explanations have been varied. For instance, Kenny and colleagues (2006) suggested that "it appears that the correction removes artifacts that lead to a spurious correlation" (p. 333), and Acitelli, Kenny, and Weiner (2001) ventured that perhaps "a social desirability response set is a third variable that links stereotyped responding and relationship satisfaction" (p. 182). Alternatively, Funder (2001) suggests that the discrepancies may actually lie in problems with the distinctive similarity indices, which may simply be less reliable and thus less expected to correlate with anything.

Introducing the Normative-Desirability Confound

As we will show here, the major problem with using overall profile correlations to explore the correlates of similarity is the *normative-desirability confound (NDC)*: the correlates of

[4]We agree with some parts of this argument, and use the word 'artifact' sparingly.

overall similarity are generally highly confounded with the correlates of having a *normative* (i.e., mathematically average) psychological profile, which in turn is generally highly confounded with the correlates of having a *desirable* psychological profile. Thus, many apparent positive correlates of similarity may simply reflect the correlates of having desirable qualities (e.g., being nice and emotionally stable), with little relevance to similarity per se.

The NDC affects the interpretation of overall similarity indices when three conditions are present, shown in Figure 2. For simplicity, we will generically describe the NDC problem in terms of dyadic similarity, although the problem generalizes to other applications (e.g., consistency across time, judgment accuracy or agreement). First, there needs to be a fairly high level of normativeness in the sample. If normativeness is high – i.e., if individual profiles tend to correspond with the normative profile quite highly – we will find a substantial correlation between an individual's level of normativeness and their similarity to their partner when using 'overall similarity' indices (Path *a*). Second, an individual's level of normativeness must be correlated with their level of desirability (Path *b*). If these two conditions are present, then individuals who simply respond desirably will tend to have high profile similarity to their partner. Third, individuals with desirable profiles must be more likely to experience or report desirable outcomes (Path *c*).

Each path has been known separately for a long time. The first condition is relatively unexceptional: individual profiles simply *have to* have some expected level of normativeness unless all items in the inventory are equally endorsed on average (e.g., the mean for every item is the scale midpoint). If there is variability in item means, then by definition people tend to respond normatively, and levels of normativeness become larger as items differ more in their average endorsement (Krueger, 1998). By extension, if you have a more normative profile, you are more likely to be indexed as similar to any random person via an overall profile correlation. Past work has indicated that for most people, their ratings on an inventory will be positively related to the normative profile, and the level of correspondence can be quite high (Hoch, 1987; Krueger, 1998). The second condition is also very well-documented: a number of studies have shown that a normative profile is very highly correlated with the profile of item desirabilities. That is, items with higher average scores tend to be more desirable. For example, investigators have found that average self-ratings of trait items regularly correlate .80 or higher with the rated desirability of the items (A. L. Edwards, 1957; Goldberg, 1982). The third condition is also well-documented and intuitive – individuals with more desirable characteristics tend to have a wide range of desirable outcomes (Ozer & Benet-Martínez, 2006). For instance, individuals that are more agreeable, conscientious, emotionally stable, and open tend to have more satisfying romantic and friend relationships (Decuyper et al., 2012; Furler, Gomez, & Grob, 2014; Watson et al., 2004; Wortman & Wood, 2011). Importantly, these are main effects; that is, they concern simply the individual's level of these traits, not whether they are similar to their partner on these traits.

Although these conditions have long been recognized separately, what has been less appreciated is how they combine to impact the meaning of profile similarity. Concerning the joint effects of the first two conditions (Paths $a \times b$ in Figure 2): if we find an individual has

responded to a personality inventory the way their partner has, as evidenced by a high *q*-correlation, this individual is likely to have responded desirably. Although two individuals can and sometimes do achieve high profile similarity by being similar on undesirable traits (e.g., both individuals are dominant, disagreeable, sloppy, and neurotic), high similarity usually indicates that both individuals have desirable profiles, and low similarity indicates that at least one of the individuals has an undesirable profile (Leising, Ostrovski, & Zimmermann, 2013). Taken another way, an individual who simply completes a measure in a socially desirable way (e.g., "I'm kind, responsible, smart…") is quite likely to have a similar profile to their partner.

If all three conditions depicted in Figure 2 are present, this creates an indirect relationship between profile similarity and desirable outcomes (Path $d \approx a \times b \times c$). More specifically, these conditions cause overall profile similarity indices to be associated with desirable characteristics and outcomes *even though similarity may not actually cause or result from these variables*. Consequently, controlling for an individual's desirability or normativeness can cause the positive correlates of similarity to largely disappear (Clement & Krueger, 1998).

This has fundamental implications for interpreting studies reporting associations between similarity and relationship or well-being outcomes. For instance, individuals who have similar personalities tend to report higher happiness and relationship satisfaction (Gonzaga et al., 2007, 2010; Luo & Klohnen, 2005). However, due to the NDC, this may be better interpreted as indicating that individuals with partners who have desirable personalities, or who themselves have desirable personalities, are happier and more satisfied with their relationships, and that there is perhaps no additional benefit of being similar to one's partner. Similarly, individuals with more stable personalities report better adjustment (Caspi & Herbener, 1990; Roberts et al., 2001). However, due to the NDC, this might be better interpreted as indicating that individuals with more desirable personalities are better adjusted.

These conditions clarify why the apparent effects of similarity hinge so dramatically on the treatment of normativeness. In overall profiles where the normative profile is not removed, items with extreme scores on both profiles are likely to be highly evaluative characteristics, like being *honest* and *dependable* versus *abusive* and *selfish* (see the Y-axis in Figure 1B). Because such items are the most extreme elements of most individuals' profiles, they robustly affect overall similarity indices. Since desirable characteristics tend to be more common, removing the normative profile increases the likelihood that the extreme elements of an individual's *distinctive* profile are more neutral or undesirable characteristics. This is clearest for Ron: in his overall profile (Figure 1A), his most extreme traits are his tendencies to be *funny* and *brave,* and not *cruel,* or *rude*. But after removing the normative profile (Figure 1C), his most distinctive traits are his tendencies to be *disorganized, ordinary,* and *brave*, and not *calm, intelligent,* or *practical.*

As in our example with Hermione and Ron, the attributes that most characterize individuals are rarely those that most distinguish individuals from one another. Overall profile similarity

estimates tends to be higher and confounded with desirability, whereas distinctive profile similarity estimates will be lower and largely independent of desirability.

Two Empirical Illustrations

The NDC might seem to have a negligible impact on the correlates of overall similarity indices, in that three different conditions must be jointly present to produce it. However, we will show that these conditions are extremely likely to be present in studies that operationalize similarity constructs using profile correlations or other 'overall similarity' indices.

We report two empirical illustrations, with the goals of concretely demonstrating the existence and nature of the NDC, the predictable effects of removing the normative profile from similarity estimates, and the broad scope of the phenomenon. In Study 1, we illustrate how the NDC causes overall profile correlation estimates of personality similarity, stability, and judgment agreement to show associations with outcomes such as well-being, mental health, and relationship satisfaction that we should attribute to the individual's own desirability. Our aim is to demonstrate that the NDC is the primary reason that overall and distinctive similarity indices differ so dramatically in their correlates. We show that the correlates of overall similarity resemble the correlates of profile normativeness and desirability – quantities which are computed without using the other partner's attributes in any manner. In Study 2, we demonstrate that the NDC is present across a very wide range of measures and constructs in social and personality psychology, and across a broad range of populations. We show that the NDC almost universally affects the apparent correlates of similarity in any wide-ranging measure, including measures of emotions, values, attitudes, personality traits, and behaviors.

Study 1: Better Detailing the Nature of the Normative-Desirability Confound

We use data from a study of participants living in freshman dormitories who described their own personalities, the personalities of three randomly selected others from their dormitory, and their own personalities again about one year later. Our goal is to illustrate the conditions that produce the NDC. First, overall similarity estimates are confounded with profile normativeness; second, profile normativeness is confounded with profile desirability; and third, profile desirability is associated with positive outcomes. We show that shifting to distinctive indices of similarity unconfounds profile similarity estimates from profile normativeness and desirability estimates, and thus serve as more meaningful tests of the effects of similarity.

Study 1 Method

Participants and Procedure—Participants were freshmen living in dormitories at Wake Forest University (WFU). Each participant completed personality measures, and was asked to indicate who they were intending to room with in the following school year. Following this, they rated three different individuals ('targets') on their dormitory floor; the targets were randomly assigned with the constraint that participants were more likely to rate individuals that lived close to them (e.g., next door neighbors). Participants were then invited

to participate again a year later. Participants were included in the present analyses only if their intended roommate completed the survey or if they completed the survey in both years of the study. This resulted in a sample of 164 participants. Personality normativeness and desirability indices were available for 164 individuals; personality profile stability indices were available for 108 individuals; roommate personality similarity indices were available for 149 individuals; and judgment agreement indices (at least one rating of the participant's personality by another participant) were available for 126 individuals.

Measures

Personality ratings: Participants provided self-ratings on an early 58-item version of the Inventory of Individual Differences in the Lexicon (IIDL; see Wood et al., 2010) on a 1 (*Extremely uncharacteristic*) to 7 (*Extremely characteristic*) scale. In order to improve the reliability of the profile, at both Year 1 and Year 2 participants also described their personalities on this instrument a second time again a short period later (M= 5.2 and 6.2 days apart in Year 1 and 2, respectively; see Wood & Wortman, 2012) and these two ratings were averaged.

Item desirabilities: The IIDL items were rated for their desirability by an independent sample of 249 WFU undergraduates on a scale ranging from 1 (*Very undesirable*) to 5 (*Very desirable*)(see Wood & Wortman, 2012).

Profile similarity indices: The participant's profile of IIDL self-ratings in Year 1 served as a basis for all profile similarity indices.

The participant's *normativeness* was indexed as the correlation between their Year 1 profile and the normative profile ($\bar{Y}$) estimated as the average self-rating on each item across all participants at Year 1. The participant's *desirability* was indexed as the correlation between their Year 1 profile and the profile of item desirabilities. To index *roommate similarity*, the participant's Year 1 profile was correlated with the Year 1 self-ratings provided by the person they indicated as their intended roommate for the following year. To index *personality stability*, the participant's Year 1 profile was correlated with their own profile in Year 2. Finally, to index *judgment agreement*, the participant's Year 1 profile was correlated with ratings of the participant made by others assigned to rate their personality

Overall and distinctive indices: Following recommendations by Furr (2008), we computed both *overall* and *distinctive* profile correlations estimates of the participant's roommate similarity, personality stability, and judgment agreement with other raters. Overall profile correlations were computed by correlating raw IIDL scores (i.e., participant responses on the original scale metric). Distinctive profile correlations were computed by correlating these scores after subtracting the normative profile from each profile.

Squared difference similarity index: To demonstrate that the NDC impacts other commonly used similarity indices, we also estimated scores on the D^2 index. This index, described by Cronbach and Gleser (1953), sums the squared differences of two profiles across each item or variable:

$$D^2=\sum_{j=1}^{k}(x_{j1}-x_{j2})^2,$$

where *j* indicates an item (or variable) of a profile of *k* total items, and 1 and 2 indicate two separate profiles. The D^2 index is a dissimilarity index, in that high scores are obtained by having dissimilar profiles, and low scores (near the minimum of zero) by having profiles with very similar scores on each item. When estimated for a single variable or item, the D^2 is simply a squared difference score (i.e., $(x_1-x_2)^2$), which is closely related to an absolute difference score (i.e., $|x_1-x_2|$).

All of these statistics are used frequently in research on similarity constructs (Kenny et al., 2006; Robins, Fraley, Roberts, & Trzesniewski, 2001). The D^2 is closely related to overall profile correlations; indeed, if items are first ipsatized (standardized within-person), the subsequent scores are equal to overall profile correlations following a simple linear transformation (Cronbach & Gleser, 1953, Equation 4).

Outcome Variables

Personality disorder scales: Participants completed subscales of narcissistic, paranoid, and borderline personality disorder symptoms from the Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, 1997). Each item was either unendorsed (1) or endorsed (2) and scales were computed using standard scoring for the MCMI scales, with the exception that items pertaining to drug use or suicidality were removed.

Satisfaction with life: Participants completed an abbreviated three-item measure of the Satisfaction with Life Scale (Pavot, Diener, Colvin, & Sandvik, 1991) consisting of the items "In most ways my life is close to my ideal," "The conditions of my life are excellent" and "I am satisfied with my life" (α = .86). Items on this and all remaining scales were rated on a 1 (*Strongly disagree*) to 5 (*Strongly agree*) scale.

Depression: Participants completed two indicators of their level of depression: "I have often felt depressed recently" and "I often feel blue and sad" (α = .87).

Roommate relationship positivity: Participants completed four items indicating "I would say that [X] and I are friends," "I would say that overall, [X] and I fit in well with one another," "I would approach [X] for support when I need help," and "I am satisfied with my experience with [X]," where the person they indicated as their future roommate was given in each question (α = .88).

Perceived roommate similarity: Participants rated how much they felt similar to their future roommate by the item "Overall, I would say that me and [X] are very similar people."

Observer reports of liking, knowing, and feeling similar to the participant: In addition to rating the personality of selected targets on their floor, raters were additionally asked to describe their relationship with the target. To measure *liking*, raters were asked how much they agreed with the statement "Overall, I would say that I like [X]"; to measure *sense of*

knowing, raters were asked "I feel that I know [X] well"; to measure *perceived personality similarity*, raters were asked "I feel that [X] and I have similar personalities."

Study 1 Results

We first show how estimates of profile stability, similarity, and judgment agreement are associated with one another and other outcomes both using overall and distinctive similarity indices. We then elaborate on how differences in these associations are due to the NDC.

Relations between Overall and Distinctive Profile Similarity Measures—As shown in Table 1, the rank-ordering of which individuals have the highest and lowest profile similarity is substantially affected by the normative profile. The correlation between overall and distinctive estimates of roommate similarity was only .42; and the corresponding correlations were .61 for personality stability and .45 for judgment agreement, respectively.

We also see that the apparent relationships between roommate similarity, personality stability, and judgment agreement differed substantially as a function of whether overall or distinctive profile correlations were examined. Paralleling some past research (e.g., Bem & Allen, 1974; Caspi & Herbener, 1990), when roommate similarity, personality stability, and judgment agreement were estimated by overall profile correlations, they were positively correlated with one another (all *rs* ≥ .26). However, when estimated by distinctive profile correlations, there were no indications that they are related to one another (all |*rs*| ≤ .08).

Finally, it has been suggested that distinctive similarity might be problematic due to greater unreliability. To explore this possibility, we estimated the reliabilities of the overall and distinctive profile correlations using a variant of Cronbach's alpha coefficient which can estimate the reliability of an ordering of correlations (Sherman & Wood, 2014). As shown in Table 1, distinctive profile correlations did in fact have lower reliabilities than overall profile correlations. However, the upper-bound of a measure's expected correlations with other variables is the square-root of its reliability (Nunnally & Bernstein, 1991). Consequently while the distinctive stability estimates appear to have a problematic level of reliability ($\sqrt{\alpha_{CP}}$=.40), there is considerable potential for the distinctive similarity and judgment agreement estimates to show valid correlations with other variables ($\sqrt{\alpha_{CP}}$'s=.62 and .69, respectively).

Correlates of roommate personality similarity: As shown in Table 1, the correlates of roommate similarity differed dramatically as a function of whether the normative profile is removed. Using overall similarity, the correlates largely paralleled the correlates of simply having a normative or desirable personality profile: individuals who were similar to their roommates reported lower pathology and depression, and higher life satisfaction, perceived similarity, and relationship satisfaction with one's roommate (all |*rs*| ≥ .17; all *ps* < .05).

However, a close look at these associations suggests that the overall similarity index is not working as it should. Specifically: individuals that had similar profiles to their roommate were considerably more likely to report good mental health (e.g., higher life satisfaction, lower depression, lower scores on personality disorder measures, all |*rs*| between .28 to .46) than to report feeling satisfied and similar to their roommate (*rs* = .20 and .17, respectively).

After removing the normative profile by computing a distinctive roommate similarity index, all relationships with personality disorders and general well-being outcomes reduced to nonsignificance (all $|rs| \leq .15$, $ps > .05$). Distinctive similarity also was no longer significantly associated with better relational positivity ($r = .10$, $p = .18$). However, the most sensible, face-valid relationship remained: distinctive similarity continued to be associated with feeling that one was similar to one's roommate ($r = .20$).

Correlates of personality stability: We identified fewer correlates of overall estimates of personality stability over one year, but these also appeared to be produced by the NDC. Paralleling previous research (Caspi & Herbener, 1990; Klimstra et al., 2010; Roberts et al., 2001), raw estimates of personality stability were associated with better adjustment: individuals with more stable profiles tended to have lower levels of depression ($r = -.27$) and less indications of avoidant personality disorder ($r = -.27$).

However, paralleling results reported by Klimstra and colleagues (2010), after removing the normative profile, these relationships disappeared. Instead, the only significant predictor of personality stability was a new positive association with narcissism ($r = .21$).

Correlates of judgment agreement: Finally, individuals whose self-descriptions more closely matched with how they were described by others tended to have lower scores on paranoid, avoidant, borderline personality disorder scales and depression scales ($rs \leq -.21$), and higher life satisfaction ($r = .14$). Somewhat strangely, such individuals had better relationships with their roommates ($r = .16$) although their roommates had not been selected to rate their personality. Raters who described the target in ways that matched the target's self-perceptions also reported much greater sense of liking ($r = .69$), feeling similar ($r = .52$), and knowing ($r = .24$) the target.

Again, most of these correlates disappeared after removing the normative profile. Distinctive agreement was not associated with well-being or adjustment (all $|rs| \leq .09$). However, some sensible relationships remained. Raters whose judgments more highly patterned the individual's distinct self-perceptions still reported greater liking, similarity, and knowledge of the target, but these findings were much more modest, with *rs* reducing from .69, .52, and .24, respectively, using overall similarity estimates to the much smaller range of .18, 13, and .17, respectively, using distinctive similarity estimates.

Establishing the Conditions for the Normative-Desirability Confound (NDC)—

As we have elaborated, the dramatic shifts between the correlates of overall and distinctive profiles arise from the NDC, which should be present if (a) an individual's indexed level of 'overall profile similarity' (e.g., roommate similarity, judgment agreement) is highly correlated with their own level of normativeness, (b) their level of normativeness is highly correlated with their level of desirability, and (c) their level of desirability is correlated with positive outcomes (Figure 2). We now establish the presence of these three conditions.

Profile normativeness is associated with overall similarity coefficients: There is high normativeness in how members of this sample responded to this inventory. As shown in Table 1, a given participant's profile correlated on average $M(q\bar{Y}_{Y_1}) = .81$ with the normative

profile. This average normative correlation allows us to estimate the level of similarity expected to occur between two profiles that are randomly paired. As described by Wood and Brumbaugh (2009) the average expected correlation between the profiles of two randomly selected individuals is approximated by the square of the average normative correlation.[5] Consequently the IIDL profile of any two individuals was expected to correlate $.81^2$, or .66, which closely approximates the level of profile similarity actually found among roommates: $M(q_{Y_1R_1}) = .68$.

As we have noted, normative individuals are likely to be similar to any randomly selected individual. Table 1 shows this is clearly the case. Individuals with normative personality profiles were more likely to be similar to their roommate ($r = .63$). Additionally, normative individuals tended to have more stable profiles over a year ($r = .55$), and were described by others the way they described themselves ($r = .39$) as indexed by overall profile correlations. Thus, we see that Path *a* in Figure 2 is strong.

As distinctive similarity indices are formed by subtracting the normative profile, it should be unsurprising that distinctive similarity, stability, and judgment agreement indices were almost entirely unassociated with normativeness (all $|rs| \leq .04$).

Relations between profile normativeness and profile desirability: As Table 1 shows, an individual's level of normativeness was nearly identical to their level of profile desirability ($r = 97$). Thus, Path *b* in Figure 2 is exceptionally strong: individuals describing themselves in a normative manner are essentially describing themselves as having desirable attributes.

To understand this relationship, it is useful to understand the items that anchor the normative profile. Within this sample, the most-endorsed items (i.e., highest mean self-ratings) were *courteous/polite*, *dependable/reliable*, *competent/capable*, *kind-hearted/caring*, and *truthful/honest* (means between 78 and 81% of the maximum scale range). The least endorsed items were self-ratings of being *cruel/abusive, dumb/stupid, unstable/disturbed,* and *unsympathetic/unfriendly* (means between 10% to 17% of the maximum scale range).

Because of the exceptionally strong relationship between profile normativeness and desirability, the patterns of associations between involving normativeness and desirability were nearly interchangeable. Profile desirability was associated with higher overall roommate personality similarity ($r = .61$), personality stability ($r = .55$) and judgment agreement ($r = .38$), and was unassociated with all of the distinctive profile similarity measures (all $|rs| \leq .06$). This indicates that in this sample, removing the normative profile effectively creates an index of similarity no longer confounded with the effects of simply being a generally desirable person.

Correlates of profile normativeness and desirability: As just shown, normative individuals tend to be much more similar to their roommates (Path *a* in Figure 2), and have much more desirable profiles (Path *b*). If having a desirable personality is associated with positive outcomes (Path *c*), then this will satisfy the third condition for producing the NDC

[5]This is analogous to the fact that the square root of the average inter-item correlation equals the expected correlation between a given item with the latent variable or the perfectly reliable "true score" in classical test theory (Nunnally & Bernstein, 1991).

(Path $d = a \times b \times c$). As expected, profile desirability was in fact associated with positive outcomes: individuals with desirable profiles tended to describe themselves as having substantially lower levels of paranoid, avoidant, and borderline symptoms and depression (*rs* ≤ −.36), higher levels of life satisfaction ($r = .36$), and more positive relationships with their roommate ($r = .16$, $p = .06$). As expected, the correlates of normativeness were nearly identical to those of profile desirability; this can be seen by comparing Columns 1 and 2 in Table 1.

Because all three conditions for the NDC were present in the current sample, we see that the correlates of overall similarity parallel the correlates of profile desirability. By comparing Columns 2 with Columns 3 and 7 in Table 1, we see that the correlates of profile desirability closely paralleled the correlates of overall roommate similarity and overall judgment agreement. (Somewhat strangely given the high .55 correlation between profile desirability and profile stability, the correlates of profile stability over a year showed less similarity to the correlates of profile desirability.) In contrast, the correlates of distinctive indices of similarity showed far less resemblance to the correlates of profile desirability.

Are the Conditions for the NDC Present for Other Overall Similarity Coefficients?—Finally, we explored the extent to which D^2 indices are confounded with profile normativeness and desirability. Recall that D^2 indices are *dissimilarity* measures, and should thus show negative correlations with these measures. The D^2 indices of roommate similarity and personality stability showed $r = -.77$ and −.83 associations with corresponding overall profile correlations; $r = -.36$ and −.34 associations with the normative profile; and $r = -.35$ and −.33 associations with the desirability profile, respectively.

These results thus indicate that D^2 indices show smaller associations with normativeness and desirability than overall profile correlations, but these associations were nonetheless substantial. Thus, D^2 indices may also say more about the correlates of having a desirable profile than about the correlates of having a stable, similar, or easily judged profile.

Are the Conditions for the NDC Present in Other Samples?—One critique of this study might be that certain attributes of the measure may have made the NDC a greater problem than it generally would be in other contexts. The IIDL instrument is designed to have a broader and more extreme range of items (e.g., *cruel/abusive*) than in many similar measures, and self-ratings were collected twice a couple days apart and averaged. Both of these features should serve to increase levels of normativeness. It could be that inventories with a less extreme range of items and administered in a more usual fashion (by obtaining self-rating items only once) will be less susceptible to this confound.

To explore this, a brief replication using the widely-used Big Five Inventory (BFI; John, Naumann, & Soto, 2008) and a sample of college roommates followed over a year is detailed in the Supplementary Materials S1. As shown there, overall profile stability and roommate similarity were considerably associated with normativeness and desirability, were associated with life satisfaction and university burnout, and were associated with one another. All of these associations were substantially reduced – almost invariably to insignificance – when shifting to distinctive profile similarity indices. Paralleling results

shown in Table 1, distinctive roommate similarity continued to be associated with roommate relationship quality and perceived personality match. This indicates again that while relationships with well-being measures tend to decrease by shifting from overall to distinctive similarity indices, they tend to remain correlated with highly face-valid and sensible outcomes, such as global perceptions of similarity.

Study 1 Discussion

Consistent with previous findings, the correlates of various similarity constructs differed dramatically as a function of whether the indices removed the normative profile. Our goal was to show why: when the normative profile is not removed, similarity indices are strong proxies for desirability.

Due to the NDC, the correlates of commonly-used similarity indices such as overall profile correlations and D^2 coefficients may have little to do with "similarity" and more to do with having desirable attributes. This problem is illustrated by comparing the correlates of overall roommate similarity to the correlates of profile desirability and the correlates of distinctive roommate similarity (i.e., Column 3 with Columns 2 and 4, respectively in Table 1). Tellingly, the correlates of overall similarity indices parallel the correlates of having a desirable personality profile more closely than the correlates of similarity indices that have removed the normative profile. This was also well illustrated with overall and distinctive estimates of self-other judgment agreement (i.e., comparing Columns 2, 7, and 8 in Table 1).

Given that removing the normative profile when examining profile similarity is relatively novel (Furr, 2008; Shoda, Mischel, & Wright, 1994), it is worth highlighting evidence that such indices are valid. Despite the fact that these indices were less reliable than overall similarity indices, they were reliable enough to produce sensible correlations with other variables. Although profile similarity shed most associations with measures of well-being and other desirable outcomes once the normative profile was removed, sensible relations remained (e.g., with perceptions of similarity). There are many reasons to think that distinctive similarity measures are valid measures of the types of similarity they are supposed to measure.

Study 2: The Normative-Desirability Confound is Almost Everywhere

We continue by further demonstrating the generality of the conditions that produce the NDC across other personality inventories, and across different classes of constructs such as values, interests, emotions, and behaviors. Although a strong link between item desirabilities and *average* item endorsements has been found with a range of inventories and item pools (e.g., Edwards, 1957; Goldberg, 1982; Wood & Wortman, 2012), past research has generally not addressed the question of how strongly item desirabilities are correlated with a *single individual's* endorsement of personality items. As described earlier, if there is a relatively low level of normativeness in response patterns – that is, if individuals do not tend to respond particularly similarly to one another – then the desirability profile could correlate with the average response profile very highly, but not be particularly predictive of a single randomly-selected individual's response. Given the conditions detailed as necessary to produce the NDC in Figure 2, finding normativeness to be low on certain inventories would

mean that the NDC might not bias the meaning of overall profile similarity coefficients with these inventories.

Additionally, although we have illustrated that the conditions producing the NDC exist in personality inventories, researchers index similarity in terms of many other constructs – such as emotions (C. Anderson et al., 2003; Townsend et al., 2013), values (Boer et al., 2011), attitudes (Byrne, 1971), behaviors (Furr & Funder, 2004), and situations (Sherman, Nave, & Funder, 2010). It is less clear whether the NDC influences similarity indices of these constructs. To address this, we analyzed a large number of commonly-used measures for a range of constructs to estimate whether (a) individuals tend to show high normativeness and (b) the normative profile is correlated with profiles of desirabilities. These parallel the conditions labeled Paths *a* and *b* for producing the NDC as outlined earlier in Figure 1, respectively.

Study 2 Methods, Results, and Discussion

Estimating these two properties – the level of normativeness, and the correlation between normativeness and desirability – requires only raw data from a measure, and secondary coding of its items' desirability. We thus obtained scores and desirability ratings from a wide range of personality inventories, and from measures of other types of constructs: including behaviors, situational characteristics, emotions, and attitudes. We detail the measures and sample characteristics used to obtain scores and desirability ratings in Supplementary Materials S2.

Because these measures reflect a range of scales (e.g., 1 to 5 scales, 1 to 9 scales) we transformed each to a 'percentage of maximum possible' metric (P. Cohen, Cohen, Aiken, & West, 1999), to place them on a common 0–100 range. Tables 2 and 3 present the items with the highest and lowest averages for each measure, as they most illustrate the meaning of normativeness for that particular measure.

We first discuss whether the NDC influences overall similarity coefficients in the context of personality inventories, and then across measures of other classes of constructs.

Personality Inventories

Levels of normativeness, and correlations between normative and desirability profiles: We report (a) the average normativeness of a single participant's personality self-ratings (the $q_{\bar{Y}Y_i}$ values shown in Table 2), estimated as the square-root of the average inter-rater response similarity, and (b) the correlation between the profile of item desirabilities and item average endorsements (the $q_{D\bar{Y}}$ values in Table 2). The product of these two values estimates the average profile desirability of a single individual (the q_{DY_i} values in Table 2).

As seen in Table 2, all personality inventories showed considerable levels of normativeness in these samples. In the non-clinical samples, the expected profile normativeness of a single rater ranged from a high of .76 for the PDA instrument to a low of .46 with the HEXACO instrument. Even among participants diagnosed with BPD who completed the HEXACO, this value was not appreciably lower (.41). We thus found that an individual's level of normativeness is fairly high when using any of the personality inventories examined. This

leads to an expectation that when overall profile correlations or squared difference scores are used, an individual will likely have high similarity to others, high stability over time, and be described by others as they describe themselves, just by having a normative personality.

The relationship between the normative ($\bar{Y}$) and desirability (D) profiles was also extremely high. For non-clinical samples, $q_{D\bar{Y}}$ values ranged from .84 to .95. For the clinical sample, the correlation was lower but still high ($q_{D\bar{Y}}$ = .53). This indicates that in most contexts and inventories, an individual who has a normative personality profile effectively has a desirable personality profile. And as shown in Table 2, an individual who has a desirable personality, or simply describes oneself in a desirable manner, is much more likely to be indexed as similar to another randomly selected person on all inventories (all q_{DY_i} values between .38 and .66).

The nature of the normative profile: The available datasets do not allow us to formally examine the third condition necessary for the NDC: that the desirability of an individual's profile was associated with desirable outcomes (Path *c* in Figure 2). However, Table 2 provides the most and least endorsed items in each inventory, which most illustrate the meaning of normativeness.

These items reveal why an individual's normativeness and desirability relate positively to well-being and relational outcomes in a slightly different way than seen thus far. Specifically, it is likely that a major reason overall similarity indices correlate with measures of emotional and social adjustment is because indicators of emotional and social adjustment largely anchor the normative profile. Here: individuals are estimated as having a normative personality profile in large part *because* they say that they are "considerate and kind to almost everyone" (BFI), are "reliable" and "trustworthy" (PDA), and are not "cruel, abusive" (IIDL), "self-pitying" (CAQ), "depressed, blue" (BFI), or someone who "seeks conflict" (BFAS).

This is a variant of the problem of criterion-predictor overlap described by Nicholls, Licht, and Pearl (1982), who noted that large correlations often exist between two scales for the uninteresting reason that nearly identical items are found in both scales. As applied here, one key reason that overall similarity indices might correlate with mental health and relational outcomes is because in many inventories, reporting that one has good mental health and relationship qualities is very nearly necessary to being estimated as having a highly normative profile.

Non-Personality Inventories—Investigators often use overall similarity indices (e.g., overall profile correlations, D^2) to index similarities for classes of constructs other than personality traits. We next explore whether the conditions that produce the NDC apply to other important classes of constructs. Specifically, we estimated normativeness for a range of different measures and samples, and report items with the highest and lowest means in each measure. Results are in Table 3.

Behaviors: The Riverside Behavioral Q-Sort (RBQ) is a wide-ranging measure of behaviors (Funder, Furr, & Colvin, 2000). Table 3 shows that individuals' RBQ profiles in an

unstructured interaction have normativeness levels comparable to personality inventories ($q_{\bar{Y}Yi}$ = .73), and the normative RBQ profile was highly desirable ($q_{D\bar{Y}}$ = .74). The items with the lowest means (*Ms* ≤ 28) reflected tendencies to act in a disengaged and oppositional fashion; items with the highest means (*Ms* ≥ 73) involved tendencies to behave in an engaged and positive fashion. These values indicate that having similar patterns of behavior across situations or with others is a strong proxy for simply exhibiting a prosocial, positive behavioral style.

Situation characteristics: In the similar Riverside Situational Q-Sort (RSQ; Sherman et al., 2010), an individual's description of the features of a sampled situation ("describe what you were doing yesterday at this time") tended to have a considerable level of normativeness ($q_{\bar{Y}Yi}$ = .45), and the normative profile tended to covary highly with the desirability profile ($q_{D\bar{Y}}$ = .57). The least endorsed items (*Ms* ≤ 20) involved the presence of physical or emotional threats, whereas the most endorsed items (*Ms* ≥ 63) involved relatively simple and enjoyable situations. These values indicate that high scores of overall similarity indices across two randomly sampled situations should be moderately effective proxies of being in pleasant, non-threatening situations.

Emotions: The Differential Emotion Scale-Revised (Fredrickson, Tugade, Waugh, & Larkin, 2003) assesses a wide range of emotions. In reports of how frequently people felt a list of emotions across the past two weeks, the level of normativeness was very high ($q_{\bar{Y}Yi}$ = .63), and the normative profile covaried extremely highly with the desirability profile ($q_{D\bar{Y}}$ = .90). The least endorsed items (*Ms* ≤ 32) concerned tendencies to feel *ashamed, guilty*, and *disdainful*, and the most endorsed items (*Ms* ≥ 71) concerned tendencies to feel *glad, optimistic,* and *love*. Together, these indicate that high scores on overall indices of emotional similarity is a very strong indicator that the individual is simply feeling more positive than negative emotions.

Attitudes: The Saucier Dictionary-based Isms-46 (Saucier, 2014; Saucier et al., 2014) assesses a range of attitude or belief dimensions. The level of normativeness in attitudes was very high ($q_{\bar{Y}Yi}$ = .66), and the normative profile correlated very highly with the desirability profile ($q_{D\bar{Y}}$ = .85). The least endorsed items (*Ms* ≤ 20) concerned hedonism, materialism, and ethnic superiority, and the most endorsed items (*M* ≥ 80) concerned egalitarianism, constitutional government, and of consideration of others' interests. These values indicate that high scores on overall indices of attitude similarity are strong indicators that the individual is other-regarding rather than entirely self-regarding in his or her belief system.

Values: As indicated elsewhere (Schwartz & Bardi, 2001), there were high levels of normativeness regarding which values were most and least important ($q_{\bar{Y}Yi}$ = .61). The least endorsed values (*Ms* ≤ 47) especially concern valuing power, authority, whereas the most endorsed values (*Ms* ≥ 86) concern being responsible, self-respecting, and honest. This indicates that high scores on overall indices of value similarity are strong indirect indicators that the individual has more other-regarding, communal values than selfish, agentic values.

Music preferences: Levels of normativeness were lower in this domain ($q_{\bar{Y}Yi}$ = .46). In a sample of college undergraduates, the least endorsed music preferences (*Ms* ≤ 41) concerned

preference for folk, religious, heavy metal, and country music, and the most endorsed (*Ms* ≥ 68) concerned preferences for rock, alternative, and pop music.

Vocational interests: Levels of normativeness were the lowest in this domain ($q_{\bar{Y}Yi}$ = .37). The least endorsed vocational interests (*Ms* ≤ 30) concerned unengaging, repetitive jobs (working on an assembly line, sorting mail), whereas the most endorsed (*Ms* ≥ 59) concerned more agentic or interpersonal jobs (e.g., starting a business, volunteering, teaching).

Summary—For almost every class of construct we examined, our findings indicate that overall similarity indices are highly confounded with normativeness and desirability, which should substantially influence how they relate to other variables. For similarity in values, behaviors, attitudes, emotions, and situation characteristics, high scores on common 'overall similarity' coefficients indicate not just that the two individuals are *similar* but that each individual is *desirable* – especially by being more well-adjusted and other-regarding.

General Discussion

Our aim has been to clarify that common indices of overall similarity (e.g., simple profile correlations, squared difference scores) will regularly correlate with desirable outcomes simply due to being confounded with the individual's general desirability. This *normative-desirability confound (NDC)* is the major reason for dramatic shifts in the apparent correlates of similarity as a function of how similarity is indexed.

Understanding the NDC is important for making decisions regarding how to index similarity, and for drawing appropriate theoretical conclusions. We continue by discussing the breadth of the NDC, how we might regard the confounding of normativeness, desirability, and similarity at a more conceptual level, and recommendations for research on similarity constructs.

When is the Normative-Desirability Confound Likely and Unlikely to Occur?

There are circumstances where the NDC may be stronger or weaker; these can be seen to some extent in Tables 2 and 3. First, the NDC should impact findings more strongly when similarity is indexed using measures containing more extremely desirable or undesirable items. Additionally, as indicated in Table 2, normativeness should be more correlated with desirability in 'normal' populations than clinical populations.

The NDC should not problematically impact similarity indices for *all* classes of constructs. This is most clear in the case of music preferences in Table 3. Normative music preferences in the college population sampled seemed to be for rock, pop, and alternative music and against folk, religious, and country music. There is little reason to think that individuals reporting such normative music preferences show greater emotional adjustment and prosociality in the same way as found for normative personality traits, attitudes, emotions, and behaviors. Because the correlation between normativeness and desirability is central to the NDC (Path *b* in Figure 2), one way to address whether the NDC might affect the correlates of overall similarity is to code an inventory's items for desirability. If the

correlation between item means and desirabilities is low ($q \approx 0$), then overall similarity should not be confounded with desirability.

Importantly, the NDC may influence many more constructs than most researchers may realize. It seems almost universally present when examining similarity across personality attributes (Table 2; A. L. Edwards, 1957; Goldberg, 1982). Given a normative tendency to report communal, other-serving values (Park, Peterson, & Seligman, 2006; Schwartz & Bardi, 2001), the correlates of 'overall value similarity' should be confounded with such values. This confounding of similarity and desirability is likely for several other classes of constructs, such as emotions (Townsend et al., 2013), behavioral profiles (Furr, 2009; Shoda et al., 1994), situation characterizations (Sherman et al., 2010), parental styles (Deal, Halverson, & Wampler, 1999), and self-presentational personas (Leary & Allen, 2011).

A Deeper Appreciation of Normativeness

> "Dyadic analysis is a breeding ground for artifacts."
>
> – Lee Cronbach (1958)

> "One man's artifact may be another man's main effect."
>
> – William McGuire (2009)

Despite the ways in which normativeness can lead to artifactual correlations between similarity and desirable outcomes, some theorists who recognize these effects have argued against removing normativeness. The logic is that similarities between random individuals reflect real rather than artifactual similarities between everyone, and consequently removing the normative profile removes valid variance. For example, when considering whether to "correct" for the normative profile when examining the accuracy of personality judgments, Funder (2001) notes:

> The word "correction" implies that what one has afterwards is "correct," and in this context that is not always the case… it is no trivial feat to understand what people in general are like. If a judge of personality attains some of his or her accuracy through a profound understanding of general human nature, should this be held against him or her? … In the absence of specific knowledge, the most reasonable strategy is to guess the baseline or average … This is not an artifact – it is part of being a good judge. (p. 326)

Relatedly, some investigators note that because of normativeness, accuracy in predictions of characteristics of individuals or groups can often be increased by using one's own characteristics as a basis of prediction (Hoch, 1987).

We are sympathetic to these arguments, in that there is an important sense in which the personalities of most people *really do tend to be similar*. In the same way that two randomly selected humans can be expected to have over 99% of their genetic profile in common (Venter et al., 2001), any two individuals tend to have psychological and behavioral tendencies that are at least somewhat similar. This is why we have described the NDC as a *confound* rather than an *artifact*. All three components we've outlined as necessary to produce the NDC (Figure 2) reflect real phenomena: people *really do* show base-rate

similarities to one another, the normative profile *really is* desirable, and having a desirable profile *really does* result in real outcomes. The problem is not that any of paths are artifactual, but rather that together they often result in similarity indices being associated with outcomes for reasons that have nothing to do with the effects of being similar to someone.

The idea that there are real base-rate similarities in personality traits may be jarring to some psychologists, where 'personality traits' may be nearly equated with 'dimensions of individual differences' (McCrae & Costa, 2008). However, normative psychological and behavioral attributes likely play a critical role in many psychological phenomena. For instance, the greater frequency of positive traits in the population is necessary for certain explanations of the greater effect of negative trait information on behavior (Kellermann, 1984) and the faster processing of positive trait information (Unkelbach, Fiedler, Bayer, Stegmüller, & Danner, 2008). Similarly, some prosocial policies regarding how to interact with others – 'it is better to forgive and forget'; 'give and you shall receive'; 'expect the best in people' – may work mainly *because* the social environment is normatively positive. In less positive environments, less prosocial policies are often advisable (McNulty & Fincham, 2012).

To reiterate: overall similarity indices are problematic not because they fail to index real similarities, but rather because the characteristics people share with one another when indexed by these indices are disproportionately desirable ones. Consequently, it becomes difficult to attribute correlations involving overall similarity indices to the effects of similarity – the desirability of normative attributes may be doing all the work. Both for the sake of attempting to understand similarity as unconfounded with general desirability, and for the sake of understanding normativeness as a potentially important psychological phenomenon in its own right, researchers should adopt methods that differentiate these components as much as possible.

Recommendations in Operationalizing Similarity Constructs

Focus on distinctive similarity indices—Because overall similarity indices confound similarity, normativeness, and desirability, we believe the natural implication is that overall similarity indices should not be the primary or default index of similarity when exploring correlates of similarity. Rather, indices that control for normativeness (and, consequently desirability) should be the primary focus in profile-based examinations of the correlates of similarity constructs (e.g., agreement, stability, consistency).

Interpretation of distinctive similarity indices should be based on the recognition that they are conservative (Furr, 2008). Consider self-other agreement in personality judgments, where a target participant rates herself on a set of traits and is rated by an acquaintance on the same traits. If the target is quite like the average person and the judge correctly recognizes that fact, then their "distinctive agreement" correlation will likely be small. This result is an accurate reflection of the fact that the target and judge do not share similar perceptions of any non-normative qualities of the target. However, it is indeed a conservative perspective on agreement.

Thus, neither overall nor distinctive indices are perfect indicators of "pure" similarity. However, we believe that despite their conservative nature, distinctive indices are clearly preferable to indices that fail to account for the substantial confounding effects of normativeness and desirability when one is interested in establishing the correlates of similarity.

Joint examination of overall similarity, distinctive similarity, and normativeness—Even while focusing on distinctive indices, researchers can benefit from a broader perspective that includes multiple indices of similarity and normativeness. Much can be learned by joint examination of normativeness and overall similarity indices alongside distinctive similarity indices. Not only does an inclusive approach alleviate some concerns about focusing on a relatively conservative index of distinctive similarity, it also allows researchers to learn about the meaningful role normativeness often plays in many important phenomena, as described above.

There is an emerging recognition that examining all coefficients in the same analysis can reveal interesting effects (Bleidorn et al., 2012; Borkenau & Zaltauskas, 2009; Furr et al., 2007, 2007; Furr & Funder, 2004; Furr & Wood, 2013; Leary & Allen, 2011; Locke, Zheng, & Smith, 2014). For instance, Biesanz and Human (2010) found that experimentally manipulating motivation among judges to attain accurate perceptions of targets was not associated with increased judgment agreement as indexed by an overall profile correlation. However, this null association masked tendencies for accuracy motivation to increase distinctive agreement and to decrease tendencies to rate targets in a normative manner. Thus, failing to separate normativeness from similarity can considerably compromise our understanding of key phenomena, as normativeness and distinctive similarity can have independent or even opposing relationships with other variables.

Other techniques that successfully account for normativeness—Although we have focused on illustrating how distinctive profile correlations can be used to address the NDC (Furr, 2008), other techniques are available. First, Biesanz and colleagues use linear mixed modeling and regression to control normativeness when examining personality judgment accuracy and agreement (e.g., Human & Biesanz, 2011b; Rogers & Biesanz, 2015). Other investigators use simultaneous regressions or partial correlations to control normativeness or desirability when examining associations of overall similarity estimates with other variables (Clement & Krueger, 1998, 2002; Luo & Snider, 2009; Sherman et al., 2010). The relative merits of these various methods deserve further attention, but they largely should show similar results.[6]

As shown by many of these investigators, an advantage of multiple regression or linear mixed modeling is that they can separate the effects of normativeness and desirability. Although the correspondence between the normative and desirability profiles is regularly at levels of $q = .85$ or higher, this leaves a surprising level of room for item means to differ

[6]In Study 1, regressing overall profile correlation estimates of roommate similarity on self and roommate normativeness results in residuals which correlated .74 with distinctive roommate similarity estimates shown in Table 1. Regressing overall personality similarity on Time 1 and Time 2 normativeness resulted in residuals which correlated .68 with the distinctive stability estimates in Table 1.

from item desirabilities. For instance, many undesirable characteristics (e.g., *anxious, depressed, procrastinating*) are more endorsed and many desirable characteristics (e.g., *gorgeous, hilarious, influential*) less endorsed than their rated desirability would predict (N. H. Anderson, 1968). Most effects of normativeness seem due to its confounding with desirability, but there are indications that normativeness can have independent effects when both are examined together in a single analysis. For instance, Rogers and Biesanz (2015) found that well-adjusted judges provide ratings matching target self-perceptions not just by rating targets in a desirable fashion, but also by employing knowledge of less desirable characteristics that describe most people.

A second method addresses the NDC problem by avoiding profiles altogether, examining each trait separately (i.e., a "variable-centered" approach, Furr et al., 2007). For example, researchers might examine self-other agreement via moderated regression, in which two scores on matched variables (e.g., self and partner scores on Extraversion) and their interaction are entered as predictors of an outcome in a regression model (Aiken & West, 1991; Cronbach, 1958; J. R. Edwards & Parry, 1993). The interaction term indicates the importance of similarity or matching. However, this approach introduces other complexities. For instance, decomposing the 51-item IIDL instrument used in Study 1 to its elements can create a bewildering number of main effects and interactions (Wortman et al., 2014). Moreover, interaction terms are likely to have low reliability even if items are first aggregated into a smaller number of scales (Busemeyer & Jones, 1983). The nature of relations between "variable-centered" and "person-centered" approaches are complex, and is an important area for further psychometric research.[7]

Difference score indices of similarity do not avoid normativeness—As described in Study 1, a relatively common and intuitive index of similarity is squared or absolute differences between scores, as with the D^2 index (e.g., Cronbach & Gleser, 1953). A desirable property of D^2 is that it can be estimated for a single attribute or across many attributes.

As shown in Study 1 and elsewhere both empirically and algebraically, squared and absolute difference indices tend to be extremely highly associated with overall profile correlations (Cronbach & Gleser, 1953). The manner in which the effects of squared difference scores (i.e., $(X_1-X_2)^2$) become confounded with main effects of each variable (i.e., effects of X_1 and X_2) has been detailed elsewhere (Cronbach, 1958; J. R. Edwards, 1993, 2001).[8] This in turn creates the same problems as with overall profile correlations: when squared difference

[7]Distinctive profile correlations have interesting relations with the cross-product terms in moderated regressions (Wortman et al., 2014). For instance, in Study 1, the sum of the $(Y_1 - M_{Y_1})(Y_2 - M_{Y_2})$ interaction terms across elements – i.e., the sum of the terms serving as the critical terms for testing similarity effects in moderated regression, following recommended mean-centering (Cohen et al., 1999) – correlated .93 with distinctive profile correlation estimates of roommate similarity and .68 with distinctive correlations in the case of personality stability estimates. Thus, distinctive profile correlations can be seen as roughly an omnibus test of the general importance of similarity or matching across all attributes, which can be useful for dredging out small effects that might be hard to establish at the univariate level. Further research to connect univariate moderated regression approaches and multivariate profile approaches is warranted.

[8]To see why, consider that when desirable items are highly endorsed, as typical in many types of inventories (Tables 2 and 3), large D^2 scores (indicating dissimilarity) are obtained when one person has undesirable characteristics. For example, because people tend to agree that they are *reliable* and *considerate* at about 80% or more of the maximum scale range (Table 2), the only way to obtain large D^2 scores on such characteristics when paired with an average partner is not by endorsing these items even more highly (e.g., at 100% of the maximum scale range), but by endorsing them at a very low level (e.g., at 0%–20% of the maximum scale range). For highly

scores are correlated with positive outcomes, it thus is regularly indicative of the effects of simply having more desirable characteristics rather than the effects of similarity (Baird et al., 2006).

Reliability considerations—As noted earlier, distinctive similarity indices likely have lower variability and reliability than corresponding 'overall similarity' indices (Biesanz, 2010; Funder, 2001; Kenny & Acitelli, 1994). Because low reliability increases the risk of false acceptance of the null hypothesis, this might lead some investigators to avoid removing the normative profile and use overall similarity indices.

This is inadvisable for several reasons. First, there are ways to boost reliability. Most importantly, researchers should use long profiles when estimating similarity. For instance, computing a profile correlation using only Big Five scores would result in a correlation based on five items which should be exceptionally unstable. Disaggregating multi-item scales to their constituent items, as we have done here, should generally increase reliability. Second, distinctive similarity indices should not be avoided just because using them effectively is hard. If one is interested in making valid inferences about correlates of similarity, distinctive indices almost *need* to be used to avoid important and pervasive confounds. In a sense, this is a trade-off between lower reliability and substantially-compromised discriminant validity. Third, although the reliability of distinctive similarity indices may be lower than ideal, those indices often correlate with variables in a sensible manner which have the capacity to reveal deep psychological insights. As shown here and elsewhere (e.g., Furr & Wood, 2013; Beisanz, West, & Millevoi, 2007), distinctive similarity indices can even correlate with other variables in a *more* sensible manner than do overall indices. Finally, the reliability of similarity indices should simply be reported. Unfortunately, this is rarely done, almost certainly because the development of tools for estimating the reliability of profile correlations is fairly recent (Sherman & Wood, 2014).

Should we use measures with less extreme items?—An alternative approach to reducing normativeness is to remove highly desirable and consensually endorsed items from our measures. Indeed, this has been recognized and suggested as a possible practice (Bäckström, Björklund, & Larsson, 2009; Paunonen & LeBel, 2012). This would by definition reduce normativeness. In the extreme, as we noted, if the normative profile has almost no variance (i.e., all items have nearly equal means), an individual's expected level of normativeness will be low (Krueger, 1998), which will eliminate the first condition underlying the NDC.

We think this is the psychometric equivalent of throwing the baby out with the bathwater. A trait's desirability strongly predicts its correlation with being liked by others (Leising et al., 2013; Wood, 2015; Wortman & Wood, 2011). Therefore, the most extreme items on a psychological inventory generally have the greatest associations with relational outcomes. Consequently, constructing inventories by removing highly endorsed and unendorsed items

normative characteristics such as these, the highest D^2 scores necessarily indicate not just "the dyad is dissimilar on this dimension" but that "one of the dyad members has highly undesirable levels on this dimension." (Baird et al., 2006) show that this issue generalizes to standard deviation measures of variability across more than two profiles (e.g., variability in agreeableness levels across a half dozen life contexts or interaction partners).

of the sort shown in Tables 2 and 3 might solve the NDC problem via the unfortunate route of only measuring characteristics that are relatively inconsequential. As we have detailed, there are many effective means of addressing the issues created by the NDC. Discarding some of the most valuable items from the inventory does not need to be one of them.

Conclusion

We expect two themes to emerge as overall similarity indices are replaced by indices that account for the NDC in examinations of the correlates of similarity. First, such correlates should look more sensible. For example, as we have shown, overall similarity indices can indicate that dyadic similarity correlates more with life satisfaction than with dyadic satisfaction – implying that roommate similarity improves roommate satisfaction, but it even more greatly improves one's satisfaction with life! We demonstrated that such effects arise because overall similarity indices generally reflect a person's own desirability, adjustment, or prosociality more so than similarity per se. Attention to the NDC should reduce such strange effects.

Second, we expect that similarity indices will show associations with desirable outcomes such as adjustment and relationship positivity which are substantially smaller than when the NDC is not accounted for. However, we do not expect all benefits of similarity to disappear when distinctive indices are used. For instance, as we reported, distinctive estimates of personality judgment agreement are associated with perceivers' sense of knowing and liking targets. However, these relationships may be considerably reduced.

Apart from the theoretical consequences, misattributing the benefits of desirability to similarity also has important practical consequences. Overall similarity indices might lead us to recommend that individuals seek similar roommates or romantic partners to improve relationship quality and general well-being. However, awareness of the NDC might lead to dramatically different advice: the better route to well-being and relationship satisfaction may be to seek partners who are agreeable, conscientious, emotionally stable, and open to experiences. This may be recommended advice even for individuals who do not have these traits themselves (Watson et al., 2004).

As we have shown, the tendency for common similarity indices to be extremely confounded with desirability creates considerable problems when exploring the putative effects of similarity. It is simply too easy to see how having desirable characteristics can explain correlations found with common similarity indices in ways that have nothing to do with the *matching* or *correspondence* of attributes which is central to the meaning of similarity.

If similarity (or consistency, fit, stability, agreement, accuracy, etc) truly matters, it must matter beyond the effects we can attribute to simply having desirable characteristics. This requires shifting from overall similarity indices and toward distinctive indices of similarity, or indices and methodologies that otherwise control for normativeness. At the moment, appropriate adjustments are made inconsistently; as they are increasingly used, our understanding of the nature of similarity constructs is likely to change substantially in a range of literatures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Katherine Rogers for helpful comments on an earlier draft of this manuscript. We would also like to thank Nicholas Brown, Peter Harms, Bell Jones, Ryne Sherman, Christian Waugh, and the Oregon Research Institute for providing normative data used in some analyses in Study 2. This work was supported, in part, by the National Institute of Mental Health Grant R01 MH70571.

References

Acitelli LK, Kenny DA, Weiner D. The importance of similarity and understanding of partners' marital ideals to relationship satisfaction. Personal Relationships. 2001; 8:167–185.

Aiken, LS.; West, SG. Multiple regression: Testing and interpreting interactions. Thousand Oaks, CA: Sage; 1991.

Anderson C, Keltner D, John OP. Emotional convergence between people over time. Journal of Personality and Social Psychology. 2003; 84:1054–1068. [PubMed: 12757148]

Anderson NH. Likableness ratings of 555 personality-trait words. Journal of Personality and Social Psychology. 1968; 9:272. [PubMed: 5666976]

Bäckström M, Björklund F, Larsson MR. Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. Journal of Research in Personality. 2009; 43:335–344.

Baird BM, Le K, Lucas RE. On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. Journal of Personality and Social Psychology. 2006; 90:512–527. [PubMed: 16594835]

Barni D, Knafo A, Ben-Arieh A, Haj-Yahia MM. Parent–child value similarity across and within cultures. Journal of Cross-Cultural Psychology. 2014; 45:853–867.

Bem DJ, Allen A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. Psychological Review. 1974; 81:506–520.

Biesanz JC. The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. Multivariate Behavioral Research. 2010; 45:853–885. [PubMed: 26795267]

Biesanz JC, Human LJ. The cost of forming more accurate impressions accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. Psychological Science. 2010; 21:589–594. [PubMed: 20424106]

Bleidorn W, Kandler C, Riemann R, Angleitner A, Spinath FM. Genetic and environmental influences on personality profile stability: Unraveling the normativeness problem. Journal of Personality. 2012; 80:1029–1060. [PubMed: 22092116]

Block, J. The Q-sort method in personality assessment and psychiatric research. Thomas Springfield; IL: 1961.

Boer D, Fischer R, Strack M, Bond MH, Lo E, Lam J. How shared preferences in music create bonds between people: Values as the missing link. Personality and Social Psychology Bulletin. 2011:1159–1171. [PubMed: 21543650]

Borkenau P, Zaltauskas K. Effects of self-enhancement on agreement on personality profiles. European Journal of Personality. 2009; 23:107–123.

Busemeyer JR, Jones LE. Analysis of multiplicative combination rules when the causal variables are measured with error. Psychological Bulletin. 1983; 93:549–562.

Byrne, DE. The attraction paradigm. Vol. 11. Academic Press; 1971.

Campbell JD, Assanand S, Di Paula A. The structure of the self-concept and its relation to psychological adjustment. Journal of Personality. 2003; 71:115–140. [PubMed: 12597239]

Carlson EN, Furr RM. Resolution of meta-accuracy: Should people trust their beliefs about how others see them? Social Psychological and Personality Science. 2013; 4:419–426.

Caspi A, Herbener ES. Continuity and change: Assortative marriage and the consistency of personality in adulthood. Journal of Personality and Social Psychology. 1990; 58:250–258. [PubMed: 2319443]

Cattell RB. The three basic factor-analytic research designs—their interrelations and derivatives. Psychological Bulletin. 1952; 49:499–520. [PubMed: 12993927]

Clement RW, Krueger J. Liking persons versus liking groups: a dual-process hypothesis. European Journal of Social Psychology. 1998; 28:457–469.

Clement RW, Krueger J. Social categorization moderates social projection. Journal of Experimental Social Psychology. 2002; 38:219–231.

Cohen J. rc: A profile similarity coefficient invariant over variable reflection. Psychological Bulletin. 1969; 71:281–284. [PubMed: 5790938]

Cohen P, Cohen J, Aiken LS, West SG. The problem of units and the circumstances for POMP. Multivariate Behavioral Research. 1999; 34:315–346.

Colvin CR. “Judgable” people: Personality, behavior, and competing explanations. Journal of Personality and Social Psychology. 1993; 64:861–873. [PubMed: 8505714]

Cronbach LJ. Processes affecting scores on “understanding of others” and“ assumed similarity. Psychological Bulletin. 1955; 52:177–193. [PubMed: 14371889]

Cronbach, LJ. Proposals toward the analytic treatment of social perception scores. In: Tagiui, R.; Petrullo, L., editors. Person perception. Stanford, CA: Stanford University Press; 1958. p. 353-378.

Cronbach LJ, Gleser GC. Assessing similarity between profiles. Psychological Bulletin. 1953; 50:456–473. [PubMed: 13112334]

Deal JE, Halverson CF, Wampler KS. Parental similarity on child-rearing orientations: Effects of stereotype similarity. Journal of Social and Personal Relationships. 1999; 16:87–102.

Decuyper M, De Bolle M, De Fruyt F. Personality similarity, perceptual accuracy, and relationship satisfaction in dating and married couples. Personal Relationships. 2012; 19:128–145.

Donahue EM, Robins RW, Roberts BW, John OP. The divided self: Concurrent and longitudinal effects of psychological adjustment and social roles on self-concept differentiation. Journal of Personality and Social Psychology. 1993; 64:834–846. [PubMed: 8505712]

Donnellan MB, Conger RD, Burzette RG. Personality development from late adolescence to young adulthood: Differential stability, normative maturity, and evidence for the maturity-stability hypothesis. Journal of Personality. 2007; 75:237–263. [PubMed: 17359238]

Edwards, AL. The social desirability variable in personality assessment and research. Ft Worth, TX: Dryden; 1957.

Edwards JR. Problems with the use of profile similarity indices in the study of congruence in organizational research. Personnel Psychology. 1993; 46:641–665.

Edwards JR. Ten difference score myths. Organizational Research Methods. 2001; 4:265–287.

Edwards JR, Parry ME. On the use of polynomial regression equations as an alternative to difference scores in organizational research. Academy of Management Journal. 1993; 36:1577–1613.

Fredrickson BL, Tugade MM, Waugh CE, Larkin GR. What good are positive emotions in crisis? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. Journal of Personality and Social Psychology. 2003; 84:365–376. [PubMed: 12585810]

Funder, DC. Three trends in current research on person perception: Positivity, realism, and sophistication. In: Hall, JA.; Bernieri, FJ., editors. Interpersonal sensitivity: Theory and measurement. Mahwah, NJ: Erlbaum; 2001. p. 319-331.

Funder DC, Furr RM, Colvin CR. The Riverside Behavioral Q-sort: A tool for the description of social behavior. Journal of Personality. 2000; 68:451–489. [PubMed: 10831309]

Furler K, Gomez V, Grob A. Personality similarity and life satisfaction in couples. Journal of Research in Personality. 2014; 47:369–375.

Furr RM. Framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. Journal of Personality. 2008; 76:1267–1316. [PubMed: 18705644]

Furr RM. Profile analysis in person-situation integration. Journal of Research in Personality. 2009; 43:196–207.

Furr RM. The double-entry intraclass correlation as an index of profile similarity: Meaning, limitations, and alternatives. Journal of Personality Assessment. 2010; 92:1–15. [PubMed: 20013451]

Furr RM, Dougherty DM, Marsh DM, Mathias CW. Personality judgment and personality pathology: Self-other agreement in adolescents with conduct disorder. Journal of Personality. 2007; 75:629–662. [PubMed: 17489894]

Furr RM, Funder DC. Situational similarity and behavioral consistency: Subjective, objective, variable-centered, and person-centered approaches. Journal of Research in Personality. 2004; 38:421–447.

Furr RM, Wood D. On the similarity between exchangeable profiles: A psychometric model, analytic strategy, and empirical illustration. Journal of Research in Personality. 2013; 47:233–247. [PubMed: 24039314]

Gerstorf D, Windsor TD, Hoppmann CA, Butterworth P. Longitudinal change in spousal similarities in mental health: Between-couple and within-couple perspectives. Psychology and Aging. 2013; 28:540–554. [PubMed: 23795766]

Goldberg, LR. From ace to zombie: Some explorations in the language of personality. In: Spielberger, CD.; Butcher, JN., editors. Advances in personality assessment. Vol. 1. Hillsdale, NJ: Erlbaum; 1982. p. 203-234.

Gonzaga GC, Campos B, Bradbury T. Similarity, convergence, and relationship satisfaction in dating and married couples. Journal of Personality and Social Psychology. 2007; 93:34–48. [PubMed: 17605587]

Gonzaga GC, Carter S, Buckwalter JG. Assortative mating, convergence, and satisfaction in married couples. Personal Relationships. 2010; 17:634–644.

Hoch SJ. Perceived consensus and predictive accuracy: The pros and cons of projection. Journal of Personality and Social Psychology. 1987; 53:221–234.

Human LJ, Biesanz JC. Through the looking glass clearly: Accuracy and assumed similarity in well-adjusted individuals' first impressions. Journal of Personality and Social Psychology. 2011; 100:349–364. [PubMed: 21299315]

Izard C. Personality similarity and friendship. The Journal of Abnormal and Social Psychology. 1960; 61:47–51. [PubMed: 14406216]

John, OP.; Naumann, LP.; Soto, CJ. Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In: John, OP.; Robins, RW.; Pervin, LA., editors. Handbook of personality: Theory and research. 3. New York: Guilford; 2008. p. 114-158.

Karniol R. Egocentrism versus protocentrism: The status of self in social prediction. Psychological Review. 2003; 110:564–580. [PubMed: 12885115]

Kellermann K. The negativity effect and its implications for initial interaction. Communication Monographs. 1984; 51:37–55.

Kenny DA, Acitelli LK. Measuring similarity in couples. Journal of Family Psychology. 1994; 8:417–431.

Kenny DA, Albright L. Accuracy in interpersonal perception: a social relations analysis. Psychological Bulletin. 1987; 102:390–402. [PubMed: 3317468]

Kenny, DA.; Kashy, DA.; Cook, WL. Dyadic data analysis. New York: Guilford; 2006.

Klimstra TA, Luyckx K, Hale WWI, Goossens L, Meeus WHJ. Longitudinal associations between personality profile stability and adjustment in college students: Distinguishing among overall stability, distinctive stability, and within-time normativeness. Journal of Personality. 2010; 78:1163–1184. [PubMed: 20545819]

Krueger J. On the perception of social consensus. Advances in Experimental Social Psychology. 1998; 30:164–240.

Leary MR, Allen AB. Self-presentational persona: Simultaneous management of multiple impressions. Journal of Personality and Social Psychology. 2011; 101:1033–1049. [PubMed: 21688923]

Leising D, Erbs J, Fritz U. The letter of recommendation effect in informant ratings of personality. Journal of Personality and Social Psychology. 2010; 98:668–682. [PubMed: 20307137]

Leising D, Ostrovski O, Zimmermann J. "Are we talking about the same person here?": Interrater agreement in judgments of personality varies dramatically with how much the perceivers like the targets. Social Psychological and Personality Science. 2013; 4:468–474.

Locke KD, Zheng D, Smith J. Establishing commonality versus affirming distinctiveness: Patterns of personality judgments in China and the United States. Social Psychological and Personality Science. 2014; 5:389–397.

Luo S, Klohnen EC. Assortative mating and marital quality in newlyweds: A couple-centered approach. Journal of Personality and Social Psychology. 2005; 88:304–326. [PubMed: 15841861]

Luo S, Snider AG. Accuracy and biases in newlyweds' perceptions of each other: Not mutually exclusive but mutually beneficial. Psychological Science. 2009; 20:1332–1339. [PubMed: 19818046]

McCrae, RR.; Costa, PTJ. The five-factor theory of personality. In: John, OP.; Robins, RW.; Pervin, LA., editors. Handbook of personality: Theory and research. 3. New York: Guilford; 2008. p. 159-181.

McCrae RR, Terracciano A. Personality profiles of cultures: Aggregate personality traits. Journal of Personality and Social Psychology. 2005; 89:407–425. [PubMed: 16248722]

McGuire, WJ. Suspiciousness of experimenter's intent. In: Rosenthal, R.; Rosnow, RL., editors. Artifacts in Behavioral Research. New York: Oxford; 2009. p. 15-47.

McNulty JK, Fincham FD. Beyond positive psychology? Toward a contextual view of psychological processes and well-being. American Psychologist. 2012; 67:101–110. [PubMed: 21787036]

Nicholls JG, Licht BG, Pearl RA. Some dangers of using personality questionnaires to study personality. Psychological Bulletin. 1982; 92:572–580.

Nunnally, JC.; Bernstein, IH. Psychometric theory. New York: McGraw; 1991.

Ozer DJ, Benet-Martínez V. Personality and the prediction of consequential outcomes. Annual Review of Psychology. 2006; 57:401–421.

Park N, Peterson C, Seligman ME. Character strengths in fifty-four nations and the fifty US states. The Journal of Positive Psychology. 2006; 1:118–129.

Paunonen SV, LeBel EP. Socially desirable responding and its elusive effects on the validity of personality assessments. Journal of Personality and Social Psychology. 2012; 103:158–175. [PubMed: 22564012]

Pavot WG, Diener E, Colvin CR, Sandvik E. Further validation of the Satisfaction With Life Scale: Evidence for the cross-method convergence of well-being measures. Journal of Personality Assessment. 1991; 57:149–161. [PubMed: 1920028]

Roberts BW, Caspi A, Moffitt TE. The kids are alright: Growth and stability in personality development from adolescence to adulthood. Journal of Personality and Social Psychology. 2001; 81:670–683. [PubMed: 11642353]

Robins RW, Fraley RC, Roberts BW, Trzesniewski KH. A longitudinal study of personality change in young adulthood. Journal of Personality. 2001; 69:617–640. [PubMed: 11497032]

Rogers, KH.; Biesanz, JC. Knowing versus liking: Separating normative knowledge from social desirability in personality impressions. University of British Columbia; Vancouver: 2015.

Saucier, G. Public-domain measures of generalized attitude dispositions. 2014. Retrieved from http://pages.uoregon.edu/gsaucier/gsau4b.htm

Saucier G, Thalmayer AG, Payne DL, Carlson R, Sanogo L, Ole-Kotikash L, Szarota P. A basic bivariate structure of personality attributes evident across nine languages. Journal of Personality. 2014:1–14.

Schwartz SH, Bardi A. Value hierarchies across cultures taking a similarities perspective. Journal of Cross-Cultural Psychology. 2001; 32:268–290.

Sherman RA, Nave CS, Funder DC. Situational similarity and personality predict behavioral consistency. Journal of Personality and Social Psychology. 2010; 99:330–343. [PubMed: 20658847]

Sherman RA, Nave CS, Funder DC. Situational construal is related to personality and gender. Journal of Research in Personality. 2013; 47:1–14.

Sherman RA, Wood D. Estimating the expected replicability of a pattern of correlations and other measures of association. Multivariate Behavioral Research. 2014; 49:17–40. [PubMed: 26745671]

Shoda Y, Mischel W, Wright JC. Intraindividual stability in the organization and patterning of behavior: Incorporating psychological situations into the idiographic analysis of personality. Journal of Personality and Social Psychology. 1994; 67:674–687. [PubMed: 7965613]

Stephenson, W. The study of behavior; Q-technique and its methodology. Chicago, IL: University of Chicago Press; 1953.

Suh EM. Culture, identity consistency, and subjective well-being. Journal of Personality and Social Psychology. 2002; 83:1378–1391. [PubMed: 12500819]

Terracciano A, McCrae RR, Costa PT. Intra-individual change in personality stability and age. Journal of Research in Personality. 2010; 44:31–37. [PubMed: 20305728]

Townsend SS, Kim HS, Mesquita B. Are you feeling what I'm feeling? Emotional similarity buffers stress. Social Psychological and Personality Science. 2013

Unkelbach C, Fiedler K, Bayer M, Stegmüller M, Danner D. Why positive information is processed faster: The density hypothesis. Journal of Personality and Social Psychology. 2008; 95:36–49. [PubMed: 18605850]

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Holt RA. The sequence of the human genome. Science. 2001; 291:1304–1351. [PubMed: 11181995]

Watson D, Klohnen EC, Casillas A, Nus Simms E, Haig J, Berry DS. Match makers and deal breakers: Analyses of assortative mating in newlywed couples. Journal of Personality. 2004; 72:1029–1068. [PubMed: 15335336]

Wood D. Testing the lexical hypothesis: Are socially important traits more densely reflected in the English lexicon? Journal of Personality and Social Psychology. 2015; 108:317–335. [PubMed: 25603378]

Wood D, Brumbaugh CC. Using revealed mate preferences to evaluate market force and differential preference explanations for mate selection. Journal of Personality and Social Psychology. 2009; 96:1226–1244. [PubMed: 19469598]

Wood D, Nye CD, Saucier G. Identification and measurement of a more comprehensive set of person-descriptive trait markers from the English lexicon. Journal of Research in Personality. 2010; 44:258–272.

Wood D, Wortman J. Trait means and desirabilities as artifactual and real sources of differential stability of personality traits. Journal of Personality. 2012; 80:665–701. [PubMed: 22092250]

Wortman J, Wood D. The personality traits of liked people. Journal of Research in Personality. 2011; 45:519–528.

Wortman J, Wood D, Furr RM, Fanciullo J, Harms PD. The relationship between actual and perceived similarity. Journal of Research in Personality. 2014; 49:31–46.

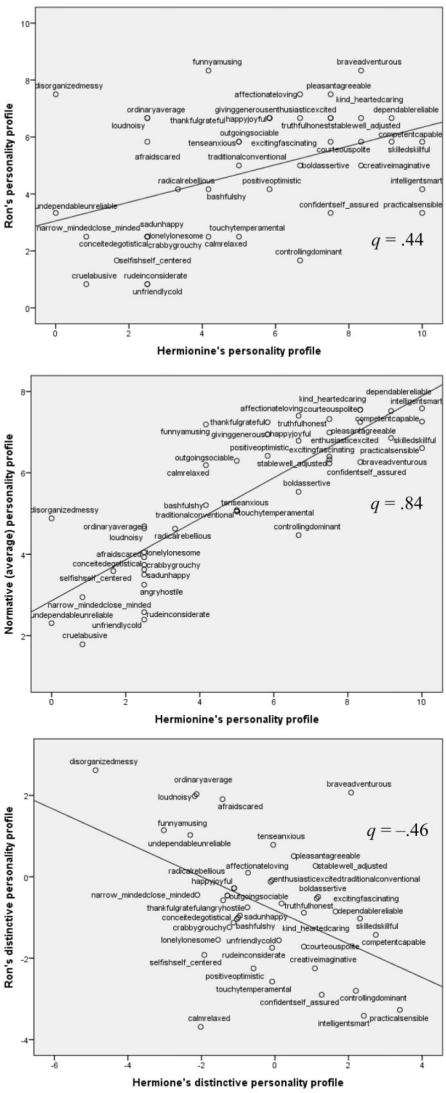

Figure 1.

1A: Estimated “overall profile similarity” between Hermione and Ron.

1B: Similarity between Hermione’s profile and the normative (average) personality profile.

1C: Estimated “distinctive profile similarity” between Hermione after removing the normative profile

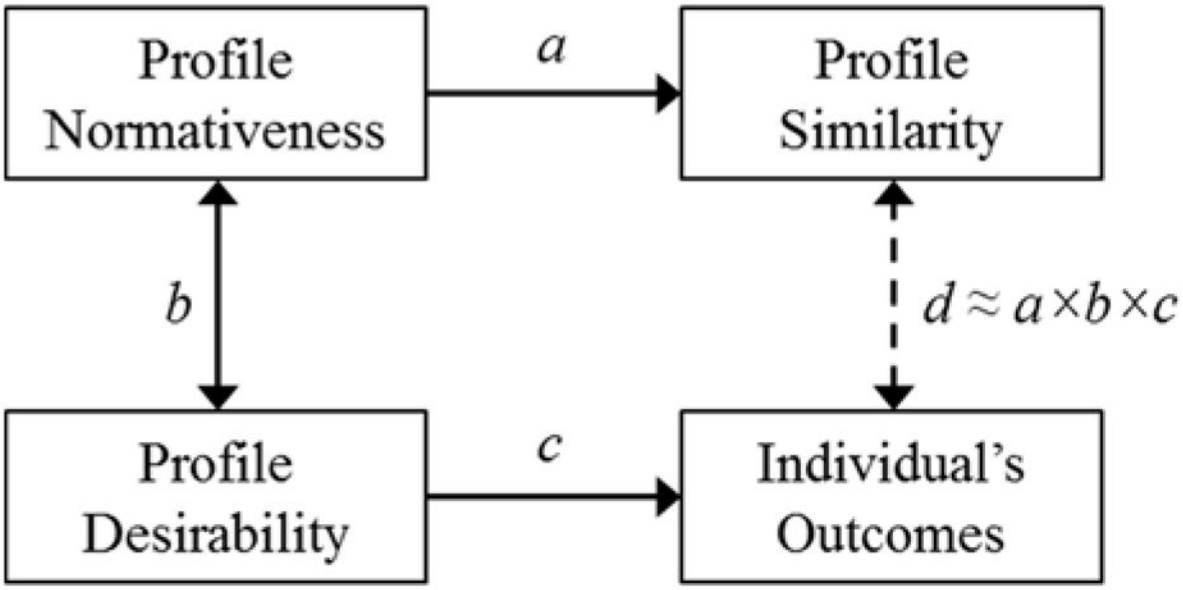

Figure 2.

Illustration of how similarity as indexed via a profile correlation may be associated with individual difference or dyadic outcome variables through the three-part Normative-Desirability Confound (NDC) rather than due to an effect of the individual being more similar to their dyadic partner than expected by chance. Specifically: 'overall similarity' will frequently correlate with an individual's outcomes (Path *d*) in large part because (1) normative individuals are more expected to be similar to anyone (Path *a*), (2) more normative individuals tend to have more desirable profiles (Path *b*), and (3) more desirable individuals tend to have more positive outcomes (Path *c*).

Table 1

Relationships between IIDL Overall and Distinctive Profile Similarity, Stability, and Judgment Agreement Estimates with Measures of Mental Health, Well-Being, and Relationship Quality

Variable	Mean (SD)	α	1	2	3	4	5	6	7	8
Profile Similarity Measure (q)										
1. Normativeness: ($q_{\bar{Y}Y_1}$)	.81* (.15)	*.75*								
2. Desirability (q_{DY_1})	.78* (.18)	*.81*	.97*							
3. Roommate Similarity ($q_{Y_1R_1}$)	.68* (.18)	*.71*	.63*	.61*						
4. Distinctive Rmt. Similarity ($q_{Y*_1R*_1}$)	.10* (.27)	*.39*	.04	.06	.42*					
5. Stability ($q_{Y_1Y_2}$)	.87* (.12)	*.63*	.55*	.55*	.31*	−.02				
6. Distinctive Stability ($q_{Y*_1Y*_2}$)	.64* (.19)	*.16*	−.01	.03	−.10	.00	.61*			
7. Judgment Agreement ($q_{Y_1P(Y)1)}$)†	.60* (.28)	*.87*	.39*	.38*	.28*	−.09	.26*	.04		
8. Distinctive J. Agreement ($q_{Y*_1P(Y)*_1}$)†	.21* (.27)	*.48*	−.02	−.02	−.01	.08	−.04	−.04	.45*	
Outcome Measures										
MCMI Narcissism	1.44 (.12)	--	.05	.18*	.00	−.07	.16	.21*	.04	.05
MCMI Paranoid	1.13 (.15)	--	−.37*	−.36*	−.38*	−.15	−.13	−.04	−.29*	−.02
MCMI Avoidant	1.16 (.18)	--	−.55*	−.61*	−.42*	.02	−.27*	−.03	−.32*	.00
MCMI Borderline	1.19 (.21)	--	−.59*	−.57*	−.46*	−.04	−.18	.01	−.30*	.09
Satisfaction with Life	4.09 (.67)	*.86*	.33*	.36*	.28*	−.01	.01	−.05	.14*	−.06
Depression	2.03 (1.00)	*.87*	−.46*	−.50*	−.35*	−.10	−.27*	−.02	−.21*	.09
Rmt. Relationship Positivity	4.58 (.48)	*.88*	.16*	.16	.20*	.10	−.07	−.06	.16*	.07
Perceived Similarity to Rmt.	3.90 (.99)	--	.11	.11	.17*	.20*	.03	−.07	.04	.03
Rater's Perceived Sim. to Participant†	3.02 (1.28)	--	.14*	.14*	.03	−.09	.03	.02	.52*	.13*
Rater's Sense of Knowing Participant†	3.60 (1.24)	--	.03	.03	−.02	−.04	−.07	−.03	.24*	.17*
Rater's Liking of Participant†	4.22 (1.02)	--	.11	.10	.08	−.06	−.07	−.01	.69*	.18*

Note. Subscripts of *q* indicate the profiles being correlated. *D* indicates the profile of item desirabilities, *Y1* indicates participant self-ratings from Time 1 and *Y2* from Time 2, *R1* indicates ratings from the participant's roommate at Time 1, *P(Y)1* indicates ratings of the participant by another observer in Year 1, and *Y*, R*, P(Y)** indicates the same distinctive profiles. All correlations are based on *N* ranging from 93 to 151.

$^{*}p < .05$.

†indicates that correlations with these variables are tested for significance using a multilevel model, where separate ratings of the individual by different raters are not aggregated.

"--" indicates values are unavailable or scale is a single item. Note that some of the relationships involving rater perceptions of similarity in this dataset have been explored in Wortman et al. (2014).

Pers Soc Psychol Rev. Author manuscript; available in PMC 2017 May 01.

Table 2

Nature of Normative Self-Rating Profiles in Various Personality Trait Inventories

Inventory (sample)	$q_{\bar{Y}Y_i}$	$q_{D\bar{Y}}$	q_{DY_i}	Least Normative Items $_{\text{POMP M(Y)/M(D)}}$	Most Normative Items $_{\text{POMP M(Y), M(D)}}$
IIDL (online)	.67	.95	.63	cruel, abusive$_{9/5}$/narrow-minded, close-minded$_{17/11}$/rude, inconsiderate$_{19/11}$	intelligent, smart$_{82/88}$/kind-hearted, caring$_{79/84}$/truthful, honest$_{78/85}$
BFI (college freshman)	.60	.86	.51	Starts quarrels with others$_{17/16}$/Is depressed, blue $_{20/18}$/Can be cold and aloof$_{29/27}$	Is a reliable worker$_{87/80}$/Is considerate & kind to almost everyone$_{82/81}$/Is curious about many different things$_{82/83}$
TIPI (online)	.52	.90	.47	Conventional, uncreative$_{27/23}$/Disorganized, careless$_{39/22}$/Critical (to others), quarrelsome$_{43/24}$	Sympathetic, warm$_{78/81}$/Open to new experiences, complex$_{75/81}$/Dependable, self-disciplined$_{71/81}$
PDA (community)	.76	.87	.66	homeless$_{2/4}$, retarded$_{5/14}$, evil$_{5/1}$, good-for-nothing$_{5/4}$/insane$_{5/9}$,/dumb$_{6/4}$/corrupt$_{6/4}$	reliable$_{93/88}$, trustworthy$_{93/93}$, honest$_{92/97}$/faithful$_{91/87}$/truthful$_{91/91}$
CAQ (undergraduates)	.62	.87	.54	Is guileful and deceitful, manipulative, opportunistic$_{17/14}$/Feels cheated and victimized by life; self-pitying$_{18/11}$	Responds to humor$_{90/91}$/Is a genuinely dependable and responsible person$_{88/88}$
BFAS (community)	.64	.93	.60	Insult people$_{9/23}$/Seek conflict$_{14/21}$/Take advantage of others$_{15/19}$	Enjoy the beauty of nature$_{94/83}$/Show my feelings when I'm happy$_{83/79}$/Like order$_{82/66}$
NEO-FFI (undergraduates)	.51	.85	.43	I believe letting students hear controversial speakers can only confuse and mislead them$_{15/20}$/ Some people think of me as cold and calculating$_{18,14}$/I often get into arguments with my family & co-workers$_{28/17}$	I generally try to be thoughtful and considerate$_{83/93}$/I laugh easily$_{82/90}$/I try to be courteous to everyone I meet$_{82/91}$/I work hard to accomplish my goals$_{80/92}$
HEXACO-PI (non-BPD sample)$_a$	.46	.84	.38	It wouldn't bother me to harm someone I didn't like$_{18/29}$/I'd be tempted to use counterfeit	I have sympathy for people who are less fortunate than I am$_{81/76}$/If someone has cheated me once, I
HEXACO-PI (BPD sample)$_a$	.41	.53	.22	money, if I were sure that I could get away with it$_{21/28}$/I do only the minimum amount of work needed to get by$_{23/26}$	will always feel suspicious of that person$_{76/70}$/I think that most people like some aspects of my personality$_{76/70}$

Note. $q\bar{Y}Y_i$ indicates average *normativeness*: the expected correlation between a single individual (Y_i) and the normative profile ($\bar{Y}$). $q_{D\bar{Y}}$ indicates the observed correlation between the normative profile and desirability profile (D), and q_{DY_i} the expected correlation between a single individual's profile and the desirability profile, estimated as $q_{DY_i} = q\bar{Y}Y_i \times q_{D\bar{Y}}$. In the last two columns, item endorsement and desirability means are given in subscripts in POMP metric (Percentage of Maximum Possible; range of 0 to 100). Subscript "a" indicates that the items shown from the HEXACO represent most/least normative items within the non-BPD sample.

Pers Soc Psychol Rev. Author manuscript; available in PMC 2017 May 01.

Table 3

Nature of Normative Profiles in Additional Types of Inventories

Type of Measure (Inventory)	$q_{\bar{Y}Y_i}$	$q_{D\bar{Y}}$	q_{DY_i}	Least Normative Items $_{\text{POMP M(Y)/M(D)}}$	Most Normative Items $_{\text{POMP M(Y)/M(D)}}$
Behaviors (RBQ)	.73	.74	.54	Seems detached from the interaction$_{23/13}$/Talks at rather than with partner(s)$_{24/10}$/Acts irritated$_{25/10}$/Behaves in a fearful, timid manner$_{28/25}$/Tries to undermine, sabotage, or obstruct$_{28/0}$	Exhibits social skills$_{77/95}$/Engages in constant eye contact with partner(s)$_{76/75}$/Interviews his or her partner(s)$_{75/43}$/Is talkative$_{74/68}$/Seems interested in what partner(s) has to say$_{74/88}$/Smiles frequently$_{73/85}$
Situation characteristics (RSQ)	.45	.57	.26	P is being abused$_{10/2}$/Situation contains physical threats$_{17/10}$/Independence of P is questioned or threatened$_{18/19}$/Situation contains emotional threats$_{20/14}$	Talking is permitted$_{75/69}$/Situation is simple & clear-cut$_{72/76}$/Situation is potentially enjoyable$_{69/81}$/A job needs to be done$_{64/43}$/Social interaction is possible$_{63/69}$
Emotions (DES)	.63	.90	.57	ashamed, humiliated, disgraced$_{26/5}$/repentant, guilty, blameworthy$_{32/13}$/disgust, distaste, revulsion$_{32/10}$/contemptuous, scornful, disdainful$_{32/18}$	glad, happy, joyful$_{75/93}$/amused, fun-loving, silly$_{74/95}$/hopeful, optimistic, encouraged$_{72/88}$/interested, alert, curious$_{71/88}$/love, closeness, trust$_{71/90}$
Attitudes (SDI-46)	.66	.85	.56	Worldly possessions are the greatest good and the highest value in life$_{8/35}$/People of different races and nationalities should live in different places apart from one another$_{10/3}$/I believe in the superiority of my own ethnic group$_{11/13}$/What is good can only be judged by the gratification of the senses$_{20/38}$	My own race is not superior to any other race$_{90/93}$/I believe in government by law with the consent of those people governed$_{88/80}$/People ought to be motivated by something beyond their own self-interest$_{88/85}$/I am in favor of a constitutional form of government$_{87/85}$/There is a higher good than the pleasures of the senses$_{80/73}$
Values (SVS)	.61	NA	NA	Social power$_{18}$/Detachment$_{38}$/Daring$_{45}$/Authority$_{45}$/Preserving my public image$_{47}$	Family security$_{90}$/Honest$_{90}$/Healthy$_{89}$/Self-respect$_{89}$/Freedom$_{86}$/Responsible$_{86}$
Music preferences (STOMP)	.46	NA	NA	Folk$_{35}$/Religious$_{37}$/Heavy metal$_{39}$/Country$_{41}$	Rock$_{79}$/Alternative$_{73}$/Soundtracks$_{72}$/Pop$_{68}$
Vocational Interests (O*NET Interest Profiler)	.37	NA	NA	Monitor a machine on an assembly line$_{21}$/Stamp, sort, and distribute mail for an organization$_{24}$/Drive a truck to deliver packages to offices and homes$_{25}$/Lay brick or tile$_{28}$/Raise fish in a fish hachery$_{30}$	Start your own business$_{69}$/Teach a high-school class$_{61}$/Play a musical instrument$_{61}$/Teach children how to read$_{60}$/Do volunteer work at a non-profit organization$_{59}$

Note. $q_{\bar{Y}Y_i}$ indicates average *normativeness*: the expected correlation between a single individual (Y_i) and the normative profile ($\bar{Y}$), $q_{D\bar{Y}}$ the observed correlation between the normative and desirability profiles (D), and q_{DY_i} the expected correlation between a single individual's profile and the desirability profile, estimated as $q_{DY_i} = q_{\bar{Y}Y_i} \times q_{D\bar{Y}}$. In the last two columns, item endorsement and desirability means are given in subscripts in POMP metric (Percentage of Maximum Possible; range of 0 to 100). "P" indicates the participant in the situation or interaction. "NA" indicates desirability ratings of stimuli were not collected.