Article

JSPR

Journal of Social and Personal Relationships
2018, Vol. 35(1) 89–111

DOI: 10.1177/0265407517712901
journals.sagepub.com/home/spr
SAGE

Assessing accuracy in close relationships research: A truth and bias approach

Chadly Stern[1] and Tessa V. West[2]

Abstract

Do people accurately perceive their romantic partner? What are the implications of perceiving one's partner accurately or inaccurately? These questions are frequently debated and researched in psychology, and so researchers need to have methods for assessing accuracy that are flexible enough to answer different theoretical questions. Researchers frequently utilize a variety of different approaches to assess two different forms of accuracy: mean-level bias and correlational accuracy. The main goal of this article is to provide recommendations for the best approaches that relationship researchers can use to assess these types of accuracy. We focus on statistical approaches employing advances in multilevel modeling and, in particular, how West and Kenny's Truth and Bias model can be especially useful for testing questions of bias and accuracy in perceptions in close relationships. We provide step-by-step approaches of how to implement the models we outline.

Keywords

Accuracy, bias, close relationships, social perception, Truth and Bias model

> *What types of accuracy can researchers assess in close relationships, and can these different types be examined in the same statistical model? How can researchers examine both accuracy and bias in perceptions of one's relationship partner?*

[1] University of Illinois, Urbana-Champaign, USA
[2] New York University, USA

Corresponding author:
Chadly Stern, University of Illinois, Urbana-Champaign, 603 E. Daniel Street, Champaign, IL 61820, USA.
Email: chadly@illinois.edu

Questions of how people come to perceive the world and whether those perceptions are accurate have been debated and researched in a variety of disciplines. Psychologists have taken an empirical approach to addressing the question of accuracy in person perception and have focused primarily on whether people accurately perceive *other people* (e.g., Gage & Cronbach, 1955; Taft, 1955; for a review see Hall, Schmidt-Mast, & West, 2016). For example, psychologists have asked questions related to whether people accurately perceive others' emotional states (Ickes, 1997), personalities, (Funder & Colvin, 1988), and group memberships (Tskhay & Rule, 2013).

One area of work that has received a large amount of empirical attention concerns how accurate people are in their perceptions of their close relationships partners. For example, researchers have examined whether people have "positive illusions" or "unrealistic idealizations" about their relationship partners (Murray, Holmes, & Griffin, 1996a), and whether perceivers can be both biased in their judgments (e.g., drawn toward seeing their partner as having a positive mood) and accurate (e.g., accurately detect changes in their partner's mood; see Fletcher & Kerr, 2010 and Gagne & Lydon, 2004). Accuracy has also been treated as a predictor of important relationship outcomes, such as satisfaction (Murray, Holmes, & Griffin, 1996b) and security (Lemay, 2014).

Close relationship researchers frequently use different approaches to assess accuracy and often raise legitimate questions concerning how to best statistically assess accuracy. Additionally, when discussing factors that relate to whether perceivers make accurate judgments, researchers frequently assume that biased judgments are necessarily inaccurate ones, without considering the possibility that bias in some cases could facilitate accuracy. A main goal of this article is to provide recommendations for the best approaches that relationship researchers can use to assess questions of accuracy in close relationships (although the approaches discussed here could be utilized in any area of research) as well as how to examine the relationship between bias and accuracy in perceptions.

In this article, we discuss two types of accuracy that close relationship researchers commonly examine: mean-level bias and correlational accuracy. We discuss the conceptual meaning of each type of accuracy, provide recommendations for the types of questions that mean-level bias and correlational accuracy can be used to answer, outline step-by-step procedures that researchers can follow to assess these types of accuracy, and note the limitations of approaches to assessing each type of accuracy. We then dedicate the remainder of the article to discussing a more recently developed model for assessing accuracy: The Truth and Bias (T&B) model of judgment (West & Kenny, 2011). We discuss the conceptual background of this model as a means of examining accuracy and outline the benefits of using this model over other approaches. We additionally outline in a step-by-step manner how researchers can implement the T&B model in their own work.

Mean-level bias and correlational accuracy

Mean-level bias and correlational accuracy address theoretically different questions (Fletcher & Kerr, 2010). The mean-level bias approach examines the extent to which perceivers' responses diverge from the truth. For example, the question of whether people consistently overestimate or underestimate their partner's satisfaction concerns mean-level bias. The correlational accuracy approach examines the extent to which perceivers' responses correspond to or "track" the truth. For example, the question of how much people's estimates of their partner's relationship satisfaction correspond with their partner's self-reported satisfaction concerns correlational accuracy.

It is possible to examine accuracy across the whole sample (referred to as a "nomothetic" approach) and also compute accuracy scores for each individual (referred to as an "idiographic" approach). For example, a researcher might ask whether people in general accurately perceive their relationship partners (e.g., Fletcher & Kerr, 2010), in which case accuracy could be assessed on the level of the sample. A researcher might also be interested in whether accuracy is associated with certain personality characteristics or relationship qualities, in which case they would want to calculate accuracy scores for each individual in their sample. In this article, we describe how researchers can calculate mean-level bias and correlational accuracy both on the level of the sample and the level of the individual. We outline procedures using multilevel modeling to examine accuracy in close relationships because it provides clear benefits for researchers. First, in dyadic contexts (e.g., romantic couples), a researcher is interested in assessing accuracy for *both people* in the couple. The scores of the people in the couple are likely nonindependent, meaning that the partner's scores are correlated with each other. Although nonindependence can be assessed for any variable that both members of the dyad provide responses on (predictors and outcome variables), we are primarily concerned with nonindependence in the outcome variables. Failing to account for the nonindependence in outcome variables can lead to biased standard errors for any of the fixed effects parameters in the model (and in turn significance tests). Although a full description of how standard errors can be biased is beyond the scope of this article, it is possible for both Type I and Type II errors to occur, depending on the type of predictor variable (between-dyad, within-dyad, or mixed), the direction of the nonindependence (positive or negative) as well as the strength of the nonindependence (Kenny, Kashy, & Cook, 2006). When researchers have accounted for the nonindependence in dyads in the past, they have used a variety of different approaches, such as repeated-measures analysis of variance (Hall, Rosip, LeBeau, Horgan, & Carter, 2006), structural equation modeling (Murray et al., 1996b), and providing correlations on the level of the individual that adjust for the dyad-level mean (Ickes, Stinson, Bissonnette, & Garcia, 1990). We seek to provide a straightforward approach using multilevel modeling in which close relationships researchers can account for nonindependence.

Second, scholars have emphasized the importance of including both fixed and random effects in statistical analyses (e.g., Judd, Westfall, & Kenny, 2012). As we elaborate throughout this article, multilevel modeling provides researchers the opportunity to specify different types of random effects (depending on their design and dyad types) as well as fixed effects that test the role of constructs that are theoretically expected to influence judgments and accuracy. Altogether, the methods we outline provide a high level of flexibility in what researchers are able to include and test in their model.

In the first half of this article, we describe models that researchers can use to separately test for mean-level bias and correlational accuracy. We provide these models for readers who may only be interested in examining one type of accuracy and note the limitations of these approaches. We then turn to the T&B model, in which we outline how these two forms of accuracy can be simultaneously examined and discuss additional questions that can be examined in the truth and bias model.

Example data set

In the first half of the article, we will use the following fictitious data set to conceptually outline and demonstrate how to assess mean-level bias and correlational accuracy using multilevel modeling. This data set is included in the excel sheet in the online supplemental materials, and syntax for all analyses described using this data set is included in the online supplement. Imagine that a researcher is interested in assessing relationship satisfaction in couples where each person belongs to a different political party. These data are considered *distinguishable* dyads because all dyads can be ordered based on a meaningful dichotomous variable (i.e., political party, because all dyads contain one Democrat and one Republican; see Iida, Seidman, & Shrout, In Press, and Kenny et al., 2006, for further discussion). If the data were *indistinguishable*, some or all of the dyads could not be ordered based on a meaningful variable. For instance, if the data set contained some couples of the same political party, then the dyads would be considered indistinguishable. Note here that *all* dyads in the data set must be distinguishable in order for the data to be analyzed as distinguishable dyadic data.

The researcher recruited 10 married heterosexual couples and assessed the relationship satisfaction of both people in the couple once a month over the course of a year. Each month, the researcher also asked participants to rate their partner's satisfaction. Participants provided both their own satisfaction and their perception of their partner's satisfaction on a 1 (*not at all satisfied*) to 7 (*very satisfied*) scale. For each person in the relationship, this procedure results in 12 ratings of actual satisfaction and 12 perceptions of one's partner's satisfaction (see West & Kenny, 2011, on the importance of measuring the truth and the judgment using the same scale). The number of couples and repeated assessments in this hypothetical data set were arbitrarily selected for the sake of example. However, researchers collecting data for this type of study—and dyadic studies in general—should consider both the number of dyads and repeated measurements to maximize statistical power in their design (Lane & Hennes, In Press).

In this example, we use relationship satisfaction assessed over time. However, all of the methods described below could also be used if researchers assessed ratings on multiple items at the same time point (e.g., satisfaction with different aspects of a relationship), as the different items would be considered the "repeated measure." It is also important to note that although we use an example with reciprocal ratings (e.g., each person in the couple rates one another's satisfaction), the models we outline can be used for nonreciprocal ratings as well (e.g., each person in a romantic couple estimates the number of years that they have been together).

Popular political strategists James Carville (a liberal Democrat) and his wife Mary Matalin (a conservative Republican) hypothetically participated in this study. The data for James and Mary are shown in Table 1. The first column indicates the number of the couple. James and Mary were the first couple in the study and so are given a value of 1. The person column indicates a code value for each person in the relationship,

Table 1. Hypothetical perceived and self-reported relationship satisfaction for James (Person 1) and Mary (Person 2) over the course of 1 year.

Couple	Person	Month	Perception of partner's satisfaction	Partner's self-reported satisfaction	Perceived – self-reported satisfaction
1	1	1	4	3	1
1	1	2	3	2	1
1	1	3	4	2	2
1	1	4	5	3	2
1	1	5	2	1	1
1	1	6	3	4	1
1	1	7	5	3	2
1	1	8	5	4	1
1	1	9	6	4	2
1	1	10	4	4	0
1	1	11	3	2	1
1	1	12	5	3	2
1	2	1	3	3	0
1	2	2	4	4	0
1	2	3	2	5	−3
1	2	4	5	2	3
1	2	5	3	4	−1
1	2	6	4	3	1
1	2	7	3	5	−2
1	2	8	3	4	−1
1	2	9	2	3	−1
1	2	10	4	2	2
1	2	11	2	3	−1
1	2	12	4	3	1

distinguished by political party. Democrats receive a value of 1 and Republicans a value of 2. The month column indicates the month, which in this example ranges 1–12. The "perception of partner's satisfaction" column indicates the person's judgment of their partner's satisfaction in that month, and the "partner's self-reported satisfaction" column indicates the partner's self-reported satisfaction in that month. James's perceptions of Mary's satisfaction and Mary's actual satisfaction are also plotted across the 12 months in Figure 1. We next discuss how a researcher can assess mean-level bias and correlational accuracy in these data. We provide sample syntax for conducting all models we outline using the MIXED procedure in SPSS in the online supplemental materials (version 24).

Mean-level bias

A researcher might be interested in examining how much people's perceptions of reality diverge from actual reality, such as whether perceptions of a partner's satisfaction diverge from a partner's actual levels of satisfaction. For example, does James tend to overestimate how satisfied Mary is with their relationship? Mean-level bias allows

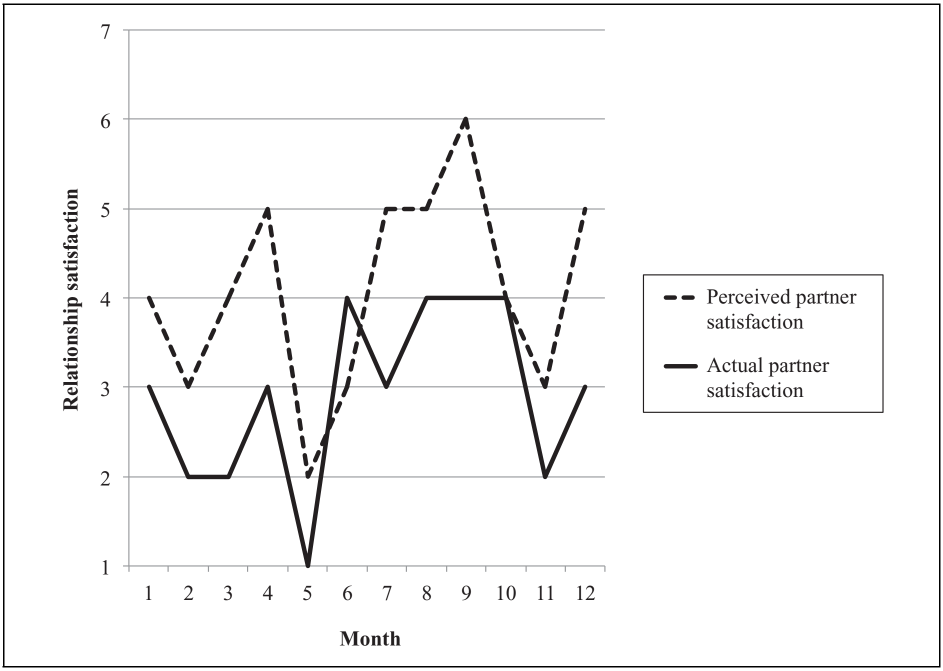

Figure 1. James' perceptions of Mary's satisfaction and Mary's actual satisfaction over the course of a year.

researchers to test this question. This approach simply consists of subtracting the truth from the perception (Fletcher & Kerr, 2010; Overall, Fletcher, & Kenny, 2012). This calculation is shown in Table 1 in the "Perceived—Self-Reported Satisfaction" column. Positive numbers indicate overestimations, negative numbers indicate underestimations, and a score of zero indicates no bias.

Level of the sample

A researcher might be interested in the question of whether people in the sample as a whole underestimate or overestimate their partner's level of satisfaction. The most basic equation for testing mean-level bias on the level of the sample can be represented as follows:

$$y_{ij} = b_{0j} + E_{ij}.$$

Consistent with notation outlined by Kenny and Kashy (2011), we use the subscript i to refer to an individual and j to refer to a dyad. y_{ij} is the difference between the perception and the truth for person i in dyad j. For example, in Table 1, this would consist of James' scores in the last column; b_0 is the intercept. The intercept indicates the value of mean-level bias when all predictor values are zero. In this model, there are no fixed effects (i.e., it is a "null model"). Testing the intercept against zero indicates whether the average mean-level bias score for the sample is different from zero. This model includes a random intercept. In this example, we estimate a "common" intercept on the level of the dyad. In turn, the random intercept estimates whether there is a significant amount of variability between dyads (which we refer to as "dyadic variability") in the outcome variable (i.e., mean-level bias). The error variance (E_{ij}) estimates variability in the outcome variable on the level of the individual.

When both members of the dyad provide data for the same repeated measures or time points, as they did here, the level of repeated measure is the same for both members of the dyad. As such, the repeated measure and person are crossed rather than being nested. Researchers sometimes make the mistake of assuming a three-level model in which time points are nested within persons and persons within dyads. However, when both members of the dyad provide data for the same repeated measures (e.g., they both make ratings on day 1, day 2, etc.), the errors between partners at each repeated measure and within each partner across the repeated measures are estimated. Only if dyad members complete data at different repeated measures should a three-level model be assumed. When a three-level model is assumed, the correlation between the dyad members' scores at the same repeated measure is assumed to be zero (e.g., the correlation between the husband's error on day 1 and the wife's error on day 1), which is often not the case. Assuming the wrong error structure can lead to biased standard errors, which can lead to incorrect p values for tests of fixed effects (Kenny et al., 2006).

For all models outlined in this article, the researcher can specify a heterogeneous compound symmetry variance–covariance matrix when the dyads are distinguishable (as in the present example). This structure of covariance matrix estimates separate variances of the residuals along the distinguishing characteristic of the dyads (e.g., all of the Democrats have the same variance and all of the Republicans have the same variance) and assumes that the covariance of the residuals between people are the same across couples (Fitzmaurice, Laird, & Ware, 2012). If the dyads were indistinguishable (e.g., same-party couples), the researcher would specify a compound symmetry covariance matrix, which assumes that both the variance and the covariance of the two people in the couple are the same across couples. In other words, a constraint is placed on the variances by forcing the error variances of the two partners to be the same.

In this example, the intercept value (.11) is positive but not significant ($p = .39$), indicating that people displayed a nonsignificant trend toward overestimating their partner's level of relationship satisfaction. The random effect of the intercept was significant, indicating that there was a significant amount of variability in mean-level bias across dyads. The significance of the random effect suggests that it could be fruitful to examine factors that shape individual variability in mean-level bias (as described in the next section).

In addition to what we have described earlier, researches could examine the within-couple covariance of accuracy, which tests whether a person's accuracy is correlated with their partner's accuracy (i.e., if I am accurate, is my partner also accurate?). West and Kenny (2011) provide details for this more elaborate model. This model is appropriate for researchers who are interested in studying accuracy as a within-couple process.

Level of the individual

It is also possible for a researcher to calculate mean-level bias scores for each individual in the sample. For example, imagine that a researcher wants to examine whether mean-level bias in perceiving a partner's relationship satisfaction is associated with individual difference variables (e.g., extraversion, attachment styles, political party membership). The most basic version of this question can be represented as follows:

$$y_{ij} = b_{0j} + b_{1j}X_{ij} + E_{ij}$$

y_{ij} is the mean-level bias score for person i in dyad j. Calculating this score consists of subtracting the truth from the perception for each time point and then creating an average score for each person (see excel sheet in online supplement). Alternatively, if the researcher only assessed satisfaction at one time point (instead of 12), this score would simply consist of subtracting the truth from the perception. b_0 is the intercept, and b_1 is a fixed effect of the individual difference variable or situational variable. In this example, it is extraversion, which is measured on a 1 (*not at all*) to 7 (*very much so*) scale. If this variable is continuous, it should be grand-mean centered. X_{ij} is the extraversion score of person i in dyad j. This model includes a random intercept, which tests whether there is a significant amount of dyadic variability in mean-level bias.

In this example, the intercept value is positive (.16) but is not significant ($p = .23$), indicating that people displayed a nonsignificant trend to overestimate their partner's level of satisfaction. The fixed effect of extraversion is positive (.30) and significant ($p < .001$), indicating that higher levels of extraversion were associated with greater overestimation of a partner's relationship satisfaction. The random effect of the intercept was also significant, indicating that after accounting for a person's level of extraversion there was still a significant amount of variability in mean-level bias.

The individual-level approach also provides the unique advantage of allowing the researcher to use mean-level bias scores in mediation models or to estimate structural models. For example, a researcher might want to test the question of whether extraversion predicts the likelihood of staying in one's relationship and whether this association is in part attributable to how people overestimate their relationship partner's satisfaction. The researcher can test these types of questions using modern tools that simplify the process of testing for mediation and are discussed in detail elsewhere (for single-level models: Hayes, 2013; for multilevel models: Bauer, Preacher, & Gil, 2006).

It is important to note that calculating individual mean-level bias scores requires researchers to collapse across the repeated unit of measurement. As such, if researchers assess judgments of the same item across multiple time points (as in the present example), they should only examine how factors that remain highly constant over time, such as personality variables, explain individual variability in accuracy. Otherwise, they would ignore important variability in factors that greatly change over time (e.g., variability in perceptions of conflict). However, if researchers collapse across perceptions on multiple attributes within the same time point, they could examine the role of any individual difference variable, as they would not be collapsing over time. If researchers did wish to examine how mean-level bias changes over time as a function of other variables that also change over time (e.g., perceived conflict), they would need to conduct a three-level model. The complexity of this model is beyond the scope of the present article, and we refer interested readers to Kenny and Kashy (2011) for a discussion of more complex usages of multilevel models in dyadic contexts.

Limitations of difference score approach

The difference score approach described earlier has limitations of what one can infer from the mean-level bias scores. Specifically, a small discrepancy between the perception and the truth (i.e., low mean-level bias) could be attributable to similar response biases among the perceiver and the target in how they make ratings, rather than the perceiver's ability to accurately appraise the target (Cronbach, 1955). For example, it is possible that James has a small mean-level bias score in perceiving Mary's satisfaction simply because they both have response biases to report higher numbers on a scale. Additionally, when examining factors that predict or are associated with difference scores, any observed association could be attributable to one part of the difference score producing the observed effect (e.g., Laird & Weems, 2011). For example, extraversion could be associated with a difference score of perceived—actual relationship satisfaction in part because of simply being associated with perceived partner satisfaction. These are important caveats to keep in mind (see Castro-Schilo & Grimm, In Press, for further discussion).

Correlational accuracy

A researcher might also be interested in examining whether people's perceptions tend to "hang with" or "track" the correct answer. Correlational accuracy tests this question (Fletcher & Kerr, 2010; Overall et al., 2012). For example, the question of whether James can detect the highs and lows of Mary's satisfaction is a test of correlational accuracy. As with mean-level bias, correlational accuracy can also be assessed on the level of the sample and the individual.

Level of the sample

Examining correlational accuracy on the level of the sample can be statistically represented as follows:

$$y_{ij} = b_{0j} + b_{1j}X_{ij} + E_{ij}.$$

y_{ij} is the judgment made by person i in dyad j, which in this example is the perception of one's partner's satisfaction (e.g., James' perception of Mary). b_0 is the intercept, and X_{ij} is the correct answer, which in this example is the self-reported satisfaction (grand-mean centered) of the partner of person i in dyad j (e.g., Mary's self-reported satisfaction). The fixed effect of b_1 estimates the relationship between the judgment and the truth, and so tests for correlational accuracy. A positive value would indicate that people's perceptions systematically track the truth (e.g., people know when their partner is relatively more or less satisfied), a negative correlation indicates that perceptions systematically diverge from the truth (e.g., people perceive their partner as relatively high in satisfaction when they are low and vice-versa), and a score of zero indicates that there is no systematic association between perceptions and the truth (e.g., perceptions of satisfaction do not in any way correspond to actual satisfaction).

This model includes a common random intercept, which tests whether there is a significant amount of dyadic variability in perceptions of a partner's satisfaction. Because there are several assessments of perceived and actual satisfaction, we also include a random effect of actual satisfaction, which tests whether there is variability in correlational accuracy across dyads.

To account for the nonindependence of the people in the couple, the researcher would conduct a two-level crossed model, as described in the mean-level bias section above. If the researcher had measured relationship satisfaction at one time point (instead of 12), they would instead conduct a two-level model where person is nested within dyad (person is level 1 and dyad is level 2).

If the dyads are distinguishable (as in this example), the researcher would specify a heterogeneous compound symmetry covariance matrix (as in the previously described model), which allows for separate error variances for each partner. If the dyads are indistinguishable, the researcher would need to force the errors to be the same for the two partners, as it is arbitrary who is Partner 1 and Partner 2 in the couple. The researcher would estimate the random effects in the same way as the distinguishable case, but an extra step is needed to force constraints on the variance–covariance matrix. West (2013) and Kenny and Kashy (2011) provided syntax in SAS for this method, and Olsen and Kenny (2006) illustrate this method in structural equation modeling. It is important to note that this analytic strategy estimates the nonindependence of the residuals of the dependent variable (i.e., the intraclass correlation). In turn, specifying the judgment or the correct answer as the dependent variable will change the effects. We would recommend specifying perceived satisfaction (i.e., the judgment) as the dependent variable.

In this example where satisfaction is assessed over 12 months, the intercept is positive (4.09) and significant ($p < .001$), which simply indicates that the average estimate of a partner's satisfaction is above zero when at the mean of actual relationship satisfaction. The fixed effect of the partner's self-reported satisfaction was negative ($-.05$) but not significant ($p = .52$), which indicates that perceptions of a partner's satisfaction contrasted away from the partner's self-reported satisfaction. In other words, there was not a significant degree of correlational accuracy on the level of the sample. The random effect of the intercept was positive but not significant, indicating that, after accounting for a partner's actual satisfaction, there was not a significant degree of variability in perceptions of satisfaction. The random effect of actual satisfaction was positive and significant, indicating that there was variability in the extent to which perceptions tracked the truth. In other words, some dyads had greater correlational accuracy than others. This random effect is informative for determining whether it would be fruitful to further examine factors that could explain individual variability in accuracy (as explained earlier).

Level of the individual

It is also possible to obtain individual correlational accuracy scores and examine factors (e.g., personality variables) that help to explain individual variability. The most basic version of the question for doing so can be represented as follows:

$$y_{ij} = b_{0j} + b_{1j}X_{ij} + E_{ij}.$$

y_{ij} is the correlational accuracy score for person i in dyad j. b_1 is the fixed effect of an individual difference factor. In this example, the variable is extraversion. X_{ij} is the grand-mean centered extraversion score for person i in dyad j. In this model, there is also a random intercept, which tests whether there is a significant degree of dyadic variability in the outcome variable (i.e., correlational accuracy).

To calculate a correlational accuracy score for each person, there are two requirements that need to be met. First, the perception and the truth need to be assessed at multiple times for the same item (e.g., relationships satisfaction over 12 months) or multiple items at the same time point (e.g., satisfaction with different aspects of the relationship). Second, there needs to be variability in *both* the perception and the truth (i.e., the perception and truth *cannot* be the same for all items or time points; that is, it cannot be a constant). If there is not variability in both the perception and the correct answer, correlational accuracy cannot be assessed because the correlation between the perception and the correct answer will be undefined.

To calculate individual accuracy scores, the researcher would correlate the perception with the truth for each person (see excel sheet in online supplement; see also Krueger & Zeiger, 1993). Before using this score in analyses, the researcher must unbound the correlation by transforming it to a Fisher's Z score. In the rare situation that a person has obtained a score of either 1 or -1 (which cannot be transformed to a Fisher's Z score), we recommend simply replacing that person's correlational accuracy score with the next score in the sample that is closest to 1 or -1, respectively, before conducting r-to-Z transformations.

A positive score indicates that perceptions track the correct answer (e.g., James perceives Mary as higher in satisfaction when she is higher), a negative score indicates that perceptions systematically diverge from the correct answer (e.g., James perceives Mary as lower in satisfaction when she is higher), and a score of zero indicates that there is no systematic association between perceptions and the correct answer (e.g., James' perceptions of Mary's satisfaction are not associated with Mary's self-reported satisfaction). This procedure results in a score of .75 for James, indicating that James' perceptions of Mary's satisfaction track her actual satisfaction.

Using individual scores to test questions of correlational accuracy requires the same analytic strategy as described for assessing individual mean-level bias. In this example, the intercept is negative ($-.07$) but is not significant ($p = .93$), indicating that overall people's perceptions were not associated with the truth. The fixed effect of extraversion was positive (.11) and significant ($p = .02$), indicating that higher levels of extraversion were associated with greater correlational accuracy. The random intercept was also significant, indicating that after accounting for the effect of extraversion, there was still a significant amount of variability in correlational accuracy.

Creating individual-level correlational accuracy scores has the advantage of providing a standardized size of the relationship between two variables. When accuracy on individual time points (or items) is assessed, the overall pattern of effects can often appear to be unsystematic. For example, in Month 4, James' perception of Mary's satisfaction increased and her self-reported satisfaction also increased, but in Month 7, James' perception of Mary's satisfaction increased and her self-reported satisfaction decreased (see Figure 1). Rather than trying to "eyeball" an overall pattern of accuracy across the assessments of satisfaction, the correlation indicates the overall trajectory of accuracy across all assessments.

Limitations of correlational approach

The correlational approach described earlier also possesses limitations in terms of what researchers can conclude from the obtained scores. Specifically, when perceptions of a target are correlated with the truth, correlational accuracy could be attributable to reliance on "stereotypes" or "normative profiles" about the judgment at hand, rather than the perceiver picking up on a pattern for the specific target in question (Biesanz, 2010; Cronbach, 1955; Rogers, Wood, & Furr, In Press). For example, if James is relatively accurate in tracking Mary's relationship satisfaction, it is unclear whether James' judgments are specific to Mary's pattern of satisfaction or whether he relied on a generic stereotype of how satisfaction fluctuates over the course of the year for *anyone* (e.g., James thinks that satisfaction is lowest in the coldest months and highest in the warmest months, and this pattern happens to be true for most people, including Mary). Researchers should keep this caveat in mind when using correlational scores to examine accuracy.

The T&B model

In the first half of this article, we discussed how researchers can separately examine mean-level bias and correlational accuracy. West and Kenny (2011) developed the T&B model as a comprehensive framework that integrates the previously described approaches and advances upon them in several ways. First, the T&B model allows researchers to simultaneously estimate mean-level bias and correlational accuracy. Second, the model allows researchers to potentially overcome some of the aforementioned issues of estimating mean-level and correlational accuracy for perceptions of a single target through statistically adjusting for potential sources of bias in judgments. Third, the model allows researchers to statistically estimate the extent to which accuracy in judgments is attributable to a direct influence of the truth on a judgment and indirect influences of biases on judgments. We elaborate on each of these points in the process of describing the conceptual and statistical aspects of the model subsequently.

Conceptual and definitional aspects of the model

Throughout this section of the article, we use language developed by West and Kenny (2011). Additionally, we provide details of the analytics of the model that are covered in

West and Kenny (2011). To describe the variables that are included in the T&B model, we use the example of James perceiving Mary's satisfaction. An example equation representing the variables that predict James' perception could be as follows:

$$\begin{aligned}\text{James}' \text{ Perception of Mary}'\text{s Satisfaction } &= \text{Mary}'\text{s Self-Reported Satisfaction}\\ &+ \text{James}' \text{ Self-Reported Satisfaction}\\ &+ \text{ Friend}'\text{s Communication about Mary.}\end{aligned}$$

We refer to the variable that measures the correct answer as the *truth variable*, and we refer to any variable that is not the correct answer but that perceivers use in the judgment process as a *bias variable*. In many cases, the truth is challenging (or impossible) to observe directly, such as when the construct of interest is a psychological phenomenon like relationship satisfaction. However, given that it is frequently possible for targets to reveal the truth in a direct way to perceivers, we conceptualize something as a truth variable if it theoretically possesses a one-to-one relationship with the underlying reality. In this example, we make the assumption that a person is honestly stating their relationship satisfaction. As such, the truth variable would be Mary's self-reported satisfaction. Bias variables would be factors other than Mary's satisfaction that James uses to estimate Mary's satisfaction. For example, James might use his own satisfaction to estimate Mary's satisfaction, or Mary's best friend might provide her own opinions to James on whether Mary is satisfied with the relationship. James' own satisfaction and the opinions of Mary's friend are *bias variables* in this example because they are variables that can affect the judgment and that are not the truth (see Table 1 in West & Kenny, 2011, for definitions of truth and bias variables). Importantly, we define a variable as a bias variable independent of whether it is associated with the truth. Bias variables must simply be conceptually distinct from the truth. James' satisfaction might be highly correlated with Mary's satisfaction, and the opinions of Mary's friend might be uncorrelated (or even negatively correlated) with Mary's satisfaction, but these are both bias variables.

In discussing the processes through which perceivers make judgments, we define how strongly the truth variable affects perceivers' judgments as the *truth force*, and we define how strongly a bias variable affects perceivers' judgments as the *bias force*. Mary might have informed James that she is highly satisfied with their relationship, but Mary's friend might have informed James that Mary is not very satisfied. The extent to which James relies on Mary's stated satisfaction would be the strength of the truth force, and the extent to which he relies on the opinions of Mary's friend would be the strength of the bias force.

Both truth and bias forces contribute to the ultimate judgments that perceivers make as well as how accurate those judgments are. Historically, researchers have discussed biased processing strategies as inherently leading to inaccurate judgments (e.g., Jones & Harris, 1967; Ross, Greene, & House, 1977). However, researchers of social perception have recently begun to discuss how bias is not synonymous with error and that perceivers can be both biased and accurate in their judgments (e.g., Gagné & Lydon, 2004; Jussim, 2012; Kenny & Acitelli, 2001). The T&B model conceptually and statistically allows for the possibility that perceivers can achieve accuracy either through having direct access to the truth or through reliance on bias variables. In other words, the T&B model allows for the possibility that perceivers can be both biased and accurate, as well as that perceivers could achieve some degree of accuracy through being biased.

In some cases, accuracy is achieved via direct access to the truth. For example, a romantic partner's actual job satisfaction plays a large role in what people think their partner's satisfaction is (Kenny & Acitelli, 2001), likely because people in a relationship frequently discuss their jobs with one another. In other situations, perceivers do not have direct access to the truth when making judgments. In these cases, people tend to rely on other information that is conceptually distinct from the truth (i.e., bias variables) to make judgments. For example, people often rely on their own sexual satisfaction to infer their romantic partner's satisfaction (Kenny & Acitelli, 2001), potentially because couples are too shy to discuss their sex life and so do not possess direct access to their partner's actual satisfaction. People could also rely on stereotypes and nonverbal cues to make inferences about their partner's sexual satisfaction, which would be considered bias variables because, for example, gender stereotypes about sexual satisfaction or nonverbal cues that relate to a person's satisfaction are not a person's actual satisfaction. In this example, we view cues like nonverbal behaviors as bias variables because they are likely to be derived from multiple factors (rather than solely the underlying truth), and they are unlikely to be conceptually redundant with the truth variable in question.

Relying on bias variables to make judgments can lead people to indirectly achieve accuracy when the bias variable and truth are positively correlated. In other words, whether or not a bias variable will lead to accurate judgments depends on the validity of the bias. For example, if the perceiver's own sexual satisfaction is strongly associated with their partner's satisfaction, then relying on their own feelings to assess their partner's feelings will lead them to make accurate judgments through being biased.

It is important to point out that the bias and truth variables could be correlated for several different reasons. The bias variable could be correlated with the truth variable simply due to chance or because of a shared causal source. The truth variable could also be associated with the bias variable because it has a direct causal influence on the bias variable (e.g., a person's relationships satisfaction has a causal influence on their partner's level of satisfaction). However, our argument is simply that if the bias variable is positively correlated with the truth variable for any reason, then relying on the bias variable to make judgments will lead to enhanced accuracy.

The T&B model also allows for moderator variables, which are defined as variables that influence the strength of the truth and bias forces. For example, how close people are in a relationship might influence how strongly they rely on truth and bias variables to estimate their partner's relationship satisfaction. James might feel very close to Mary and so more heavily rely on her stated satisfaction (the truth variable) than her friend's assessment of Mary's satisfaction (a bias variable) to estimate Mary's satisfaction.

To conduct a truth and bias analysis, the researcher at minimum (generally) needs to have specified the truth variable, the bias variable, and the judgment being made. Depending on the researcher's question of interest, they might also include moderator variables in their analysis. Interested readers are referred to Stern, West, and Schoenthaler (2013) for a more in-depth theoretical discussion of each of these types of variables and a step-by-step approach of how to conceptually and operationally define these variables.

Overcoming issues in difference score and correlational approaches

As mentioned in the previous sections on limitations of difference score and correlational approaches, individual variability in response bias and usage of general forms of knowledge (e.g., stereotypes) can obscure the interpretation of whether people are accurately perceiving their relationship partner. Researchers have proposed multiple statistical approaches to addressing these issues including polynomial regression (Laird & De Los Reyes, 2013) and response surface analysis (Barranti, Carlson, & Furr, 2016). Some statistical approaches of how to address these issues require people to make ratings of multiple targets (e.g., Kenny's [1994] social relations model and Biesanz's [2010] social accuracy model). Given that people often have one partner in a close relationship (e.g., being in a monogamous romantic relationship, having one college roommate), and this coupling (sometimes) lasts for a considerable length of time, obtaining ratings of multiple targets that are within the same conceptual category (e.g., romantic partner) is frequently not feasible. Additionally, obtaining multiple ratings through having perceivers rate other people's relationship partners (i.e., using pseudo-couples) creates conceptual differences between targets that can obscure interpretation of results (but see Rogers et al., In Press, for benefits of this approach).

We wish to emphasize that a primary purpose of the T&B model is to simultaneously examine different forms of accuracy and also to address the factors that shape people's judgments. Previously mentioned caveats for using difference score and correlational approaches should still be kept in mind when using the T&B model. However, in the context of using the T&B model, components of the model may also help researchers to overcome some potential issues of examining accuracy for ratings of one target through the inclusion of bias variables. Specifically, a researcher can include bias variables in the model that help to account for response bias and usage of normative profiles. This process requires the researcher to use theory to determine which bias variables should be included. In terms of response bias, for example, a researcher might think that social desirability shapes response bias on their scales and so could include a measure of social desirability as a bias variable in the model. Doing so would both adjust for response bias when examining accuracy and statistically estimate its influence on the judgment. In terms of normative profiles, for example, if previous research has found that men and women differ in their normative profiles of relationship satisfaction (both in their perceptions of satisfaction and actual satisfaction), gender could be included as a bias variable.

Another advantage of the T&B model is that it can simultaneously estimate mean-level bias and correlational accuracy in the same model. Examining mean-level bias separate from correlational accuracy assumes that the effect of the truth on the judgment is 1 (West & Kenny, 2011). However, the truth force is unlikely to ever be one. As such, a benefit of the T&B approach is that it is possible to simultaneously calculate mean-level bias and correlational accuracy while accounting for one another.

Relation to other models of social perception

It is important to briefly note how the T&B model relates to other models of social perception. The T&B model has some conceptual similarities to other models. For example, the bias force is analogous to cue and attribute utilization processes that are central to other models of social perception, such as Brunswick's (1955) lens model, Funder's (1995) realistic accuracy model, Jussim's (1991) reflection-construction model, and Zebrowitz and Collin's (1997) Gibsonian approach. Because the T&B model shares some conceptual overlap with other models of social perception, the T&B model can be used to test components of multiple models. However, the T&B model is also unique from other models in two key ways. First, models of social perception traditionally argue that one's environment is perceived entirely in an indirect manner through the usage of cues (e.g., Brunswick, 1955). In other words, they assume that the truth force is zero. In contrast, the T&B model allows for the possibility that the truth has a direct effect on the outcome. Second, the T&B model translates theoretical factors that contribute to social perception into a statistical model. In turn, the model allows researchers to statistically estimate the relative influence of both truth and bias variables on judgments.

Overview of sections on the statistical model

We next introduce the statistical aspects of the T&B model. First, we discuss the model in its most basic form and outline the different components that are typically included in the model. Second, we provide an example using the most basic form of the model and discuss how researchers should interpret the parameters in the model. Third, we discuss how researchers can add moderators to the model to examine both main and interaction effect predictions. Fourth, we outline how researchers can estimate both direct and indirect accuracy in the model. These first four sections focus on using the model to examine effects on the level of the sample. As such, in the final section, we discuss how researchers can also use the model to calculate individual scores for each person in their sample.

The statistical model

The most basic version of the T&B model can be written as follows:

$$y = b_0 + tT + bB + E.$$

y is the value of the outcome variable. The outcome variable can be any variable that the researcher is interested in assessing, and for the sake of example in this article, we refer to the outcome variable as a judgment. b_0 is the intercept, which indicates the value of the judgment when all predictor values are zero. If the truth, bias, and judgment variables are all assessed on the same scale, then subtracting the mean of the truth variable from all variables renders the intercept an assessment of *directional bias*. In other words, the test of the intercept against zero concerns how much the average judgment deviates from the average value of the truth. When the model is assessed for the entire sample, directional bias is analogous to mean-level bias on the level of the sample. t represents the truth force, or how strongly the truth is associated with the judgment, and T represents a person's score on the truth variable. When the model is assessed for the entire sample, the truth force is analogous to correlational accuracy on the level of the sample. b represents the bias force, or how strongly the bias variable is associated with the judgment, and B represents a person's score on the bias variable. It is important to note that although we have included only one bias variable in this equation, researchers should include all of the bias variables that they believe impact the judgment and are theoretically relevant for their research question. E is the error variance, which consists of both systematic and nonsystematic variance in the judgment that is not explained by the truth or bias variables.

Examining the role of truth and bias variables on the judgment

We next describe how to carry out a truth and bias analysis by elaborating on example analyses and results initially presented in West and Kenny (2011), Cases 1 and 2. The data set used consists of 65 pairs of college roommates. Each roommate provided reports twice a week for 5.5 weeks concerning (a) how hurt they felt by their roommate and (b) how hurt they thought their roommate felt on a 1 (*not at all*) to 7 (*very much*) scale. This data set provides self-ratings of hurt feelings and perceptions of one's roommates hurt feelings. As such, in each roommate pair, each person both perceives their roommate and is perceived by their roommate. We refer to the person who is making a rating of their roommate as the perceiver roommate and the person who is being rated as the target roommate. The dyads are indistinguishable, and so the two partners cannot be ordered on a meaningful variable.

We examine factors that shape judgments of the target roommate's feelings. Because the judgment is of the target's feelings, the truth variable is the target's self-report of how hurt they feel. The bias variable is the perceiver's feelings, as it is conceptually distinct from the truth (the target's feelings) and could be expected to impact judgments of the target's feelings (via assumed similarity). In turn, we test how the target's self-reported hurt feelings (the truth variable) and the perceiver's self-reported feelings (the bias variable) shape judgments of the target's hurt feelings (the dependent variable). We subtracted the grand-mean of target roommate's feelings from perceiver roommate's feelings, target roommate's feelings, and judgments of target roommate's feelings. In doing so, the intercept is now a test of directional bias. Plugging these variables in the truth and bias equation reported above, the equation can be written as follows:

$$\begin{aligned}\text{Judgment of Target's Feelings } &= \text{Directional Bias } + t(\text{Target's Feelings}) \\ &\quad + b(\text{Perceiver's Feelings}) \ + \ \text{Error.}\end{aligned}$$

In a dyadic context, the researcher would conduct this analysis for both people in the relationship (e.g., roommates' perceptions of one another) and so would need to use a procedure that accounts for the nonindependence in judgments of roommates' feelings. The target's and perceiver's feelings would be entered as fixed effects, and judgments of the target's feelings would be entered as the dependent variable. The researcher would specify a compound symmetry variance–covariance matrix. Because the dyads are indistinguishable, in this model, we estimate random effects for each dyad (rather than an effect for each person), which we describe in detail in the online supplement. Sample syntax for conducting a truth and bias analysis for this model in SAS can be found in the online supplement. For reference, we have also created syntax for a distinguishable case in the online supplement.

The resulting estimate of directional bias (the intercept) in this model is −.03. This value represents the difference between the mean of targets' self-reported feelings and the mean of perceivers' estimates of targets' feelings. The significance test for the intercept indicates whether directional bias is different from zero. In this example, it is not. As such, the test of directional bias indicates that perceivers' judgments of their roommates' feelings do not, on average, systematically diverge from their roommates' actual feelings. In other words, perceivers, on average, did not display directional bias in their judgments.

The truth force parameter (.19) is significant ($p = .02$). This value indicates that the truth is uniquely associated with perceivers' judgments. The positive value of the truth force indicates that people's judgments are falling in accordance with the truth (e.g., roommates with hurt feelings are being perceived as having hurt feelings). This finding also indicates that perceivers, on average, have correlational accuracy in their judgments.

The bias force parameter (.60) is significant ($p < .001$). The positive value indicates that perceivers' judgments are falling in accordance with the bias variable. In this example, the judgment and the bias variable are both hurt feelings, one for each roommate. As such, a positive bias force indicates that perceivers are assuming similarity: Roommates with hurt feelings are perceiving their roommate as hurt.

Now that the parameters of the model have been obtained, the researcher can place them in the equation to predict perceivers' judgments of their roommates:

$$\begin{aligned}\text{Judgment of Target Roommate's Feelings} = &-.03 + .19(\text{Target's Feelings})\\ &+ .60(\text{Perceiver's Feelings}) + \text{Error}.\end{aligned}$$

For example, a perceiver whose roommate self-reported being very hurt (a score of 7) but personally did not feel hurt (a score of 2) would be predicted to judge their roommate as not having very hurt feelings (a score of 2.5):

$$2.5 = -.03 + .19(7) + .60(2).$$

This process could be used to plot the trajectory of judgments among the possible scores of perceiver and target roommate's feelings.

Adding a moderator variable

The first equation provides researchers with the most basic pieces of information that can be obtained from a truth and bias analysis: information about direct accuracy (the truth force parameter) and how influential bias variables are in shaping perceivers' judgments (the bias force parameter). However, researchers can also conduct more complex models to test whether moderator variables shape the strength of the truth and bias forces. For example, does relationship closeness impact how strongly perceivers rely on their own feelings to estimate their roommates' feelings? To illustrate this idea, we use the same example as above and add relationship closeness as a moderator variable, which we represent as M:

$$y = b_0 + tT + bB + mM + nTM + pBM + E.$$

In this example, n is the parameter estimate of the interaction term between the moderator and truth force, p is the parameter estimate of the interaction term between the moderator and bias force, and M is the person's score on the moderator variable. Unlike the truth and bias variables in the model, the researcher should *not* subtract the mean of the truth from the moderator variable. Instead, the moderator variable should be grand mean centered. Using the variables in this example, the equation can be rewritten as follows:

$$\begin{aligned} \text{Judgment of Target's Feelings } = {} & \text{Directional Bias } + t(\text{Target's Feelings}) \\ & + b(\text{Perceiver's Feelings}) \\ & + m(\text{Relationship Closeness}) \\ & + n(\text{Target's Feelings} * \text{Relationship Closeness}) \\ & + p(\text{Perceiver's Feelings} * \text{Relationship Closeness}) \\ & + \text{ Error.} \end{aligned}$$

To obtain the parameter estimates for this equation, the researcher can conduct a model in the same way as for the previous equation (i.e., a MIXED model for both roommates' perceptions of one another). The directional bias estimate (.02) is non-significant ($p = .87$), which indicates that, on average, perceptions of targets' feelings did not significantly diverge from targets' self-reported feelings (i.e., there is not a significant degree of mean-level bias). The truth force parameter (.22) is positive and significant ($p = .05$), which indicates that targets' actual feelings were associated with how they were perceived (i.e., there is correlational accuracy). The bias parameter (.72) is positive and significant ($p < .001$), which indicates that perceivers viewed their roommates as having similar feelings as they personally did. The closeness parameter (.08) is positive but not significant ($p = .41$), indicating that feelings of closeness, on average, were not associated with judgments of targets' feelings. The interaction term between closeness and the truth force ($-.05$) was also not significant ($p = .52$), indicating that relationship closeness did not significantly impact how strongly targets' self-reported feelings were associated with perceptions of their feelings. In other words, people high and low in relationship closeness perceived targets' feelings in a manner that tracked targets' self-reported feelings to a similar extent. However, the interaction term between closeness and the bias force (.16) was positive and significant ($p = .03$), indicating that perceivers' own feelings were more strongly associated with perceptions of targets' feelings as relationship closeness increased. In other words, people high in relationship closeness were more likely to perceive their roommates as having similar feelings as themselves than were people low in relationship closeness.

Assessing indirect accuracy

We next discuss how researchers can determine the extent to which the bias force contributes to accuracy in judgments, which we refer to as *indirect accuracy*. The researcher would need to conduct a model in which the bias variable is the dependent variable. In doing so, the researcher obtains an estimate of how much concordance there is between the bias variable and the truth variable (e.g., how strongly are perceiver roommates' own feelings associated with target roommates' feelings). The equation with the bias variable as the dependent variable can be represented as follows:

$$B = b_0 + tT + E$$

or, filling in the variable names:

$$\text{Perceiver's Feelings} = \text{Feelings of Average Roommate} + t(\text{Target's Feelings}) + \text{Error.}$$

The truth force parameter ($-.17$) was negative but nonsignificant ($p = .14$), indicating that there was a trending negative association between the truth and the bias variables. In other words, roommates tended to systematically diverge in their feelings (i.e., people who felt hurt had roommates who did not feel hurt). After obtaining this value, the researcher can assess the components of accuracy that are attributable to the truth (direct accuracy) and the bias (indirect accuracy). Direct accuracy is the parameter estimate of the truth force predicting the judgment, which we obtained above and has a value of .19. Indirect accuracy is calculated by multiplying the effect of the bias force on the judgment and the effect of the truth force on the bias. This results in an indirect accuracy of $(.60)*(-.17) = -.10$. The fact that indirect accuracy is negative indicates that the bias variable (the perceiver's own feelings) is associated with greater inaccuracy in judgments. In other words, the more people relied on their own feelings to judge their roommate's feelings, the more inaccurate they became. After the researcher has obtained the values for direct and indirect accuracy, they can then calculate an overall accuracy score that accounts for the unique contributions of direct and indirect factors shaping accuracy. Total accuracy is the sum of the direct and indirect effects, and so in this example, total accuracy is $.19 + (-.10) = .09$.

Individual scores in the truth and bias model

So far, we have only discussed using the T&B model to assess the role of truth and bias forces on the judgment for the *entire sample*. However, it is also possible to use this model to calculate individual scores if each person in a sample made multiple judgments (it would not be possible to calculate parameter estimates for each person if they only made one judgment).

The approach to do so would simply consist of conducting the same models discussed above, but for each person. The researcher would need to keep in mind that the truth, bias, and outcome variables should be centered at the mean of the truth variable *for that person* (rather than for the entire sample). In turn, the intercept in the model would indicate mean-level bias and the truth force parameter would indicate correlational accuracy for that person. These values can also be used to calculate direct, indirect, and overall accuracy scores for each person. These scores can then be used as either independent or dependent variables in analyses to assess factors that predict individual differences in mean-level bias and correlational, direct, indirect, and overall accuracy as well as the downstream consequences of these different types of accuracy.

Using this same approach, the T&B model can also be used to create individual scores assessing the role of bias variables on the judgment. For example, a researcher could calculate individual scores indicating how strongly each person in their sample relied on their own feelings to estimate their roommate's feelings. They could then correlate these scores with other variables to assess the factors that predict whether people engage in biased processing, as well as how engaging in biased processing predicts relationship outcomes.

Conclusion

In this article, we have discussed several different approaches that accuracy researchers can use to assess the veridicality of people's judgments. We have discussed two different forms of accuracy—mean-level bias and correlational accuracy—and the types of questions that they can be used to address. We have also highlighted how both of these types of accuracy can be assessed on the level of the sample and the level of the individual.

In the first half of this article, we outlined multilevel modeling approaches that can be used to separately examine mean-level bias and correlational accuracy in close relationships. While difference score and correlational approaches have received criticism in the past (e.g., Cronbach, 1955), we believe that it is still possible for researchers to utilize these approaches as long as they acknowledge the potential limitations of doing so. However, we would also recommend that researchers employ the T&B model whenever possible. We have discussed how both mean-level bias and correlational accuracy can be simultaneously examined using the T&B model and have outlined how this model can potentially be used to overcome some of the limitations of using difference score and correlational approaches to examine accuracy on ratings of a single target. The T&B model also has the unique advantage of allowing researchers to statistically assess the possibility that perceivers are both accurate and biased in their judgments and that accuracy can be achieved both through having direct access to the truth and indirectly through relying on biased processes. We hope that relationship researchers capitalize on these models to examine the roles of truth and bias in their own research, as doing so will provide the opportunity to shed new light on the processes and outcomes of accuracy in close relationships.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Barranti, M., Carlson, E. N., & Furr, R. M. (2016). Disagreement About Moral Character Is Linked to Interpersonal Costs. *Social Psychological and Personality Science*, *7*, 806–817.

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*, 142–163.

Biesanz, J. C. (2010). The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy. *Multivariate Behavioral Research*, *45*, 853–885.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217.

Castro-Schilo, L., & Grimm, K. (In Press). Using residualized change versus difference scores for longitudinal research. *Journal of Social and Personal Relationships*.

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

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol.998). Hoboken, NJ: John Wiley & Sons.

Fletcher, G. J., & Kerr, P. S. (2010). Through the eyes of love: Reality and illusion in intimate relationships. *Psychological Bulletin*, *136*, 627–658.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, *55*, 149–158.

Gage, N. L., & Cronbach, L. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Review*, *62*, 411–422.

Gagné, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative review. *Personality and Social Psychology Review*, *8*, 322–338.

Hall, J. A., Rosip, J. C., LeBeau, L. S., Horgan, T. G., & Carter, J. D. (2006). Attributing the sources of accuracy in unequal-power dyadic communication: Who is better and why? *Journal of Experimental Social Psychology*, *42*, 18–27.

Hall, J. A., Schmid Mast, M., & West, T. V. (2016). *The social psychology of perceiving others accurately*. Cambridge, MA: Cambridge University Press.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Iida, M., Seidman, G., & Shrout, P. E. (In Press). Models of interdependent individuals and dyadic process in relationship research. *Journal of Social and Personal Relationships*.

Ickes, W. J. (Ed.). (1997). *Empathic accuracy*. New York, NY: Guilford Press.

Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, *59*, 730–742.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental SocialPsychology*, *3*, 1–24.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in socialpsychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54–69.

Jussim, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, *98*, 54–73.

Jussim, L. (2012). *Social perception and social reality: Why accuracy dominates bias and self-fulfilling prophecy*. New York, NY: Oxford University Press.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York, NY: Guilford Press.

Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology*, *80*, 439–448.

Kenny, D. A., & Kashy, D. A. (2010). Dyadic data analysis using multilevel modeling. In J. J. Hox & J. K. Roberts (Eds.), *The handbook of advanced multilevel analysis* (pp. 355–371). New York: Routledge.

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

Krueger, J., & Zeiger, J. S. (1993). Social categorization and the truly false consensus effect. *Journal of Personality and Social Psychology*, *65*, 670–680.

Laird, R. D., & De Los Reyes, A. (2013). Testing informant discrepancies as predictors of early adolescent psychopathology: Why difference scores cannot tell you what you want to know and how polynomial regression may. *Journal of Abnormal Child Psychology*, *41*, 1–14.

Laird, R. D., & Weems, C. F. (2011). The equivalence of regression models using difference scores and models using separate scores for each informant: Implications for the study of informant discrepancies. *Psychological Assessment*, *23*, 388–397.

Lane, S.P., & Hennes, E.P. (In Press). Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships*.

Lemay Jr, E. P. (2014). Accuracy and bias in self-perceptions of responsive behavior: Implications for security in romantic relationships. *Journal of Personality and Social Psychology*, *107*, 638–656.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996a). The self-fulfilling nature of positive illusions in romantic relationships: Love is not blind, but prescient. *Journal of Personality and Social Psychology*, *71*, 1155–1180.

Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996b). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology*, *70*, 79–98.

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. *Psychological Methods*, *11*, 127–141.

Overall, N. C., Fletcher, G. J., & Kenny, D. A. (2012). When bias and insecurity promote accuracy: Mean-level bias and tracking accuracy in couples' conflict discussions. *Personality and Social Psychology Bulletin*, *38*, 642–655.

Rogers, K., Wood, D., & Furr, M. (In Press). Assessment of similarity and self-other agreement in dyadic relationships: A guide to best practices. *Journal of Social and Personal Relationships*.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, *13*, 279–301.

Stern, C., West, T. V., & Schoenthaler, A. (2013). The dynamic relationship between accuracy and bias in social perception research. *Social and Personality Psychology Compass*, *7*, 303–314.

Taft, R. (1955). The ability to judge people. *Psychological Bulletin*, *52*, 1–23.

Tskhay, K. O., & Rule, N. O. (2013). Accuracy in categorizing perceptually ambiguous groups a review and meta-analysis. *Personality and Social Psychology Review*, *17*, 72–86.

West, T. V. (2013). Repeated-measure methods with dyads. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 731–749). New York, NY: OxfordUniversity Press.

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

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: The affordances of a Gibsonian approach. *Personality and Social Psychology Review*, *1*, 204–223.