

The Truth and Bias Model of Judgment

Tessa V. West
New York University

David A. Kenny
University of Connecticut

We present a new model for the general study of how the truth and biases affect human judgment. In the truth and bias model, judgments about the world are pulled by 2 primary forces, the truth force and the bias force, and these 2 forces are interrelated. The truth and bias model differentiates force and value, where the force is the strength of the attraction and the value is the location toward which the judgment is attracted. The model also makes a formal theoretical distinction between bias and moderator variables. Two major classes of biases are discussed: biases that are measured with variables (e.g., assumed similarity) and directional bias, which refers to the extent to which judgments are pulled toward 1 end of the judgment continuum. Moderator variables are conceptualized as variables that affect the accuracy and bias forces but that do not affect judgments directly. We illustrate the model with 4 examples. We discuss the theoretical, empirical, methodological, measurement, and design implications of the model.

Keywords: accuracy, bias, perception

Sara and John are a young couple who have just moved in together. Prior to cohabitating, they got along well and rarely argued, but lately they have found themselves arguing quite frequently about basic household issues. John has accused Sara of not picking up after herself, and Sara has accused John of being overly concerned about the orderliness of their home. Sara is strongly committed to her relationship with John, and she is worried that these small quarrels will damage their relationship. John and Sara are taking part in a study conducted by Kevin, a psychologist, who is interested in examining the processes of accuracy and bias for perceptions made in close relationships. Kevin has couples rate, after each quarrel and over the course of several days, how frustrated they are with their partners and how frustrated they believe their partners are with them. With these data, Kevin can examine several questions that address the processes of accuracy and bias in interpersonal perception, questions that have received considerable attention from accuracy researchers.

However, Kevin is finding that the study of accuracy and bias is not so straightforward. To begin with, he is unsure of how to best conceptualize bias in perception, given the many different meanings that the term *bias* has taken in the literature. Should bias be conceptualized as the degree to which perceivers under- or overestimate their partner's frustration (i.e., whether there is mean-level bias; Fletcher & Kerr, 2010)? Or perhaps bias is the degree to which perceivers assume that their partner feels similar to themselves (i.e., whether they assume similarity; Kenny & Acitelli, 2001). Moreover, Kevin wonders whether relationship closeness should be considered a variable that taps bias or, based on a

motivated reasoning framework, a variable that moderates bias (Murray, 1999). Similarly, Kevin is unsure of how to best conceptualize accuracy. Should accuracy be conceptualized as an exact match between the judgment and the truth? Or should accuracy be defined as the ability of perceivers to track their partner's ups and downs over the course of the study (Fletcher & Kerr, 2010)?

Last, Kevin is interested in the relationships between bias and accuracy. Do accuracy and bias trade off, in that as accuracy increases bias decreases? Or do accuracy and bias coexist (Gagné & Lydon, 2004)? In addition, does the relationship between bias and accuracy differ as a function of relationship closeness? If perceivers who feel close to their partner systematically overestimate their partner's negative feelings, do they also assume similarity and are they necessarily also less accurate? Kevin is also interested in the relationships between bias and accuracy as within-person and between-partners processes. If one relationship partner is accurate, is the other partner also accurate? Also, what about the relationships between different types of bias? Is it possible to examine whether the tendency to overestimate one's partner's feelings is related to the tendency to assume similarity?

Although we have used an example from the domain of close relationships to illustrate these questions as a special case, the basic questions of how accuracy and bias operate and the nature of their interdependence have sustained the interest of psychologists for over half a century. For example, a developmental psychologist might examine the accuracy of numerical estimation in young children (e.g., Berteletti, Lucangeli, Piazza, & Dehaene, 2010); an ecological psychologist might study accuracy of perceptions of an object's width and height (Turvey, Burton, Amazeen, Butwill, & Carello, 1998); and a clinical psychologist might ask whether people with autism have the ability to accurately infer the mental states of others (Roeyers & Demurie, 2010).

Yet, despite a widely held interest in the processes of accuracy and bias, to date, there is not a single framework that can be used to define and measure them. Rather, given such broad interest in these processes, theoretical models have been developed to address accuracy and bias within particular domains of psychology. For

Tessa V. West, Department of Psychology, New York University;
David A. Kenny, Department of Psychology, University of Connecticut.

We thank Cameron Anderson, Julia Blau, Sam Gosling, Edward Lemay, Joe Magee, Paula Pietromonaco, and Peter Wilhelm for their helpful comments and suggestions.

Correspondence concerning this article should be addressed to Tessa V. West, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: tessa.west@nyu.edu

example, if researchers are interested in studying the accuracy of what an object affords, they might apply a Gibsonian approach in which the truth is considered to have a direct effect on judgment (Gibson, 1979); if they are interested in examining which cues perceivers utilize in making judgments of a target's personality, they might apply a Brunswikian approach in which accuracy (i.e., achievement) is theorized to be completely mediated by these cues (Brunswik, 1955); and if they are interested in testing whether a perceiver is lying or telling the truth, they might utilize signal detection theory (Green & Swets, 1974). Moreover, given that the terms *accuracy* and *bias* have taken on many different theoretical meanings, the methodological approaches used to examine them have also varied substantially within the field. For example, accuracy has been defined as the overall mean difference between a judgment and a truth criterion (i.e., the divergence between the mean of the truth and the mean of the judgment), and it has also been defined as the correlation between a judgment and a truth criterion. However, as several recent reviews of the accuracy literature have demonstrated (Fletcher & Kerr, 2010; Gagné & Lydon, 2004), these two processes might well be orthogonal (i.e., perceivers can be accurate in terms of both of these definitions, one but not the other, or neither of them).

We do not believe that different theoretical and methodological approaches to defining and measuring accuracy and bias are wrong or misleading; quite the contrary, they each appreciate distinct and important nuances of bias and accuracy. However, the field of psychology could benefit from a single framework that permits the systematic examination of accuracy and multiple types of bias and is adaptable to the domain of study.

Our goal in this paper to provide a single, integrative framework for the study of accuracy and bias that can be applied widely across domains within psychology. We present a new model, termed the *truth and bias (T&B) model*, which can be utilized to address multiple types of accuracy and bias within human perception. The model provides exact theoretical definitions of parameters that are of interest in the study of accuracy and bias, and so it can be used to streamline science's basic understanding of how accuracy and bias operate, regardless of the researcher's a priori theoretical orientation. For example, one researcher may believe that Gibson's approach to accuracy is correct, and so she applies Gibsonian theory to her study of accuracy of affordances; another researcher may believe that Brunswik's ideas about how perceivers become accurate are correct, and so he adopts a Brunswikian approach to study accuracy of perceptions of traits. The T&B model provides a general framework that both of these researchers can use to test their Gibsonian and Brunswikian ideas, respectively, and to determine if the data are either consistent or inconsistent with their theories. To this end, the model provides a vehicle with which to test competing theories about the processes of accuracy and bias.

The T&B model, unlike any prior approach, places a strong emphasis on the relationship between accuracy and bias. It does not assume a priori that bias and accuracy are inversely related processes; they may in fact be positively, negatively, or not at all related, depending on which psychological factors influence judgment. We propose, in line with the reasoning of Kunda (1990) and others (Kruglanski, 1989), that certain psychological mechanisms lead perceivers to be both accurate and biased, other mechanisms lead to more accuracy and less bias, and other mechanisms have their effects on accuracy or bias but not both. For a given study, a

T&B analysis may reveal that accuracy and bias have a positive relationship, a negative relationship, or no relationship at all, depending on the underlying psychological mechanism.

The T&B model can be relatively easily operationalized. All of the theoretical parameters and variables of the model can be directly translated to empirical ones. Thus, the T&B model not only will help streamline psychologists' theoretical understanding of how accuracy and bias operate as general processes but also will assist in how psychologists proceed in testing them.

We start with the basic theoretical principle that judgments are being pulled by two central forces. These two forces are the *truth force* and the *bias force*, and the key questions are their strength and relationship. Moreover, the strength of these forces and the relationship between them may be moderated by factors that vary by the perceiver and by the context. We begin with a basic conceptualization of our model, followed by an integrative review of previous theories of accuracy and bias in human judgment; we borrow from each of these theories in the conceptualization of the T&B model. Next, we translate the theory to practice by demonstrating how the T&B model can be tested empirically by providing several elaborated case examples. Finally, we return to the model and discuss advanced methodological and theoretical considerations.

Basic Principles of the T&B Model

The Truth Force

The T&B model begins with the basic assumption that in human judgment, perceptions are determined by the truth or reality. Judgments are driven very strongly by reality in some cases and weakly if at all in others. For example, people in general are quite accurate in ordering the length and width of manually wielded objects (Arzamarski, Isenhower, Kay, Turvey, & Michaels, 2010). However, in judgments of lie detection, perceivers perform only slightly better than chance (Bond & DePaulo, 2008), indicating that the effect of truth on judgment is weak. We refer to the strength of the effect of the truth on judgment as the *truth force*. The concept of the truth force is similar to Kunda's (1990) discussion of *reality constraints* in the context of motivated perception processes.¹ According to Kunda, judgments are thought to be guided by perceivers' underlying motivations to see the target of perception in a certain way (e.g., to see themselves positively on a dimension) but are still guided by reality. As such, judgments can be guided by perceivers' motivations to see a target of perception a certain way only within the constraints of reality. We note that there are many examples in the social perception literature in which the truth is not explicitly considered as having a direct effect on judgment. For example, in Jones and Harris's (1967) classic study of attributional bias, participants who were told that a writer was forced to write a pro-Castro essay inferred that the writer was actually pro-Castro. This finding has been used as support for the correspondence bias (or, as it is also called, the fundamental attribution error). However, because the essays were constructed by the experimenter, the truth did not even exist in this context.

¹ We thank Reviewer 2 for suggesting we incorporate Ziva Kunda's work into our manuscript.

Also, Ross, Greene, and House (1977) in their study of false consensus bias did not measure the actual percentage of Stanford students who would be willing to wear a sandwich board advertising a local business establishment.

Our point is not to claim that all judgments are totally guided by the truth or that the perceiver has direct access to the truth but that judgments may be guided in part by the truth and that accuracy researchers can and should measure it. The model that we present herein is appropriate for studies of human judgment in which the truth is known and is measured. Although some have argued that the truth can be difficult to define, know, and measure (Gilbert, 1997), we argue that if researchers are interested in accuracy of perception, they need to operationally define and measure it.

Defining the Truth

Across different subfields in psychology, scholars have wrestled with the problem of determining the best method of defining proper truth criteria for different types of judgment. Reviewing all of these theoretical approaches is beyond the scope of this paper (for more detailed reviews, see Funder, 1987; Kruglanski, 1989), but we do review major theoretical issues that have received attention in the domains of personality and social psychology. We seek to emphasize that in order to study accuracy and bias, the researcher should develop a precise understanding of the truth and determine a way to measure the truth.

For certain types of accuracy questions, the definition of the truth is fairly straightforward; such is the case for the number of publications an academic has (Malouff, Schutte, & Priest, 2010), the temperature of a room (Ijzerman & Semin, 2010), and the length of a stick (Turvey et al., 1998). Moreover, given the proper tools (e.g., an up-to-date curriculum vitae, a functioning thermometer, and a laser ruler), the measurement of these truths is also straightforward. For other types of judgment, particularly in the domain of social perception, defining and measuring the truth is more of a challenge but not an impossible task (Hastie & Rasinski, 1988; Kruglanski, 1989). Very often, defining the truth requires rigorous theoretical development and empirical testing, and the process must be flexible depending on the context in which judgments are made (Kruglanski, 1989).

Scholars have emphasized, germane to the issue of defining the truth, that it is important to focus on what perceivers are motivated to be accurate about when studying accuracy. According to Swann's (1984) pragmatic accuracy approach, people should be particularly motivated to be accurate in their perceptions of how particular targets behave with them (i.e., circumscribed accuracy) rather than how those targets behave with people in general (i.e., global accuracy; Gill & Swann, 2004); the former helps perceivers navigate their social worlds better than does the latter. For example, it may be more pragmatic for a perceiver to know whether a bully at school would attack him or her rather than whether that bully would attack others in general (Kenny et al., 2007). The two types of accuracy within a pragmatic accuracy framework—circumscribed and global—lend themselves to two different accuracy questions that have two different truth criteria and, potentially, two different answers. With the T&B model, we can examine the truth forces of each. That is, we examine the strength of the effect of a target's behaviors across all targets on the perceiver's judgment of the target (the truth force for global

accuracy) and the strength of the effect of a target's behaviors with the perceiver in particular on the perceiver's judgment of the target (the truth force for circumscribed accuracy). In sum, Swann's pragmatic accuracy approach provides an elegant illustration of how the same type of judgment (e.g., judgments of a target's physical aggression) can inform two different accuracy questions with two different truth criteria.

In a similar vein, there is a classic debate in psychology about the importance of cross-situational consistency in defining personality (Bem & Allen, 1974). That is, if scholars are interested in determining accuracy of personality judgments, they must first establish whether the truth criterion should reflect behaviors at the level of a particular situation or at a more general level that reflects how targets behave across situations. For example, a researcher may ask, are perceivers accurate in their judgments of how neurotic their romantic partners are at home? Or alternatively, are perceivers accurate in their judgments of how neurotic their partners are across multiple situations, such as at work, with friends, and with other family members? These two accuracy questions have different truth criteria, and, subsequently, we can treat them as two different accuracy questions. For example, we could obtain a measure of the truth for how a target behaves across situations and one for how a target behaves in a specific situation and measure the truth forces of each. To this end, the importance of considering cross-situational consistency in determining the accuracy question is similar to Swann's (1984) point about considering the context in which judgments are made.

In some cases, what may appear to be the "same" accuracy question addressed by different researchers is actually several different accuracy questions, each guided by a different theoretical question requiring a different truth criterion. For example, there has been extensive research examining the accuracy of judgments of emotion. Researchers from different domains of psychology may be interested in asking how accurate perceivers are in detecting anxiety. One researcher may be interested in how accurate people are in detecting how anxious their partner feels during an interaction, and thus the interest is in subjective perceptions of the partner's anxiety. To test this accuracy question, the researcher may employ the empathic accuracy paradigm (Ickes, 1997), in which two people interact and then make ratings of how they felt during specific times in the interaction and make judgments of how they believed their partner felt during corresponding times in the interaction. The truth criterion is the target's statement of his or her feelings, because what is of interest is whether perceivers are accurate in knowing the target's subjective experience of anxiety. In another judgment context, the subjective emotional experience of the target is irrelevant. Such is the case in studies in which the interest is in accuracy of judgments of emotions with standardized interpersonal sensitivity measures in which individuals are asked to detect the emotions of an expresser based on face, body, and vocal information (e.g., the Profile of Nonverbal Sensitivity; see Hall, Andrzejewski, Murphy, Mast, & Feinstein, 2008). An example is a study in which perceivers watch an interaction between two targets, both of whom have been trained to appear either angry or not angry. For the use of standardized instruments of posed target and the videotape of targets posing anger, the subjective emotional experiences of targets are irrelevant because the targets are posing; the truth criterion is how anxious the targets appear in their behaviors.

In sum, although defining and measuring the truth is difficult for some types of judgments, it nonetheless is still possible. How the truth is defined and measured may differ depending on the theoretical question of interest, the methodological constraints of the study, and the context in which judgments are made. The key point that we make here is that to study accuracy, we must define and measure the truth; only then can we examine how strongly it affects judgment.

The Truth Force and Truth Value

The T&B model makes a distinction between the *truth force* and the *truth value* (see Table 1 for terminology and definitions of the model). Imagine that a researcher is interested in how accurate perceivers are in their judgments of the temperature, and the truth is defined as the actual temperature. For the judgment of temperature, the truth value is where on the temperature scale the truth lies (i.e., what the actual temperature is). The truth force is the degree to which judgments of the temperature are attracted toward the truth value. As another example, a perceiver may judge how pro-Castro a writer is, and how pro-Castro the writer actually is (i.e., the truth) is measured with a 1–7 self-report Likert-type scale. If the writer is actually a 4 on that scale, then the truth value would be 4, and the truth force would be how strongly the judgment is attracted toward that value. As a final example, in a study in which wives make judgments of their husband's mood across several days, the husband's actual mood on a given day is the truth value, and the degree to which wives' judgments are on average pulled by their husband's daily moods is the truth force. The truth force provides a measure of how strong the effect of the truth value is on the judgment. We presume in the T&B model that the truth is a measured or manipulated variable and that it is in the same units as the judgment.

The Bias Force and Bias Value

Judgments not only are pulled toward the truth but are also systematically pulled toward other factors that might be near the

truth but are theoretically distinct from the truth. In the T&B model, the term *bias* is used to refer to any systematic factor that judgments are being attracted toward, besides the truth. Many biases can be measured with variables, and, as with the truth, we can examine *bias forces* and *bias values*. To return to the Castro essay example, imagine that perceivers also make judgments of how pro-Castro they themselves are, and the interest is in the degree to which they assimilate such judgments. The *bias variable* would be the perceivers' self-ratings of how pro-Castro they are. The bias value, like the truth value, is the location on the scale where perceivers see themselves. So a perceiver may rate himself as a 7 on the 1–7 pro-Castro scale. Conceptually parallel to the truth force, bias forces can be thought of as the extent to which judgments are being pulled toward bias values. Perceivers' judgments of how pro-Castro writers are may be strongly determined by how pro-Castro they themselves are, and so the bias force would be quite strong.

Different fields within psychology have emphasized different bias variables for which their forces on judgments are examined. For example, in Murray and colleagues' work on projection of personal values in romantic relationships (e.g., Murray, Holmes, Bellavia, Griffin, & Dolderman, 2002; Murray, Holmes, & Griffin, 1996), perceivers' judgments of their partner's personal values not only are pulled toward the truth (i.e., the perceivers' partner's actual values) but are also pulled toward perceivers' own personal values, which is considered a bias variable in the T&B model. The strength of the effect of perceivers' own personal values on their judgments of their partner's personal values is the bias force. A second example taken from cognitive psychology is the bias of perseveration, in which people stick to their past judgments and do not update their new judgments. If a perceiver's prior judgment was 12 on a scale, that judgment can be used to predict the current judgment. If the truth was 14 on the same scale, the bias of perseveration would indicate that judgments are being strongly pulled toward the past judgment. Thus, the past judgment serves as the bias variable. A third example comes from the area of intergroup relationships, and it is the bias to see ingroup members more positively than they actually are and to see outgroup members less positively than they actually are.

Given how bias is conceptualized in the T&B model, bias variables should in principle be measured in the same units as are the judgment and use of a scale. For example, perceivers may judge themselves and their partner using a Likert-type scale on which the value of 1 indicates *not at all* and the value of 7 indicates *very much*. We later discuss in detail what is lost by not having the bias variable and the judgment measured with the same units.

The T&B model makes a sharp distinction between bias variables and truth variables. As we discuss above, it is very important that a researcher has a clear definition of the truth and how it is measured. One reason why it is important to do so is because in some cases, bias variables are correlated with the truth variable and yet they are still theoretically distinct from each other. For example, in a study in which researchers are interested in the bias of assumed similarity, perceivers' self-perceptions (i.e., the bias variable) may be highly correlated with the truth variable (i.e., where the person being perceived actually stands on a judgment dimension); however, these variables are theoretically distinct. To this end, the definition of bias in the T&B model is quite different from the notion, adopted by many researchers, that bias is defined as

Table 1
T&B Model Terms and Definitions

Term	Definition
Truth value	The value on the truth criterion toward which a judgment is attracted.
Truth force	Extent to which judgments are attracted toward the truth value.
Bias	Any value that judgments are attracted toward besides the truth.
Bias variable	Attractor variables that lead to a particular value on the judgment scale.
Bias value	The location on the bias variable toward which the judgment is attracted.
Bias force	Extent to which judgments are attracted toward the bias value.
Directional bias	Extent to which judgments are attracted toward a particular end of the judgment scale.
Moderator variable	A variable that influence the strength of the truth and the bias forces.

Note. T&B model = truth and bias model.

inaccuracy or error. Rather, our use of the term *bias* includes both valid and invalid information used in making judgments, that is, information that might lead perceivers to be accurate and information that would lead them to be inaccurate. For example, perceivers may use their own self-perceptions (e.g., their own attitudes) in forming their impressions of another person's attitude, and if the bias variable is correlated positively with the truth, perceivers will become accurate by being biased. Given this possibility, we do not define what a bias variable is by its relationship to the truth, in part because that relationship may vary across contexts. In the T&B model, any information (both valid and invalid) that determines judgment other than the truth is referred to as a bias, regardless of what its relationship to the truth is. Following Harvey, Town, and Yarkin (1981), bias in the T&B model does not necessarily imply that perceivers are wrong. Rather, according to the model, the extent to which biases lead to incorrect judgments is an empirical question that can be examined and should not be a theoretical distinction. The T&B model also makes a distinction between bias and error. Bias is theoretically distinct from error in judgment in that bias varies in a systematic, theoretically relevant, and predictable way, whereas error refers to factors that drive judgments that occur at random. Error includes theoretically uninteresting components, such as measurement error.

Although in the T&B model, it is presumed for a given research question that there is only one truth for that specific accuracy question, there are typically many types of biases that can affect judgment.² We have discussed the case in which biases are measured with variables, and the strength of the effect of the bias values on the judgment of the target is the *bias force*. Moreover, in judgment there is almost always a bias that captures how attracted perceivers are toward a particular end of the judgment scale (e.g., an acquiescent response bias). Following Kunda (1990), we refer to this bias as *directional bias*. There are many examples of these directional biases in the perception literature. For instance, when perceivers make a judgment about whether someone is telling the truth using a scale in which the truth is one end and lying is on the other, people generally are biased to think that others are usually telling the truth (DePaulo & Kashy, 1998). They are also biased to think that their romantic partner is attractive (Fletcher, Simpson, & Boyes, 2006; Gagné & Lydon, 2004) and that relationships improve over time (Karney & Coombs, 2000). For directional bias, the bias value is one pole of the dimension on which judgments are made. For instance, in ratings of partner's attractiveness on a 1–7 scale, the bias value is the 7 on the scale. Directional biases are, therefore, distinct from biases measured with variables, in that their force is not measured with a variable; rather, they are conceptualized as how strongly judgments are being pulled toward a particular end of the scale. The term directional bias (Kunda, 1990) has also been referred to as the *positivity bias* (Rusbult, Van Lange, Wildschut, Yovetich, & Verette, 2000) and *mean-level bias* (Fletcher & Kerr, 2010; Murray, 1999). As we discuss later in this paper, the term *bias* in signal detection theory is analogous to directional bias in the T&B model.

In the T&B model, directional bias simultaneously captures two pieces of information: First, it captures which end of the scale perceivers are biased toward, and second, it captures how strongly their judgments are being pulled away from the mean level of the truth. Imagine that perceivers were asked to judge the percentage of time targets were lying from 0% to 100%, and the average rating

was 10% (i.e., perceivers were biased toward the lower end of the scale). If the actual percentage of time lying was 40%, the force of the directional bias would be quite strong because judgments are being pulled away from the mean level of the truth. However, if the actual percentage of time lying was 12%, the force would be weak.

Biases measured with variables (e.g., assumed similarity) and directional bias represent two very different processes in perception, which may or may not be correlated. For example, perceivers who are more likely to assume that their partner is as physically attractive as himself (i.e., the bias force for the bias of assumed similarity is positive) may also overestimate the attractiveness of their partner. In this case, the bias force and the directional bias are positively correlated.

In Figure 1, we present diagrams that illustrate the T&B model. We have the continuum along which the judgment is made, which ranges from negative to positive, and for that judgment, we have the truth (*T*) and bias (*B*) values, always the same. The size of the circles indicate the size of the force (i.e., the larger the circle, the greater the force). For Figure 1a, the bias and truth forces are equal; for Figure 1b, the truth force is larger than the bias force; and for Figure 1c, the bias force is larger than the truth force. The judgment is the balance point, which is indicated by the triangle. For Figure 1a, that balance point is halfway between *T* and *B* because the forces are equal; for Figure 1b, it is nearer *T* because the truth force is greater; and for Figure 1c, it is nearer *B* because the bias force is greater.

Moderators of Truth and Bias Forces

So far, we have discussed truth and bias forces without considering the processes by which accuracy and bias in judgment occur. In the T&B model, moderators are used to examine these processes, and they can change the truth force, the bias force (or forces), the directional bias, or all three. Previous literature has not made a clear distinction between a bias variable and a moderator. In the T&B model, a moderator variable does not influence the judgment directly (i.e., how perceivers "see" the person or object being judged); rather, it influences the strength of the forces that do determine judgment. Thus, moderators are a key component of the T&B model, because they enable researchers to understand what controls the truth and bias forces. Take, for example, the finding that, in comparison with people in less satisfied relationships, people in satisfied relationships see their partner as more physically attractive than the partner actually is. Is relationship satisfaction a moderator or a bias variable? Within the T&B model, satisfaction is viewed as a moderator because it alters the directional bias (i.e., it affects the extent to which perceivers overestimate their partner's attractiveness). Moreover, satisfaction may interact with the truth force. Perceivers in less satisfied relationships may be more accurate about the actual attractiveness of their partner than are those in more satisfied relationships.

The study of moderators of accuracy and bias has garnered much attention in psychology. Research has examined judgments

² One could take a more postmodernist point of view and allow for two different versions of the truth, both of which might be competing to explain judgment.

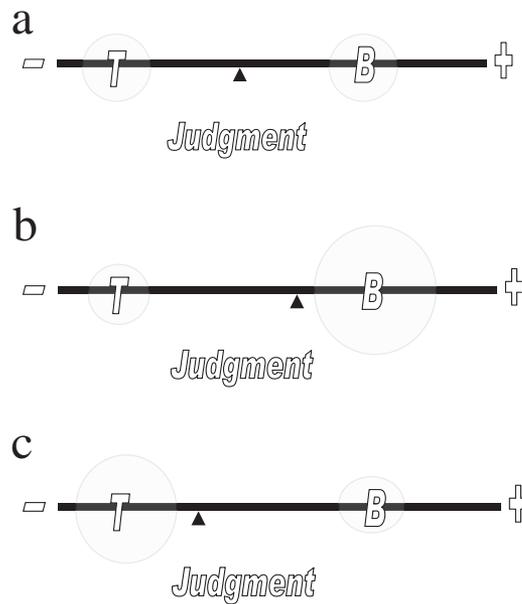


Figure 1. The schematic for the T&B model, in which the participant's judgment, the balance point, is determined by the truth (T) and the bias (B). The size of the oval represents force. For (a), the forces are equal; for (b), the bias force is larger than the truth force; for (c), the truth force is larger than the bias force.

made in close relationships (for a review, see Gagné & Lydon, 2004), impressions drawn during the stages of initial impression formation (Ambady & Rosenthal, 1992; Hall & Andrzejewski, 2008), and perceptions of in- and out-group members (Robbins & Krueger, 2005). Diverse types of variables have been examined as moderators of accuracy and bias, and previous theoretical models have placed these variables into distinct categories. For example, in Funder's (1995) realistic accuracy model (1995), accuracy in perception of personality is conceptualized to be a function of four general types of moderators: good target, good judge, good trait, and good information; these four moderators may interact with the truth force to inform the process of accuracy. These moderators are good in the sense that they are theorized to strengthen the truth force. For example, when a perceiver judges a "good target"—a target who is easier to judge than are other targets—the truth force would be stronger than it is for judgments of targets who are not good. We note that each of these moderators may well influence the bias force, as well as the truth force.

Outside the domain of person perception, research on accuracy of visual perception has examined physical characteristics of perceivers as moderators of accuracy. For example, the steepness of a hill seems more extreme after vigorous jogging (Bhalla & Proffitt, 1999), and the distance from one point to another seems longer after strapping on a heavy backpack (Proffitt, Stefanucci, Banton, & Epstein, 2003). Extending research on motivated visual perception to the social context, Balci et al. (2006) found evidence that motivational states (i.e., the motivation to think of oneself in a favorable way) influence biases in the visual perception of ambiguous figures (Balci et al., 2006).

Often, moderation is examined for accuracy or bias, but usually the processes are not simultaneously examined. The T&B model

permits a simultaneous examination of the moderation of accuracy and of bias. Moreover, these processes may not necessarily be moderated in a parallel fashion; the moderator might increase both bias and accuracy or one but not the other. In the T&B model, moderators are examined for the truth and bias forces, as well as for directional bias. This provides insight concerning how the processes of accuracy and bias are related, a possibility that we discuss in the next section.

The Relationship Between Truth and Bias

The T&B model not only emphasizes the dual presence of truth and bias in judgment but also provides several different theoretical possibilities as to how truth and bias can be related. There are four ways in which truth and bias can be related: "mediation" of the truth via bias, the relative size of the forces, the extent to which they are correlated within persons, and the way in which they are differentially affected by moderators.

One way in which truth and bias are theoretically related is that accuracy may be achieved by the perceiver's being biased. For example, as empirically demonstrated by Hoch (1987), perceivers assume similarity to their romantic partner. If they and their partner are actually similar, they are accurate by being biased (discussed in Cronbach, 1955, and more recently by Neyer, Banse, & Asendorpf, 1999, and Kenny & Acitelli, 2001). The extent to which accuracy is achieved indirectly through bias is an empirical question that the T&B model can be used to address.

A second way to examine the relation between accuracy and bias is to examine the relative strength of the truth and bias forces. We might think that they ordinarily trade off: One is high or the other is low, but they cannot both be high. In fact, several recent theoretical and empirical reviews have demonstrated that this is not necessarily the case (Fletcher & Kerr, 2010; Fletcher et al., 2006; Gagné & Lydon, 2004; Luo & Snider, 2009). Accuracy and bias might both be high, and they might persist across many perceptual contexts. For example, in close relationships, as previously discussed, perceivers are motivated to be both accurate and biased. In studies of emotion detection, evidence indicates that perceivers are accurate in labeling emotions expressed by neutral faces (Sasson et al., 2010), but some individuals also show a threat bias in that they misinterpret disgust as contempt (Heuer, Lange, Isaac, Rinck, & Becker, 2010). In other research domains, there is evidence to suggest that perceivers are both accurate and biased in their nonsocial judgments. For example, a Gibsonian researcher might find that perceivers are accurate in knowing whether they can fit through a doorway, but they may also systematically overestimate the width of that doorway, thereby engaging in directional bias.

A third way to examine the relationship between accuracy and bias is to examine the extent to which these processes are moderated by *person*; that is, some perceivers might be more accurate than others, and some perceivers might be more biased than others. As such, the processes of accuracy and bias estimated within a given perceiver might be correlated: If a perceiver is more accurate than other perceivers, is he or she also biased? In addition, we can examine the relationships between different types of bias (e.g., whether perceivers who engage in directional bias also engage in the bias of assumed similarity). Examination of the correlation

between accuracy and bias requires a within-person approach, which we discuss in the illustrative example section.

A fourth way to examine the interrelation between the truth and bias forces is to examine how these forces are affected by moderators. The extent to which bias leads to accuracy for different levels of the moderator is an empirical question that can only be examined if moderation of the truth and bias forces is considered within the same analysis.

Moderation in the T&B Model

Moderation within a T&B framework informs the processes of accuracy and bias, and when we examine moderation of the truth and bias forces simultaneously, it also provides insight into the nature of the relationship between them. We turn to the work of Kunda (1990) and others (Kruglanski, 1989) on motivated reasoning as a starting point from which we form predictions of the accuracy–bias correlation.

According to a motivated reasoning perspective, perceivers can be accurate and biased at the same time because the process of engaging in a reasoning that facilitates accuracy also allows them to be biased. For example, a perceiver may be motivated by the goal to see a potential romantic partner accurately, and so she engages in a cognitive strategy of seeking individuating information that leads her to be accurate. However, by doing so, she is better able to justify her desire to see her partner positively.

Kunda (1990) provided a plethora of examples illustrating that accuracy and bias can stem from the same underlying psychological factors. On the basis of this evidence, we propose that some categories of moderators strengthen both the truth and the bias force, because they are theoretically proposed to influence both forces in the same way. Extending this hypothesis, we also propose that some moderators exert their effects on both the truth and the bias force but in opposite ways, resulting in the truth and the bias force being inversely related. Last, another category of moderators strengthens (or weakens) only the truth or the bias force, but not both, because these biases are theoretically proposed to exert their effects on one process but not the other. We briefly outline each of these three possibilities.

In some domains, such as the domain of close relationships, accuracy and bias are positively related processes because they share an underlying construct that informs both processes. As previously discussed, relationship closeness motivates perceivers to want to understand their partner, to see the partner in the best light as possible, and to see the partner as similar to themselves (Gagné & Lydon, 2004; Murray, 1999); that is, it motivates them to be accurate and biased at the same time. As another example, during interactions with new acquaintances, the desire to have a smooth interaction may motivate perceivers to see their partners accurately, because accuracy is an important component of communication. This motivation leads perceivers to engage in reasoning strategies that have been shown to increase accuracy, such as basing their judgments on individuating information rather than on a target's category membership (Neuberg & Fiske, 1987). Perceivers are also motivated to see their new partners positively because they want to get along with them (Kunda, 1990), and so they may construe their partners' behaviors in the best light possible. The desire to be accurate would lead to accuracy only if perceivers engage in reasoning strategies that actually facilitate accuracy

(e.g., attending to the correct information about a target; Harkness, DeBono, & Borgida, 1985). It is also essential that perceivers are provided with sufficient information to be accurate (Funder, 1995).

In some cases, the same construct is hypothesized to influence both the truth and the bias force but in opposite ways. For such judgments, the truth and bias forces will be inversely related—as bias increases, accuracy decreases, and vice versa. For example, *threat* has been considered as a motivational construct that decreases accuracy and increases bias in the context of close relationships. Simpson, Ickes, and Blackstone (1995) found that when perceivers rated how their partner felt during conversations in which they viewed images of highly attractive others, accuracy was weaker because believing that one's partner is attracted to another is a relationship threat. Although Simpson et al. did not specifically examine directional bias in this study, we would propose that if they did, they would have found that the more threatened the perceivers were, the more the perceivers would have underestimated how attractive their partner found the highly attractive others (see Case 4 below).

As an additional example, imagine a study in which some perceivers are given very impoverished information on which to base a judgment of a target, and other perceivers are given very detailed information. Perceivers who are given impoverished information may assume that the target is similar to the self, because the perceiver uses the self as a baseline in the absence of other relevant information (Kenny, 1994). In this case, we would expect that the truth and bias forces would be negatively correlated because the more perceivers are able to attend to relevant information about the target, the less they would rely on the self when making judgments. Here, the bias of assumed similarity is motivated by a very different psychological process than that in the context of close relationships.

Last, we hypothesize that the truth and bias forces are weakly (if at all) correlated when accuracy and bias are motivated by very different mechanisms. For example, imagine a lie-detection study in which a perceiver must make a judgment of whether a set of targets is lying or telling the truth. Perceivers may engage in a directional bias—that is, they may assume that perceivers on average are telling the truth more often than they are lying—because they are motivated by an ideological belief that the world is an honest place. However, perceivers may also be motivated by a goal of wanting to know the truth because of the high cost of being lied to, which would strengthen the truth force. In this example, there are two very different psychological mechanisms that exert independent effects on the truth and the bias force, respectively.

In sum, we theorize that examining moderators of the truth and bias forces provides insight into the mechanisms that drive accuracy and bias as independent and related processes. We have given only an outline of how an examination of the factors that influence the truth and bias forces informs the psychological processes that underlie them; it might well serve as a basis for a more elaborate specification of the process.

The Formal Model

Thus far, we have introduced the theoretical parameters of the T&B model. Here we present the formal details of the model and discuss how that model can be translated from theory into practice.

To formalize the model, we use the following example: Imagine that we are interested in the degree to which individuals' judgments of their current partner's relationship satisfaction are determined by the truth (i.e., how satisfied the partner actually is) and by bias (i.e., how satisfied perceivers are themselves). Sara judges her partner John's relationship satisfaction and her own relationship satisfaction, and John rates his own relationship satisfaction. All three variables are measured with the same units of measurement, for example, a satisfaction scale that ranges from 1 (*not at all satisfied*) to 7 (*extremely satisfied*). In the T&B model, it is essential to measure not only the truth but also the judgment and the bias with the same units of measurement. We discuss later what is gained by having the truth and judgment measured with the same units.

We first subtract from the truth variable (i.e., John's actual relationship satisfaction) its grand mean (i.e., we grand mean center the variable), and we subtract from the judgment the *truth* grand mean. When the bias is measured on the same scale as the judgment, as in our example, we also subtract from the bias variable the grand mean of the truth. Thus, all three variables—the truth, the judgment, and the bias variable—are all centered with the grand mean of the truth.

We begin by assuming a simple version of the model, one in which the truth and bias forces do not vary. This basic version of the T&B model can be formally expressed for perceiver *i* (e.g., Sara) as follows:

$$J_{Ci} = b_0 + tT_{Ci} + bB_{Ci} + E_i \quad (1)$$

where J_{Ci} is Sara's judgment of John's relationship satisfaction; b_0 is directional bias (i.e., the degree to which Sara is biased to perceive John as more satisfied than he is); T_{Ci} is the truth variable (i.e., John's actual satisfaction); t is the truth force (i.e., the strength of the effect of John's actual satisfaction on Sara's judgment of John); B_{Ci} is the bias variable (i.e., Sara's actual satisfaction); b is the bias force (i.e., the strength of the effect of Sara's satisfaction on her judgment of John); and E_i is random error.

The *C* subscript means that the variables are all centered using the grand mean of the truth. In the T&B model, b_0 , or directional bias, represents the directional value multiplied by the directional bias force. Disentangling directional bias value and force is complicated and is detailed later in the discussion. For now, the key issues are the sign of the directional bias and how far away it is from zero. Because directional bias is an intercept, it refers to the estimated judgment bias when both T and B equal zero, which is expressed in Equation 2 as follows:

$$b_0 = \bar{J} - \bar{T} - b(\bar{B} - \bar{T}). \quad (2)$$

If we had only the truth predicting judgment (i.e., no bias variable in the model), directional bias would equal the mean difference between the judgment and the truth. This is exactly how Fletcher and Kerr (2010) conceptualized directional bias, which they referred to as *mean-level bias*. When the bias variable is included in the model, the directional bias refers to the difference between the truth mean and judgment mean. The statistical test of the intercept evaluates whether the directional bias is statistically different from zero. In the above example, perceivers would underestimate their partner's satisfaction if the constant were negative and would overestimate their partner's satisfaction if it were positive.

The parameters of t and b are interpreted as regression coefficients and are normally positive in that T and B are attractors. Because the truth value is centered on the truth mean and the bias value is centered on the truth mean, t measures the truth force when the bias value is at its mean, and b measures the bias force when the truth value is at the mean. A key question is the relative size of t and b , which can be tested statistically. As shown in Figure 1, sometimes the truth force may be larger than the bias force and other times vice versa. We could allow T and B to interact; that is, the bias may be stronger for different values of the truth.

In some studies, the truth value is known, but it is not a variable that varies for each perceiver. For instance, persons might be asked the temperature of the room, and there is just one correct answer for everyone. Also, in intergroup studies, when everyone is a member of both the in-group and the out-group, the truth value for ratings of in-group members minus out-group members equals zero. For studies where T is known but is not a variable, we can still center J and B using the value of T , and so we can measure directional bias but not t , the truth force.

Indirect Accuracy

We can ask whether bias lead to accuracy. This is a key question within the Brunswikian framework: To what extent do the cues (biases) explain achievement (accuracy)? Expressed in T&B terms, the question is, how much of the effect of the truth on the judgment is explained by the bias force or forces? In statistical terms, this task is akin to what is done in tests of mediation. In this case, the bias variable is not a mediator per se, because it is usually not caused by the truth variable; rather, it is what is called a *confounder* in the quantitative literature (MacKinnon, Krull, & Lockwood, 2000), because it may be correlated with the truth variable. In Figure 2, we have a path model of the T&B model. In this model we begin with the truth, symbolized as T , causing the judgment, J , and the effect of T on J , denoted as path t . We also have in the model a path from bias, B , to the judgment, J , denoted as path b . As seen in Figure 2, there is a "path" from T to the bias variable, B , denoted as " a ." We put quotes around the path and make it a dashed and not a solid line, because it is not a causal path

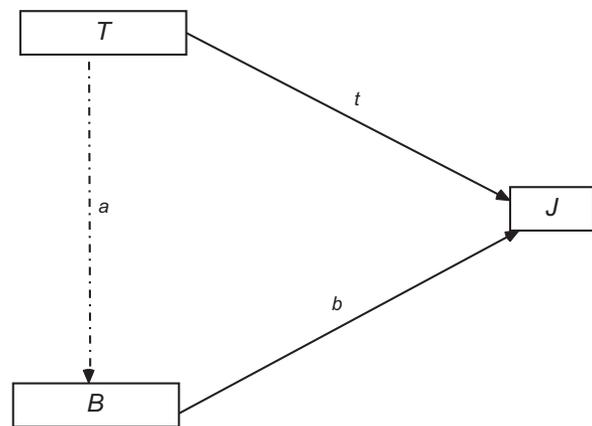


Figure 2. "Mediation" of the effect of Truth (T) on Judgment (J) via Bias (B).

but rather a predictive path. Below we discuss why it is that the two might be correlated. The *indirect effect* of T on J is equal to ab . In accordance with standard tools of mediation, total accuracy is then $t + ab$. The paths t and b can be estimated from Equation 1. To obtain a , we treat B as the outcome and T as the predictor. We illustrate the approach later in the paper when we present examples.

What is the meaning of path a ? Perhaps the simplest illustration of a is for the case in which B is the self-judgment. Consider again the case in which husbands are asked to guess at their wife's relationship satisfaction, and their judgments are predicted by their wife's actual satisfaction and by their own satisfaction (Hoch, 1987). Husbands may be biased in their guesses by their own satisfaction: "I am satisfied and so I assume that my wife is satisfied." If husband's actual satisfaction is similar to the wife's actual satisfaction, path a would be positive, an effect usually referred to as *actual similarity*. It might even be the case that path a is negative, for example, in the case in which a man makes predictions about someone he considers to be very different from himself, such as a romantic rival. When ab and t have opposite signs, introducing that bias into the equation increases the estimate of the truth force (e.g., if the perceiver assumes dissimilarity but the individuals are similar). In this case, controlling for the negative indirect effect leads to an increase in the estimation of the truth force. When ab is the same sign as t , accuracy of perception is due in part to bias (e.g., when the perceiver assumes similarity and the individuals are similar). Path a , in general, represents the validity of the bias, or what Brunswikians have called *cue validity*.

As true for mediation, perceivers can be accurate indirectly, directly, or both. Direct accuracy is the direct effect of the truth on the judgment when the bias is included in the model, which is path t from Equation 1: It is accuracy achieved "directly" after controlling for bias. *Total accuracy* can be estimated as indirect accuracy plus direct accuracy.

We can also determine the extent to which the effect of directional bias is "mediated" by the bias variable. In Equation 2, we stated that the directional bias equaled $\bar{J} - \bar{T} - b(\bar{B} - \bar{T})$, the difference between the mean of the judgment and the mean of the truth, adjusted for any other biases. Thus, "total" directional bias, or $\bar{J} - \bar{T}$, is equal to "indirect" directional bias, or $b(\bar{B} - \bar{T})$, plus "direct" directional bias of b_0 . For example, consider a study in which the mean judgment of partner satisfaction is 6.1 and the mean truth is 5.1. In a model in which no bias variable were included, directional bias would be 1.0, indicating that perceivers see their partner as one point higher on the scale of satisfaction than their partner actually reports being. However, when the bias of assumed similarity is added to the model, the directional bias might shrink to only 0.5. Because people think that their partner is more satisfied than the partner really is, which makes the bias mean greater than the truth mean, and because people are biased to see others as similar, the bias of assumed similarity partially accounts for the directional bias. In general, such a reduction happens whenever the directional bias is smaller in absolute value than the difference between the mean judgment and the mean truth. In this example, this is the case because the directional bias is 0.5 when we control for the bias of assumed similarity and the judgment mean minus the mean truth is 1.0. Because the bias variable

is included in the model, directional bias shrinks, and so it is in part "mediated" by the bias variable.

Adding Moderators to the Model

We can expand Equation 1 to include a moderator, denoted as M :

$$J_{Ci} = (b_0 + mM_i) + (tT_{Ci} + t_M T_{Ci} M_i) + (bB_{Ci} + b_M B_{Ci} M_i) + E_i. \quad (3)$$

Here, consistent with Equation 1, we have directional bias (b_0), the centered truth value for person i (T_{Ci}), the truth force (t), the centered bias value for person i (B_{Ci}), and the bias force (b). Imagine that we were interested in whether gender of the perceiver moderates the truth force and the bias forces in individuals' perceptions of their partner's satisfaction. Here, Sara's gender is denoted as M_i , and the overall effect of Sara's gender on her judgment of John is m , which estimates the moderating effect of gender on directional bias. The interaction between gender and the truth variable is denoted as $T_{Ci} M_i$, and the effect of the moderator on the truth variable is denoted as t_M , which, in this example, reflects whether the truth force is stronger for men or for women. Last, the interaction between the moderator variable and the bias variable is denoted as $B_{Ci} M_i$, and the effect of the moderator on the bias force is denoted as b_M , which estimates whether the bias force is stronger for men or for women. By including moderation of the truth and the bias force simultaneously, we can examine whether the pattern of moderation is the same for accuracy and bias. If b_M and t_M are of the same sign (assuming that b is positive), the processes of accuracy and bias are moderated in the same direction by gender (e.g., that women are both more accurate and more biased than are men in their judgments of their partner's satisfaction). If b_M and t_M are of opposite signs, bias and accuracy trade off for men or for women. Perhaps women are more accurate than men and men are more biased than women. A variable may moderate the effect of *indirect accuracy* by moderating path a (the path from T to B in Figure 2), path b , or both, and it may moderate *direct accuracy* by moderating the truth force, creating a form of moderated "mediation." We later illustrate these different types of moderation.

We can also examine *person* as a moderator of accuracy and bias. When perceivers make multiple judgments (e.g., make judgments at different times), we can examine within-person differences in the processes of accuracy and bias. Here, we assume that the multiple judgments are made of multiple targets. For perceiver i and target j , we predict J_{Cij} using T_{Cij} and B_{Cij} :

$$J_{Cij} = b_{0i} + t_i T_{Cij} + b_i B_{Cij} + E_{Cij}. \quad (4)$$

The C subscript means that the variables are all centered with the grand mean of the truth across perceivers and judgments. The three effects previously discussed in Equation 1 now vary by person: b_{0i} is the directional bias for perceiver i , t_i is the truth force for perceiver i , and b_i is the bias force for perceiver i . For this model, we can examine how much variance there is due to perceiver for the truth and bias forces. In Case 3, we discuss how the within-person correlation between the truth and the bias forces can be estimated.

We have introduced the theoretical components of the T&B model and have demonstrated how the model can be used to examine the extent to which bias leads to accuracy. In the next section, we review current theoretical approaches to the study of accuracy and bias and discuss how the theoretical components of the T&B model borrow from each of these approaches to form one integrative model.

Theoretical Approaches to the Study of Accuracy and Bias

Although the study of accuracy and bias in human judgment has been a topic among psychologists for many decades, only a limited number of formal theoretical models exist for the study of accuracy and bias as general perceptual processes. We focus our review on the three models that have dominated theoretical and empirical advancement in the study of accuracy and bias: Brunswik's lens model approach, the Gibsonian tradition, and signal detection theory. There are several other models of accuracy and bias that have stemmed from these three primary approaches (e.g., the realistic accuracy model; Funder, 1995). Moreover, there are also models of accuracy that are more methodological than theoretical (Biesanz, 2010; Cronbach, 1955, 1958; Kenny & Albright, 1987). Our presentation of these three formal models is brief, and we in no way do justice to the many intricacies and subtleties of each. Our purpose is to describe the theories very generally and explain how they relate to the T&B model. We borrowed heavily from the ideas and concepts of each of these models in formulating the T&B model. We begin with Brunswik's lens model.

Brunswikian Lens Model Approach

Brunswik's lens model (Brunswik, 1955; Castellan, 1973; Hirsch, Hammond, & Hirsch, 1964; Tucker, 1964) has been used to assess the accuracy of many types of social and nonsocial judgments (for a recent review, see Karelaia & Hogarth, 2008). In the domain of social and personality psychology, most applications of the model are found in the study of accuracy of personality judgments (e.g., Borkeau & Liebler, 1992; Gosling, Ko, Mannarelli, & Morris, 2002). According to a lens model approach, elements in the environment (i.e., proximal variables) serve as a lens through which individuals perceive underlying constructs (i.e., distal variables). For example, a perceiver who observes a target's office as clean, uncluttered, and organized may perceive a target as conscientious (Gosling et al., 2002). In the lens model, *cue validity* refers to the link between the observable cue (i.e., proximal variable) and the target's true standing on the trait (distal variable); in this example, cue validity is the extent to which the uncluttered desk is a valid indicator of conscientiousness. *Cue utilization* refers to the link between the observable cue (uncluttered desk) and the judgment (conscientiousness). The lens model presumes a probabilistic model of accuracy: No single environmental cue or combination of cues allows for the perfect prediction of a target's "true" standing on a trait, and thus accuracy is at best probabilistic. The model has been depicted within a linear framework (Hirsch et al., 1964), in which accuracy of judgment depends on the predictability of the environment and the extent to which the weights that perceivers attach to the cues match the true weights within the environment (Hogarth & Karelaia, 2007; see also An-

derson, 1981). Brunswikians do examine the relationship between truth and judgment, which they refer to as *achievement*, and determine how the set of cues explains this relationship. Ideally, the cues should entirely explain the relationship between the truth and the judgment.

From a lens model approach, accuracy is achieved indirectly through the utilization of valid cues, and so the direct effect of the truth on judgment is implicitly presumed to be zero. Accuracy is then assumed to be fully mediated by the cues. Thus, the greatest difference between Brunswik's approach and the T&B model is as follows: In the latter, the truth is directly measured and is considered as a predictor of judgment, and the extent to which it directly predicts judgment is an empirical question. In Brunswikian analysis, however, the truth is not used to predict the judgment. The two models do share a core similarity: The concept of *cues* from a Brunswikian approach is akin to the concept of *biases* in the T&B model. In Brunswik, cues are variables that may or may not be correlated with the truth, and the extent to which they are utilized predicts accuracy. Cue utilization from a lens model perspective parallels the concept of the bias force in the T&B model. To the extent to which cues are utilized by the perceiver, the bias force would be strong. Cue validity in Brunswik has its parallel in the *a* path in T&B. If these cues were valid, accuracy would be achieved indirectly via bias (i.e., accuracy would be fully "mediated" by the bias force). Thus, the notion that cues fully mediate accuracy can be tested within the T&B framework. If a Brunswikian model is supported, accuracy would be fully mediated by cues (i.e., there would be no "direct effect" of the truth on judgment). In a Brunswikian analysis, this mediation is typically assumed, whereas in T&B, it becomes an empirical question.

Gibsonian Approach

The Gibsonian approach to accuracy (Gibson, 1966, 1979) differs considerably from the Brunswikian approach. Within the Gibsonian tradition, the perceiver is embedded in the environment and perception is oriented to action; through performing action, the perceiver is better able to survive. According to the Gibsonian approach, perception is said to be *direct* in that the information in the object being perceived is directly known by the perceiver; it is not mediated by cues that are imperfectly correlated with the truth. Moreover, perceivers view an object in terms of the behaviors that it affords, that is, what "[the object] offers the animal, what it provides or furnishes, either for good or ill" (Gibson, 1979, p. 127). Affordances are not in the object but in the potential relationship between the perceiver and object in the environment in which they live (McArthur & Baron, 1983; Zebrowitz & Collins, 1997); for this reason, the same object may afford different things to different perceivers. For example, a large puddle in the road may afford "jumpability" to a child but "slipability" to someone who is elderly and infirm. Because from a Gibsonian perspective the focus is on accuracy of affordances (i.e., whether the perceived opportunity for acting, interacting, or being acted upon is realized by the perceiver), understanding accuracy in perception requires that the relationship between the perceiver and target of judgment is understood. Moreover, from a Gibsonian approach, the major interest is not so much in the accuracy of judgment but in the successfulness of action. For example, whether a perceiver accurately judges the width of a doorway is not so important; whether,

based on his perception, that perceiver is able to fit through the doorway is what is of interest. Even if the perceiver overestimates the width of the doorway, what is important is that he is still able to fit through it.

The T&B model adopts more of a Gibsonian approach than a Brunswikian approach in its conceptualization of how the truth affects judgment. As in Gibson's approach, in which affordances of objects are directly perceived and are not mediated by cues, in the T&B model, the truth is measured by the experimenter and its direct effect on judgment is examined. However, the conceptualization of bias in the T&B model is quite different from that within the standard Gibsonian approach. According to Gibsonians, perceivers may make systematic mistakes (i.e., may be biased), but those biases serve an adaptive function. For example, perceivers may systematically underestimate the length of objects such as sticks when trying to reach another object (McArthur & Baron, 1983); however, this bias may have an evolutionary advantage in that it is almost always better to have too long a stick than too short a stick. In the T&B model, whether the bias serves an adaptive function is an empirical question, not a theoretical one. Mistakes in perception occur when perceivers miss what an object really affords, or when perceivers' knowledge of the environment does not permit adaptive action.

Signal Detection Theory

One of the most influential theories of accuracy and bias in human judgment is signal detection theory (SDT; Green & Swets, 1974). In SDT, classically, both the truth and the judgment are dichotomous variables. For example, a perceiver might be asked to watch 50 videotapes in which 25 targets are lying and 25 are telling the truth; the perceiver is asked to judge whether each target is lying or telling the truth. In SDT, by knowing both the false alarm rate (the percentage of times in which the perceiver believes the target is lying when he or she is telling the truth) and the hit rate (the percentage of times in which the perceiver believes the target is lying when he or she is in fact lying), it is possible to estimate the two SDT parameters: the perceiver's sensitivity, called d' , and the perceiver's bias, called B (not to be confused with the B in T&B).

Error management theory (Haselton & Buss, 2000) can be viewed in terms of SDT. This model proposes that some errors are more costly in terms of reproductive success than others, and people will commit those errors in judgment that are the least costly overall. Consider, for example, a man's decision to approach a woman at a bar with the goal of mating with her (McKay & Efferson, 2010). When attempting to seduce a woman at a bar, a man might be biased to assume that the woman whom he approaches is interested in him. If the bias is wrong and she actually is not interested in him, it might leave the man with a slap in the face (a false alarm). Alternatively, he might correctly assume that the woman is interested in him and wants to mate with him (a hit). According to error management theory, the man should primarily be motivated to minimize errors that have reproductive costs (i.e., failing to try to seduce women who would mate with him).

We can directly translate SDT into the T&B model. We estimate a logistic regression in which the judgment (e.g., is the target lying or telling the truth?) is the criterion and the truth (e.g., is the target

actually lying or telling the truth?) is the predictor. The intercept in this logistic regression equation is closely related to B , and the coefficient for the truth is closely related to d' (DeCarlo, 1998). Thus, we can view SDT as similar to the T&B model, in that bias and truth both have direct effects on judgment. In SDT, d' represents the truth force and B is similar to directional bias in the T&B model. Note that B in SDT presumes both alternatives are equally likely, whereas directional bias in T&B is adjusted by the actual likelihood of the two alternatives.

Illustrative Examples

In this section, we provide four empirical case examples to illustrate the T&B model. Case 1 demonstrates the basic T&B model, which we elaborate to consider the role of a moderator in Case 2. In Case 3 we consider how the relationships between bias and accuracy are examined using a within-person approach, in which each perceiver makes multiple judgments over time. In Case 4 we demonstrate the T&B model with a moderator that is hypothesized to moderate directional bias and the truth force.

To illustrate the T&B model, we use two data sets. For Cases 1 through 3, we use a roommate data set that consists of 65 pairs of roommates who had not requested to live together (for method details, see West, Pearson, Dovidio, Shelton, & Trail, 2009). Beginning in the third week of the semester, participants made twice-weekly online ratings of their relationship with their roommate, for 5 1/2 weeks. We examine accuracy and bias for hurt feelings, which is measured with the item "Over the last several days, I felt hurt by my roommate [my roommate felt hurt by me] during our interactions" using a 1 (*not at all*) to 7 (*very much*) scale. Thus, we have self-judgments and perceived roommate judgments of hurt feelings. As a measure of closeness, a composite variable was created of the following 5 items: "It is easy to express who I really am when I am with my roommate," "I was completely myself when I was around my roommate," "I disclosed things to my roommate about my personal life," "My roommate understood me," and "My roommate is an excellent judge of my character" ($\alpha = .966$). The original scores of closeness can vary from 1 (*not at all*) to 7 (*very much*), and we have grand mean centered the variable ($M = 5.355$). The moderator we used is the average of closeness of the perceiver to the target and of the target to the perceiver.

We chose this example to illustrate how the relationship between bias (i.e., the bias of assumed similarity) and accuracy may change depending on the level of closeness. We hypothesized that when the relationship is close, the bias of assumed similarity increases accuracy because perceivers are motivated to see themselves as similar to their partner and because they actually are similar. However, when closeness is relatively low, accuracy is more likely to be achieved directly because perceivers will base their judgments on what actually they observe in their roommate.

For Case 4, we use a data set in which 74 college students and their mothers made ratings of several student behaviors. We focus on accuracy and bias for cheating behaviors. The students were asked to indicate on a 1–5 scale whether they had cheated on an exam or a class assignment (1 = *never*, 2 = *rarely*, 3 = *a few times*, 4 = *I have cheated a moderate amount*, 5 = *I have cheated often*). Within four days of students completing the survey, mothers made judgments of their child's cheating behav-

iors on the same 1–5 scale. Mothers were also asked to indicate how upset they would be if they found out that their child had cheated on a scale from 1 (*not at all upset*) to 5 (*extremely upset*). The upset variable was transformed to range from 0 (*not at all upset*) to 1 (*extremely upset*). We later refer to this variable as *threat*. Recall that we have hypothesized that the same underlying mechanism can increase bias and decrease accuracy in some cases. We chose this example to illustrate that threat—or the fear that one’s child is engaging in a bad behavior—can motivate both bias and accuracy but in opposite ways, leading to an inverse relationship between the truth and bias forces.³

Case 1: Truth and Bias Forces Studied Between Participants

In Case 1, we use the roommate data to examine accuracy and the bias of assumed similarity for judgments made at only the first time point of the study. Perceiver i is asked to predict his or her roommate’s hurt feelings, and we denote that judgment as J_i and the roommate’s actual hurt feelings as T_i . The judgment and truth have the same units of measurement. We center the judgment and truth using the mean of the truth variable, and we denote these new variables as J_{Ci} and T_{Ci} . In this example, the bias of interest is assumed similarity (i.e., assuming that one’s roommate shares the same feelings as oneself), and we denote the bias variable as B_i . In this case B_i has the same units of measurement as the truth, and so we center B_i using the truth mean and denote it as B_{Ci} . Referring to Equation 1, b_0 is the directional bias, t is the truth force, and b is the bias force. Recall that b_0 is the product of the directional bias value times the bias force. If b_0 is positive, perceivers are biased to see the roommate as having more hurt feelings than the roommate actually has, and if it is negative, perceivers are biased to see the roommate as having fewer hurt feelings than the roommate actually has. Assuming a normal distribution of errors, the proportion of those who are biased to see their partner as higher on the scale than the partner actually is (i.e., to see the partner as having *more* hurt feelings) is given by $P(Z > b_0/S_E)$, where Z is a standard normal deviate, s_E is the standard deviation in the errors of judgment (see Equation 1), and P is the probability. The proportion of those who are biased to see the roommate as lower on the scale than the roommate actually is (i.e., to see the roommate as having *fewer* hurt feelings) is given by $P(Z < b_0/S_E)$. The statistical test that b_0 equals zero evaluates whether the mean of the judgments equals the mean of the truth value, when the bias variable is at the mean of the truth.

Recall that it is possible to be accurate by being biased (Gagné & Lydon, 2004; Kenny & Acitelli, 2001). In this example, perceivers may accurately know their roommate’s hurt feelings not because they perceive this directly, but because they assume that their roommate feels like themselves and are correct in that assumption. Recall that we can decompose total accuracy into two sources of accuracy: indirect accuracy and direct accuracy.

As we described earlier, we then can use mediational methods to estimate total accuracy using

$$t_T = t + ab, \quad (5)$$

where t_T is total accuracy, and t and b are from Equation 1. Parameter a is measured from the following equation in which the truth variable predicts the bias variable:

$$B_{Ci} = b_{00} + aT_{Ci} + E_i. \quad (6)$$

Note that when B is a self-judgment (as it is for our example), a can be interpreted as the actual similarity parameter. We can do a similar decomposition of the directional bias. The equation is

$$b_{0T} = b_0 + b_{00}b, \quad (7)$$

where b_{0T} is total bias, b_0 and b are from Equation 1 (the direct accuracy model), and b_{00} is from Equation 6. As we discussed earlier, indirect directional bias, or the bias due to B , is equal to $b_{00}b$, which equals $b - (B - T)$. Indirect directional bias refers to a directional bias that is due to the bias variable, B .

We can also ask at what bias value the directional bias is canceled out (i.e., perceivers are unbiased). In this example, how hurt do perceivers need to feel (i.e., at what value on the bias variable) in order for them to be unbiased, in terms both of assumed similarity and of directional bias? To answer this question, we can combine the directional bias and the variable that measures the bias of assumed similarity (i.e., perceivers’ self-ratings) to form the variable $b_0 + bB_i$. The point at which perceivers are unbiased is when B_i equals the value $-b_0/b$. It is possible that $-b_0/b$ is not a plausible value for B_i , both in terms of its theoretical likelihood and its possible range of values given the measure, which would mean that no one is on average unbiased. The bias for perceiver i at the mean of T and a particular value of B_i is given by $b_0 + bB_i + E_i$, which has a variance of s_E^2 . We can also examine the percentages of people who are over- and underestimating their roommate’s hurt feelings. Assuming a normal distribution of errors, the proportion of overestimation is given by $P\left(Z > \frac{b_0 + bB}{s_E}\right)$, where Z is a standard normal deviate and the proportion of underestimation is given by $P\left(Z < \frac{b_0 + bB}{s_E}\right)$.

Example. For Time 1, the mean judgment is 1.543 and the mean truth value is 1.585. Because we have dyadic data and roommates are both perceivers and targets, the mean of B is the same as the mean of T . We subtracted 1.585, the truth mean, from T , J , and B .

Estimating Equation 1, we find that

$$J_{Ci} = -0.042 + 0.201T_{Ci} + 0.539B_{Ci} + E_i,$$

where the bias force (0.539) is statistically significant ($p < .001$), the truth force (0.201) is statistically significant ($p = .004$), and the intercept (-0.042) is not significantly different from zero ($p = .58$), indicating a negative but nonsignificant directional bias. The variance of E is estimated as 0.547. We see that the coefficient for the assumed similarity bias force is more than twice the size of the coefficient for the truth force and that it is statistically larger ($p < .001$). When we estimate Equation 6, we find that

$$B_{Ci} = 0.000 + 0.344T_{Ci} + F_i.$$

Indirect accuracy equals $(0.344)(0.539)$, or 0.185. Thus, 48.0% [$(100)0.185/(0.201 + 0.185)$] of accuracy is due to the bias of

³ For each of the four cases, we describe in greater detail the technical details in a supplementary appendix (available at davidakenny.net/t&b/technical.doc) and a spreadsheet illustrating the computations (available at davidakenny.net/t&b/comp.xls).

assumed similarity. We also see that indirect directional bias equals zero because b_{00} equals to zero.

At what value on the bias variable does the directional bias cancel out the bias of assumed similarity? If we use the formula $-b_0/b$, perceivers would be unbiased, both in terms of the bias of assumed similarity and in terms of directional bias if they scored 1.663 $[-(-0.042/0.539) + 1.585]$ on B (i.e., their self-rating of hurt feelings). For persons scoring at the mean of B , 52.3% underestimate their roommate's hurt feelings and 47.7% overestimate their roommate's hurt feelings. However, for those scoring 2 on B , 30.7% underestimate and 69.3% overestimate their roommate's hurt feelings. For those scoring 3 or more on B , virtually everyone overestimates the hurt feelings of their roommate. Finally for those scoring 1 on B , 80.2% underestimate and 19.8% overestimate their roommate's hurt feelings.

Case 2: Truth Force, Bias Force, and a Moderator

We now add to Case 1 the variable closeness of the two dyad members, a moderator that we denote as M_i . We also have J_{Ci} , T_{Ci} , and B_{Ci} as before. We estimate the product-term interactions between closeness and the truth variable and between closeness and the bias variable. Adding the moderator to the equation in Case 1 produces the equation for Case 2, which is given earlier as Equation 3. It can help to rearrange the terms of Equation 3 as follows:

$$J_{Ci} = (b_0 + mM_i) + (t + t_M M_i)T_{Ci} + (b + b_B M_i)B_i + E_i \quad (8)$$

With this second formulation, we see that directional bias, the truth force, and the bias force all can change as a function of the moderator. The main effects of these forces refers to their effects when the moderator equals zero, and the interactions of these forces with the moderator refers to the degree to which the processes of accuracy and bias change as a function of a one-unit difference in closeness.

When we estimate Equation 8 using the roommate data set with closeness grand mean centered, we find that

$$J_{Ci} = (0.015 + 0.078M_i) + (0.224 - 0.046M_i)T_{Ci} + (0.723 + 0.156M_i)B_i + E_i,$$

where the variance of E is estimated as 0.529. The directional bias is slightly positive and not statistically significant different from zero ($p = .87$). The bias force (0.723) is statistically significant ($p < .001$), and the truth force (0.224) is marginally significant ($p = .051$). The main effect of the moderator is not statistically significant ($p = .41$), indicating that closeness does not have a statistically significant effect on directional bias, but it does interact with the bias force ($p = .029$), indicating that the bias force increases as closeness increases. Results also indicate that closeness did not moderate the truth force ($p = .52$), but the direction of the effect is that people are less accurate as closeness increases.

We can determine direct, indirect, and total accuracy for different levels of the moderator. Recall that direct accuracy refers to the effect of T controlling for B , which we obtain from Equation 8, and it would equal $t + t_M M_i$. Determining indirect accuracy is more complicated. To do so, we first determine the "effect" of T on B for

different levels of the moderator (i.e., path a in Figure 2) by estimating the following equation:

$$B_i = b_{00} + qM_i + (p + zM_i)T_{Ci} + F_i, \quad (9)$$

where the parameter p represents a measurement of actual similarity and z represents the extent to which actual similarity is moderated by M . From Equation 8, we can see that the effect of B on J controlling for T is $b + b_B M_i$. Thus, the indirect effect of accuracy equals $(p + zM_i)(b + b_B M_i)$ or, in terms of Figure 2, the product of paths a and b , each moderated by M . When we expand this, we obtain $pb + (zb + pb_B)M_i + zb_B M_i^2$. From this expansion, we see that although we have assumed that the moderator has only linear effects on the truth force, the bias force, and the directional bias, the moderation of the indirect effect is quadratic because M_i^2 is in the equation (Edwards & Lambert, 2007). That is, when we multiply the moderated path a by the moderated path b , we obtain a quadratic effect of the moderator.⁴ Total accuracy for a given value of M_i is given by $(t_d + t_M M_i) + (b + b_B M_i)(p + zM_i)$. Total accuracy reaches a maximum (if $zb_B < 0$) or minimum (if $zb_B > 0$) when M_i is equal to $-(t_{MT} + pb_B + bz)/(2zb_B)$, and indirect accuracy reaches a maximum or minimum when M_i is equal to $-(pb_B + bz)/(2zb_B)$. For the example, we estimated the coefficients in Equation 9:

$$B_i = -0.116 - 0.405M_i + (-0.174 - 0.224M_i)T_i + F_i.$$

The coefficient for T , -0.174 , is the actual similarity parameter, which we see is negative. Moreover, as closeness increases, actual similarity decreases. Roommates are similar on hurt feelings when they are not close to each other and dissimilar when they are close to each other. With the coefficients from Equations 8 and 9, we can estimate how total, direct, and indirect truth force vary as a function of closeness, which we plot in Figure 3. We see in the figure that the relationship between direct accuracy and closeness is negative, with a small decline as closeness increases. However, indirect accuracy shows a quadratic relationship with closeness. It maximizes at 2.650 and then declines and actually becomes negative at high levels of closeness. Based on these results, it is reasonable to conclude that closeness leads to lower accuracy because of the bias of assumed similarity, just the opposite to what we expected. That is, increasing closeness leads to lower actual similarity but more assumed similarity. However, we do find, as predicted, that direct accuracy declines with closeness, but this effect is not statistically significant.

The bias term for Case 2 equals $b_0 + mM_i + (b + b_B M_i)B_i$. If we set B_i equal to T , one is unbiased when the moderator equals $-b_0/m$. When the moderator is zero, one is unbiased when B equals $-b_0/b$. For the example, one is unbiased when closeness equals 5.163 and unbiased on B when it equals 1.564. Assuming a normal distribution of errors, the proportion of overestimation for a given value of B_i and M_i is given by

$$P\left(Z > \frac{b_0 + mM_i + (b + b_B M_i)B_i}{s_E}\right),$$

⁴ If we were to estimate the total effect of T on J , we would need to include in the regression equation the quadratic, as well as the linear, effects of M .

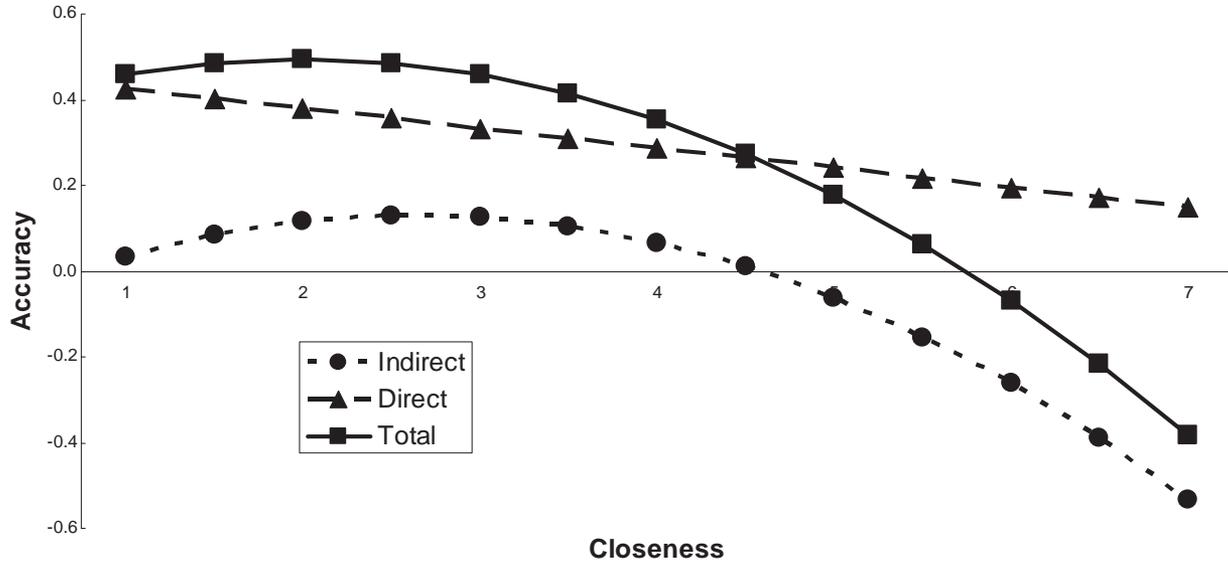


Figure 3. Indirect, direct, and total accuracy as a function of closeness for Case 2.

where Z is a standard normal deviate and the proportion of underestimation for a given value of B_i and M_i is given by

$$P\left(Z < \frac{b_0 + mM_i + (b + b_B M_i)B_i}{s_E}\right).$$

For instance, for those who are very close to their roommate ($M_i = 7$, uncentered) and have a score of 3 on B (uncentered), only 4.2% underestimate and 98.2% overestimate their roommate's hurt feelings, whereas for those who have the lowest score on closeness ($M_i = 2.47$, uncentered) and a score of 1 on B (uncentered), 61.4% underestimate and only 38.6% overestimate their roommate's hurt feelings.

Case 3: Within-Participant Analysis

In many studies of accuracy, particularly those involving nonsocial judgments, participants make many judgments, not just one as we have considered so far. As we shall see, with such data we can measure the truth and bias force for each participant. We use all of the roommate data (i.e., all 11 time points) to illustrate Case 3.

We also center the variables by subtracting the grand mean of the T across persons and times. For perceiver i , we predict J_{Cij} using T_{Cij} and B_{Cij} to obtain

$$J_{Cij} = b_0 + b_{0i}' + (t + t_i')T_{Cij} + (b + b_i')B_{Cij} + E_{ij}, \tag{10}$$

where the fixed effects are b_0 (the average directional bias across persons), t (the average truth force across persons), and b (the average bias force across persons). The model has three random effects, which vary by person: b_{0i}' is the directional bias for person i ; t_i' is the truth force for person i ; and b_i' is the bias force for person i .

The three random effects in Equation 10 can be intercorrelated. Several questions are of interest: Are persons who are more accurate more or less biased? If one person has a strong truth force,

does his or her partner have a strong truth force? The error term in Equation 10, E , represents day-to-day variation.

The mean for the Truth averaged over participants and times is equal to 1.454. All variables were centered using this value. The estimated equation is

$$J_{Cij} = -0.099 + b_{0i}' + (0.050 + t_i')T_{Cij} + (0.515 + b_i')B_{Cij} + E_{ij}.$$

We discuss the statistical significance of the coefficients below, as well as the random effects and their correlations.

The average directional bias equals -0.099 , which implies that people tend to think that their roommate has fewer hurt feelings than he or she actually does ($p = .001$). The variance in the directional bias is 0.041 ($p < .001$), and about 68.8% of the time across all days there is a negative directional bias (i.e., a bias to see their roommate as less hurt by them than he or she really is). If we look at a particular day, we have to add in the error variance, which equals 0.211. On any given day, for any given person, 57.8% of the time that person assumes that the roommate's hurt feelings are less than they really are.

The coefficient for the truth force is small, 0.050, and not statistically significant ($p = .108$). We see later that most people are accurate but that accuracy is due to the bias of assumed similarity. We find individual differences in the truth force in that the variance of the truth force is 0.064 ($p < .001$). Those who are one standard deviation above the mean in the truth variable have a truth force of 0.303 and those who are one standard deviation below the mean show negative accuracy of -0.204 . We find that about 42.3% of the participants have negative accuracy, the interpretation of which we return to in the discussion.

We do see that people are strongly biased, in terms of the bias of assumed similarity, with a coefficient of 0.515 ($p < .001$). If a perceiver has particularly hurt feelings one day, he or she sees his or her roommate the same way. There is variability in the bias

force across persons of 0.116 ($p < .001$). Those who are one standard deviation above the mean on the bias variable have a bias force of 0.855, and those who are one standard deviation below the mean on the bias variable have a bias force of 0.174. Almost all perceivers assume similarity (93.5%). We find a small and statistically nonsignificant positive correlation between the truth force and the bias force for assumed similarity ($r = .271, p = .188$), indicating that the bias force and the truth force do not trade off. However, there is a moderate negative correlation between directional bias and the truth force ($r = -.459, p = .016$), indicating that perceivers with a stronger truth force show more negative directional bias. The correlation between directional bias and the bias force for assumed similarity is .682 ($p = .002$), indicating that the two bias forces are significantly related. Thus, persons who assume more similarity are more likely to think that their roommate has hurt feelings. Finally, we note that the two roommates do show any similarity in directional bias, accuracy, or bias ($ps > .79$).

Although we do not find that people are accurate, we do find that they are accurate perhaps because they are biased. There is indeed some evidence of actual similarity 0.231 ($p = .104$), although this effect is not significant. Thus, indirect accuracy equals 0.119 (actual similarity, 0.231, times assumed similarity, 0.515). Note that indirect accuracy is more than twice as large as direct accuracy. Thus, the data may show that people are accurate because they are biased.

We see for Case 3 that a wealth of results can be obtained. The T&B model is particularly helpful here in that it serves to organize these many results.

Case 4: Threat Moderator

For this last case, we return to the case in which each perceiver makes a single judgment. As we have discussed, in some cases the moderator might be chosen to affect either the bias force or the truth force. Here we have a moderator that is presumed to increase the directional bias force and to weaken the truth force. Within the area of close relationships, such a moderator might be called a *threat* moderator. We start with the following equation, which is a simplified version of Equation 3:

$$J_{Ci} = (b_0 + mM_i) + (t + t_M M_i)T_{Ci} + E_i \quad (11)$$

where both the judgment and truth are centered by subtracting the mean of the truth from every score. We scale the moderator, M_i , such that it ranges from 0 (no threat) to 1 (maximal treat). The truth force is zero when M_i equals $-t/t_M$. Given that no threat is coded as 0, directional bias equals b_0 when there is no threat and $b_0 + m$ when threat is at the maximum. Moreover, when threat is at the maximum, the predicted value of J uncentered equals $b_0 + m + (t + t_M + 1)T$.

Example. We turn to the data set in which mothers and college-student children made ratings of the child's cheating behaviors. The mean of T (i.e., how much students reported cheating) is 1.676, and the mean of J (i.e., how much mothers reported that their child cheated) is 1.197. We centered T and J by the mean of T . The mean of how upset mothers would be if their child cheated is 4.535, and we treat this variable as *threat*. We transformed this variable so that it ranges from 0 (no threat) to 1 (maximal threat). The estimated regression equation is

$$J_{Ci} = 0.171 - 0.754M_i + 1.007T_{Ci} - 0.989M_iT_{Ci} + E_i$$

All three regression coefficients are statistically significant ($ps < .007$), but the intercept is not statistically different from zero. Because t_M is negative (-0.989), threat and accuracy trade off (i.e., as threat increases, the truth force declines). We note that the truth force is quite strong, 1.007, for those who score zero on threat, but for those who have the maximal threat ($M_i = 1$), the truth force is essentially zero (0.018) and is not statistically different from zero ($p = .79$). The more upset mothers would be if their child cheated, the less accurate they are. Recall that the point of zero accuracy on the moderator equals $-t/t_M$, or $-(1.007/-0.989)$, which equals 1.018. Thus, mothers would be totally inaccurate when they are most threatened (Threat = 1). The estimate of directional bias at maximum threat is $0.171 - 0.754M_i$, which equals -0.573 . We see that when there is no threat, the directional bias is 0.181 and is not statistically different from zero. However, when threat is at its maximum, the directional bias is negative and the predicted judgment is $-0.573 + 1.018 T$, which equals 1.14, a value virtually identical to the smallest possible value for J (uncentered) of 1. Thus, mothers are biased to see their child as having the lowest possible value of cheating.

Discussion

We presented a formal model for the study of accuracy and bias in perception, termed the T&B model. The model provides a single coherent approach, inspired by the previous theoretical approaches of Gibson, Brunswik, and the SDT, to the study of accuracy and bias. The T&B model can be used to examine many different types of research questions that are at the heart of these three aforementioned theories. For example, a researcher may be interested in understanding the extent to which perceivers use environmental cues in determining how outgoing a target is (Gosling et al., 2002), whether individuals afford the correct properties to objects such as doorways (Warren & Whang, 1987), or how accurate perceivers are in detecting a target person's sexual orientation based on only a thin slice of information (Ambady, Hallahan, & Conner, 1999). Each of these three questions is typically examined from a Gibsonian, Brunswikian, or SDT perspective, respectively, but all of them can be examined within the framework of the T&B model. We begin by discussing in detail theoretical considerations of the T&B model and then discuss measurement and design issues.

Theoretical Considerations

The importance of the truth in the T&B model. The most fundamental part of the T&B model, which is borrowed from the Gibsonian approach, is that in order for one to study accuracy and bias in perception, the truth must be measurable and measured. It is true that in some domains of perception, the truth is neither meaningful nor measurable. For example, in Balci and Dunning's (2006) research on motivated visual perception, perceivers view an ambiguous object that could be perceived as a seal or a horse in one case and the number 3 or the letter B in another case. What is of interest is not whether perceivers are "right" in their perception of the object but what they perceive the object to be depending on their motivational state. For such types of judgments, what is of interest is bias but not accuracy. Thus, the T&B

model could not be used where there is no theoretical or empirical truth criterion. However, in many areas of study the truth does exist and is measurable. We have provided a brief discussion of the importance in defining a proper truth criterion that is properly tailored to the specific accuracy question, particularly in studies of accuracy in the domains of personality and social psychology. Researchers should take into consideration many factors in determining a proper truth criterion: They need to establish and make clear what they are interested in testing the accuracy of (e.g., are they interested in examining the accuracy of a target's subjective emotional experience or what emotions that target expresses), at what level of judgment it is most appropriate to study their accuracy question (e.g., should they study accuracy in perceptions of how targets behave with particular perceivers or how targets behave across many perceivers), and what is the most appropriate context to study.

In all of the case examples, truth criteria were obtained using one method of measurement: self-report. In some domains, the truth is best captured using a set of methods that converge on the same answer. For example, in the study of accuracy of personality judgments, a researcher might be confident that she has captured a target's true standing on agreeableness only if she has measured how agreeable friends and family perceive the target to be, how agreeable the target is in a social setting, and how agreeable work colleagues perceive the target to be. Only if there is convergent validity across all of these measures can the researcher be confident that she has obtained a proper measurement of the truth.

Bias in the T&B model. In addition to emphasizing the importance of measuring the truth, the T&B model clarifies the many different meanings of bias in perception. Perhaps the most fundamental difference between the T&B model and previous theoretical approaches to accuracy is that in the T&B model, the relationship between bias and accuracy can be empirically examined. According to a Gibsonian approach, perceivers are biased when they make errors in judgment that are useful; in a Brunswikian approach, perceivers are biased when they utilize invalid cues or fail to utilize valid ones. In both of these instances, bias is defined by its relationship to accuracy. The T&B model adopts an approach that has been advocated by Funder (1987) and colleagues (Krueger & Funder, 2004): Bias, although theoretically distinct from the truth, can be correlated with the truth, and therefore perceivers can become accurate by being biased. As such, bias is theoretically distinct from error.

The T&B model makes an explicit formal distinction between bias *values* and bias *forces*. Very often in the literature the term bias is used to refer to each, yet they are different. In models in which the bias force is measured with a variable, the value of that variable by itself does not provide any information about how biased people are in their judgments of the target. In the example that we used in the first three cases, the bias variable is the self-judgment. Whether perceivers themselves experienced hurt feelings in their interactions with their roommates does not provide any insight into how biased they are in their perceptions of their roommate's hurt feelings. However, the bias force (i.e., the strength of the effect of self-perceptions on judgments) captures the presence and strength of the bias. We can measure the bias force for different perceivers, as we did in Case 3. In the T&B model, we have demonstrated a method of examining different types of bias within the same model. Although in our case exam-

ples we examined a single bias variable as a predictor of judgment, it is possible to expand the model to examine multiple bias forces simultaneously. By including multiple bias variables in one model, one examines the effect of one bias while controlling for the effects of the others (e.g., to simultaneously examine the extent to which judgments of romantic partners are driven by judgments of ideal partners and by self-judgments; Murray et al., 1996).

We also introduced the notion of a directional bias, that is, a bias in which the bias value is the same for each perceiver and represents both the direction in which judgments are being pulled and how far away judgments are from the truth. The concept of directional bias is borrowed from the concept of bias in SDT where perceivers are pulled more toward one response than the other. One way of thinking about directional bias, particularly with rating data, is that the perceiver is pulled toward the two scale endpoints.

When directional bias is measured, separating the bias force from the bias value can be quite a challenge. We denote U as the uppermost response and L as the lowermost response; the directional bias can be thought to equal $b_U U + b_L L$, where b_U and b_L are the two forces. Because J is centered using T , the directional bias, or b_0 , equals $b_U(U - \bar{T}) - b_L(\bar{T} - L)$. The problem is that we have one equation with two unknowns, b_U and b_L . If we can assume that one of these forces is zero (e.g., romantic partners are pulled only to see their partner positively; i.e., $b_L = 0$), then the bias force, or b_U , can be determined as $b_0/(U - \bar{T})$. Alternatively, we might decide to fix the value of one of the forces to a constant (e.g., to 1) and then solve for the other force. So if we set b_L to 1, the solution for b_U is $(\bar{T} - L + b_0)/(U - \bar{T})$. A final idea is that we might constrain $b_L + b_U$ to equal a set value. If we make that value be 1, then we have the weight for b_U being $(\bar{T} - L + b_0)/(U - L)$. We can see that computing the directional bias force is not very straightforward. That is why we currently recommend just looking at b_0 , which can inform us about the relative strength of the two forces without knowing quantitatively the value of that force. In principal, however, knowing the force would be instructive for determining how it compares to b and t .

Moderators in the T&B model. A key feature of the T&B model is that it makes a theoretical distinction between moderator variables and bias variables. To date, there has not been a formal distinction made between them, and so the same variable is treated as a moderator in one study and as a bias variable in another. The T&B model has specific guidelines for determining whether a particular variable is a moderator or a bias variable: Moderators do not *directly* influence judgment (e.g., whether perceivers see their roommate as hurt or not hurt; whether mothers believe that their son or daughter has cheated); rather, they inform the process of accuracy and bias by influencing the truth and bias forces. Moderator variables may influence the strength and direction of the forces independently of each other. For example, men may be more biased than women, but they may be just as accurate as women. In addition, moderators can also inform the relationship between the forces. As we illustrated in Case 2, perceivers can be accurate by being biased for some levels of the moderator but not for others. In our case examples, we demonstrated how a variable measured at the level of the relationship (i.e., closeness) moderated the forces. The T&B model can also be used to examine moderators that vary by perceiver, target, and variables that are experimentally manipulated. For example, for Case 2, we could have

considered how the perceiver's and the target's closeness each uniquely moderate the accuracy and bias forces. In addition, as we illustrated with Case 3, *person* can also be a moderator in that some persons may be more accurate and more biased than others. More interesting, the within-person variance provides insight into whether persons who are accurate are also biased. Likely of most theoretical relevance, experimental manipulations might be created to alter perceiver motivations to determine their effects on the truth and bias forces. For example, there may be a context in which the truth and bias forces are theoretically proposed to be inversely related, and so an experimental manipulation might improve the truth force and decrease the bias force.

In sum, moderator variables influence the strength, direction, and relationship between the truth and bias forces, whereas bias variables exert an effect directly on judgment by pulling it in a particular direction.

The accuracy–bias relationship in the T&B model. To understand the simultaneous processes of accuracy and bias, one must understand the relationship between them. The T&B model delineates four major ways in which the truth and bias are related. First, the truth and bias values may be correlated. This is represented in the model as path *a* in Figure 2, the “path” from *T* to *B*. If perceivers are biased and path *a* (and *b*) is positive, perceivers are accurate by being biased (Cronbach, 1955), and thus bias mediates accuracy. For example, imagine that perceivers make judgments of a target's sexual orientation based on the target's gait. Gait may be highly correlated with sexual orientation, and so perceivers may become accurate in their judgments of sexual orientation via their perceptions of gait. Brunswikians believe that a set of bias variables completely mediates the truth force, and so the effect of the truth force on the judgment is zero when the bias variables are considered. For example, judgments of a target's sexual orientation may be based on a combination of dynamic physical traits. These traits themselves are not sexual orientation, but they may mediate accurate judgments of sexual orientation.

Second, we can compare the relative strengths of the two forces. For example, in Case 1, we saw that perceivers were much more biased than they were accurate, and we evaluated statistically that the bias force was indeed larger than the truth force.

Third, we can examine the relationship between accuracy and bias across persons, when the same perceivers make multiple judgments (e.g., across multiple time points). If a perceiver is especially accurate, is he or she also especially biased? The within-person correlation between the truth and the bias forces provides insight into the intrapersonal processes of accuracy and bias. As we saw in Case 3, there is little or no relationship between the truth force and the bias force of assumed similarity. We can also examine the correlation between the directional bias and the truth force.

Fourth, when we have moderators in the study, we can examine the similarities in the psychological mechanisms that drive accuracy and bias by examining what predicts each. Kenny and West (2010) used a meta-analytic approach to demonstrate that the processes of self- and other-perception are similar if the same set of variables predicts them in the same way. If we apply Kenny and West's approach to the T&B model, it may be of interest whether the same underlying psychological processes drive people to be accurate and to be biased. As previously discussed, in the close relationships literature, there has been ample theorizing about how

perceivers are both accurate and biased and what predicts the two processes. Recent reviews have found that the same types of people are both biased and accurate (i.e., those who are in satisfied relationships and feel close to their partners; Gagné & Lydon, 2004; Murray, 1999). With the T&B model, it is possible to evaluate empirically whether the psychological processes of accuracy and bias are the same in close relationships. By examining whether the same psychological factors predict both of them in the same way, we can examine this question. For example, we found in Case 2 that closeness moderated the bias force, but it did not moderate directional bias or accuracy; these results suggest that the underlying psychological processes that drive accuracy and bias are not the same in this example.

Remaining Theoretical Issues

Judgment causing truth. In the T&B model, the truth force is formalized as the effect of the truth on the judgment. However, it can be the case that the truth is caused by the judgment, as is the case with the self-fulfilling prophecy effect (Jussim, 1991). If there were self-fulfilling prophecy effects, then part of the truth force, or *t*, would be due to self-fulfilling prophecy. It should be possible, although difficult, to separate the “real” truth force from the self-fulfilling prophecy part.

Negative accuracy. It is possible in the T&B model to obtain a negative truth force effect. For instance, in Case 3, we found that for nearly half of the participants, their accuracy in knowing whether their roommate had hurt feelings was negative. Thus, the force is quite strong, but rather than judgments being pulled toward the truth, they are being pushed away from the truth. What does it theoretically mean to have “negative accuracy”? It is useful to make the distinction between negative indirect and negative direct accuracy. It might easily happen that the indirect effect of accuracy is negative. In Case 2, perceivers assumed similarity (i.e., path *b* was positive, in Figure 2), but if they were actually dissimilar to their roommate (i.e., path *a* was negative), their indirect accuracy would be negative. Alternatively, perceivers might actually be similar to their roommate, but if they assumed dissimilarity, indirect accuracy would also be negative. When indirect accuracy is negative and direct accuracy is positive, the total effect of accuracy might be very small and even negative, especially when there is more indirect accuracy than direct accuracy.

Explaining negative direct accuracy is more difficult. We think that if a researcher finds negative direct accuracy, there is an unmeasured bias that leads to negative indirect accuracy; however, because this bias is not included in the model, there appears to be negative direct accuracy. Akin to a suppression effect, the direct effect of accuracy appears to be negative, but only because there is an unmeasured negative indirect effect. If this unmeasured bias were included in the model, the direct effect would be positive. For instance, in Case 3, it may be that some targets who had hurt feelings compensated by smiling and being pleasant. This bias variable would have a positive effect on judgments, but its moderating effect on the truth force would be negative. Thus, once this unmeasured negative indirect effect is subtracted, the remaining direct effect would be positive. Thus, the T&B model predicts that direct accuracy must be positive and that if it is estimated as negative, there must be a hidden bias that creates “negative accuracy.”

Measurement Issues in the T&B Model

Units of measurement. In our formalization of the T&B model, it is crucial to have the judgment, truth, and the bias variables all measured on the same scale. We acknowledge that for many types of judgment, the truth and the judgment are on different scales. For example, imagine a study where a researcher is interested in how accurate clinicians are in their assessments of their patients' depression levels. After an initial session, clinicians rate their patients' depression on a 1 (*not at all depressed*) to 7 (*extremely depressed*) scale. Patients then complete the Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996), and their actual depression is measured on a scale that potentially ranges from 0 (*not at all depressed*) to 63 (*extreme depression*). It is possible to use the T&B model to obtain the truth force in this example, but certain key information would be lost. First, the sign of the intercept can no longer be interpreted as either a positive or a negative bias. Thus, the researcher would not be able to test whether the clinicians are biased to over- or underestimate their patients' depression levels. Second, it would also be impossible to determine the percentage of persons who under- and overestimate (in this example, the percentage of clinicians who over- and underestimate depression). Thus, not having T and J in the same metric has some serious costs.

If the bias variable was not measured with the same scale as the truth, the researcher would be unable to compare the relative strength of the truth and bias forces. For instance, consider the case in which a researcher is interested in whether judgments of current patients' depression levels are biased by how depressed clinicians perceived their past patients to be, and judgments are made using the same 1–7 scale with which they judged their current patients. Because the bias variable is measured with the same scale as the truth, it would be possible to compare the relative sizes of the bias and the truth forces. However, if the bias variable was the average prior patients' BDI, comparison of the forces would be impossible.

Very often, the judgment scale and the truth can be measured in the same units. More difficult can be having a bias in the same metric as the judgment. Consider a simple study in which perceivers estimate the target's intelligence in IQ points and the bias (or cue) is the wearing of glasses. Say that persons with glasses are estimated to have 5 more IQ points than those without; that is, the wearing of glasses has a direct effect on judgment, and the direction and force of this effect is of interest. This 5 points represents the combination of the bias force and the difference in bias values between wearing and not wearing glasses. For instance, it might be the case that the bias force was 0.5, which would make the difference in bias values for wearing versus not wearing glasses 10 points. Or it might be that the bias force is 2, if the difference in bias values was 2.5 points. Because the bias is not measured in the same units as the judgment, we cannot separately measure the bias force and difference in bias values. What we suggest doing in this case is focusing on the direction of the force (e.g., does the wearing of glasses have a positive effect on judgments of intelligence) and its confidence interval.

Centering. In the T&B model, centering of the truth, the bias, the judgment, and the moderators affects the meaning of directional bias. Because T&B centers using the truth, directional bias gives a measure of how much judgment is shifted toward one end of the continuum. The centering of the judgment variable is crucial

for understanding the meaning of directional bias. We adopted an approach to similar to that of Fletcher and Kerr (2010) in which the truth mean was subtracted from the judgment. With such a centering strategy, the directional bias indicates how much more positive or negative judgments are from the average truth in the sample.

Subtracting the mean of T presumes that the targets in the study are representative of the population. For instance, people generally think that most people are telling the truth (Bond & DePaulo, 2008). If we had a study with representative stimuli, we would find that the mean of T would indicate that people were telling the truth most of the time. Imagine also in this study that we also find that b_0 was zero, indicating that there was no directional bias. If, however, a study were done in which the stimuli were chosen so that half were telling the truth and half were telling a lie and we center using the sample mean of T , we would probably see that the b_0 was positive in that participants were "biased" to think that people were telling the truth. We note in this case that the "bias" equals $\mu_T - \bar{T}$, where μ_T is the population mean of the truth. Thus, if μ_T is greater than \bar{T} and \bar{T} is used to center T and J , it would appear that perceivers were showing a positive directional bias. The tendency to see others as telling the truth in general leads perceivers astray in the laboratory, but in everyday life, it helps them make accurate judgments (Funder, 1987; Gigerenzer, 1991). Therefore, it is very important to consider the representativeness of the sample when determining the best method of centering the judgment. Note that if a given theory postulates little or no directional bias (e.g., Funder, 1987) and if representative targets are used, directional bias should be very close to zero.

When targets are unrepresentative, the centering should be done using the *population* mean and not the sample mean. Obviously, researchers very rarely exactly know the population mean, and the consequences of using the sample mean is a topic worthy of further study. At minimum, if researchers find directional bias, they should consider whether that "bias" might be due to the sample of targets being unrepresentative.

The issue of centering becomes even more complicated if there are moderators in the model. It might be more appropriate to center based on the moderator. Consider the moderator of gender and the study of accuracy of lie detection and presume that men tell more lies than do women. When we subtract the mean of the T from both T and J , we might use the mean for the appropriate gender (i.e., group mean centering) as opposed to the grand mean. Using the grand mean for this example, we would find the perceivers were biased to think that men lied more than women when in fact that is not an error in judgment but a reality that is correctly perceived. Researchers might consider regressing T on M , computing T , and subtracting \hat{T} and not \bar{T} . As another example, imagine that individuals make multiple ratings over the course of several days of their partner's mood and their own mood (Wilhelm & Perrez, 2004). To understand the meaning of directional bias, one might more appropriately center the judgment on the average mood for each target across days (i.e., person centering) rather than the average mood across days and across partners (i.e., grand mean centering). If the former strategy is used, directional bias refers to whether perceivers under- or overestimate their partner's mood compared to how that person usually feels. Do note that when we

discuss grand mean and person centering, we center not just T but also B and J .

Dichotomous outcomes. Thus far, we have discussed only the case in which judgments are made of continuous variables. Dichotomous outcomes can be examined within the T&B model. We earlier discussed SDT and the use of logistic regression to estimate that model. The intercept from such a model reflects bias in signal detection, which is related to directional bias in the T&B model. (For the intercept to be a measure of directional bias, we would need to find a way to center the truth and judgment variables using the “mean” of the truth.) Unlike in traditional SDT, bias variables could be added to the T&B model, and the bias variables could interact with the truth variable to predict dichotomous outcomes.

Between-person and within-person truth and bias forces. For Case 3, we considered a design in which each person has repeated measures. Here, we consider a complication for this design that was ignored earlier: It is possible to measure the truth and bias forces at two different levels. We discussed the effect of T_{Cij} on J_{Cij} (and the effect of B_{Cij} on J_{Cij}), which can be thought of as the effect being measured at the level of time. The second way in which the truth may influence judgment is by the effect of \bar{T}_i on J_{ij} , or at the level of the person. We also note that these forms of accuracy closely parallel the 1955 proposal of Cronbach to examine accuracy at different levels.

Use of discrepancy scores. It is very common in accuracy research to measure accuracy as the absolute difference between the truth and judgment, or $|J - T|$. Following Cronbach (1955) and other sources, we take a very dim view of using a discrepancy score as an outcome measure. Using discrepancy scores implicitly assumes that the truth force equals one, a very dubious assumption. We note that for all our examples, the truth force is always much less than one.

There may be situations in which one may wish to use a discrepancy variable as a *predictor* variable. Consider a test of the hypothesis that being able to read one's partner's emotions leads to fewer arguments. If we study a single interaction in which we measure accuracy and the arguments of that one interaction, it does make sense to use a discrepancy, because we can test whether there are fewer arguments if couples are accurate in *that* interaction. However, if we measure perceiver's accuracy and correlate it with reports of arguing in general (i.e., we are interested in accuracy as a process rather than being accurate at one time), a discrepancy score should not be used. Rather, we should estimate the truth force separately for each perceiver and use this estimate to predict number of arguments.

Design Issues

As we have described, the T&B model can be used when judgments are made of single targets (e.g., each perceiver rates his or her roommate) or when judgments of the same target are made at multiple time points or with the same stimuli judged by different perceivers. In relationship and group research, multiple people are making judgments of each other. Here, we consider alternative designs: the standard dyadic design and social relations model designs.

Standard dyadic design. Especially in the area of close relationships, the study of accuracy often uses the standard dyadic

design (Kenny, Kashy, & Cook, 2006): There are two people, and each member of the dyad serves as both a perceiver and a target. For instance, in heterosexual relationships, we can determine the extent to which husbands and wives are both accurate and biased in their judgments of their spouse. In fact, the roommate data sets that we analyzed in Cases 1, 2, and 3 employed the standard dyadic design. Case 4 is dyadic, but mothers serve as perceivers and adult children serve as targets.

With this design, we can elaborate the T&B model to consider the *interpersonal* processes as well. It is possible to address the following questions: If one member of a dyad is accurate, is the other member of the dyad also accurate? If one member is biased, is the other member biased? And if one member is accurate, is the other member biased? These between-person correlations, which we discussed for Case 3, provide insight into how bias and accuracy operate interpersonally, not just intrapersonally.

Also with dyadic designs, we examine how the truth and bias forces are moderated by perceiver- and target-level factors. Such a strategy would adopt an interdependence theory approach (Kelley & Thibaut, 1978) in which judgments are considered as a function not only of the perceivers' characteristics but also of the targets' characteristics. In the study of accuracy of judgments of personality, it might be of interest to examine moderators at the level of the perceiver and target, which could inform the process by which accuracy is due to perceivers being “good perceivers” (i.e., perceivers with larger truth forces) and targets being “good targets” (e.g., targets with larger truth forces; Funder, 1995). Complications arise in the statistical analysis of dyadic designs, many of which are discussed in Kenny et al. (2006), as well as the technical appendix for this paper.

Social relations model designs. A social relations model design is one in which each perceiver judges multiple targets and each target is judged by multiple perceivers. One social relations model design is the crossed design, and another is the round-robin design. We discuss each in turn.

Crossed design. We presented an example with repeated measures in Case 3. Perceivers judged their roommate on multiple days. There, we treated days as if they were nested within perceivers, a strategy commonly used for such data. Alternatively, we can treat repeated measures as crossed, not nested, within perceivers. Thinking of the design as crossed, we have a two-way analysis of variance: Perceivers \times Days.

A second example might help us better understand the nested versus crossed distinction. Imagine a study in which each participant guesses how happy 10 people are. If the participants all judge the same 10 people, the targets are said to be *crossed* with participant. If, however, each participant judges a different set of people (and these participants are not in turn judged by these people), the targets are said to be *nested* within participant.

In a crossed design with all perceivers judging the same targets, we can simultaneously estimate two different ways that the truth and bias forces vary: by perceiver and by target. In this way, we can evaluate whether some targets are easier to judge. Thus, with this design, we can in one study test Funder's realistic accuracy model hypothesis about individual differences in “good” judges and “good” targets. Moreover, we can measure the correlation of truth and bias forces across targets as well as perceivers. For the hurt feelings example, Case 3, we can consider day (or time point) as crossed within judge. When we conducted such an analysis, we

found no evidence for random variation in constant bias, truth force, and bias force by day ($ps > .05$), whereas all the variation between persons was statistically significant. Thus, we conclude that accuracy and bias forces do not vary by day, but they do vary by person.

Round-robin design. For this design, in groups of at least four persons, each perceiver judges every other member of the group. Thus, each member serves as both a perceiver and a target. Methods of examining accuracy with round-robin data have taken componential approaches, where dyadic judgments (i.e., Person A's judgment of Person B) are decomposed into different sources of variance. Kenny and Albright (1987) used a social relations model approach (Kenny & La Voie, 1984) to examine how accurate perceivers are in what they believe others think of them. Participants made judgments about how they believed others perceived them (i.e., metaperceptions) and then made judgments of everyone. Judgments and truth criterion were decomposed in three sources of variance: actor, partner, and relationship.

Imagine that John judges how much he believes that Sara likes him. His judgment can be decomposed into an *actor* effect, or the tendency for John to believe that others like him; a *partner* effect, or the tendency for people to believe that Sara likes them, and a *relationship* effect, or the tendency for John to uniquely believe that Sara likes him (adjusting for actor and partner variance). Similarly, criterion judgments (how much Sara actually likes John) can also be decomposed into these three sources of variance. Kenny and Albright (1987) correlated these sources of variance for judgments with those for the truth to estimate accuracy at different levels (for details, see their paper). One of these types of accuracy, *elevation* accuracy, measures mean differences between perceptions and the truth, which is analogous to the concept of directional bias in the T&B model.

Recently, Biesanz (2010) introduced the social accuracy model (SAM), which integrates a Cronbachian (1955) and a social relations model approach to accuracy. Unlike Kenny and Albright's (1987) approach and the T&B model, both of which examine accuracy separately for each trait, the SAM model focuses on a core question: How accurately do individuals perceive others and how accurately are they perceived by others, across traits? Like the T&B model, the SAM treats the judgment as the outcome measure and the "validation" as the truth measure, and thus perception is directly influenced by truth. As does the T&B model, the SAM includes a test of "mediation" whereby accuracy of targets is due to "generalized other" perceptions. For example, John may be accurate in his judgment that Sara likes ice cream because everyone likes ice cream.

Conclusion

We have presented a model for the general study of accuracy and bias in perception. Through four case examples, we have illustrated how the theoretical components of the model can be tested empirically. Our illustrations are by no means comprehensive of what types of questions the T&B model can be used to address. We have focused on the essential case analyses for illustrative analyses, but we emphasize that the model can be elaborated in a number of ways. For example, the T&B model can be elaborated to consider multiple moderating factors, interactions

between the truth and bias forces, and the relationship between the truth and bias forces between persons.

The T&B model complements rather than provides an alternative to the Gibsonian, Brunswikian, and SDT approaches. It is hoped that the model will be of value to a wide range of psychologists. To return to our opening example, our psychologist Kevin now has a theoretical and methodological model that will help guide his study of accuracy and bias, one that can be elaborated to examine more complex questions in his future research.

References

- Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*, 538–547. doi:10.1037/0022-3514.77.3.538
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 256–274. doi:10.1037/0033-2909.111.2.256
- Anderson, N. H. (1981). *Foundations of information integration theory*. New York NY: Academic Press.
- Arzamarski, R., Isenhower, R. W., Kay, B. A., Turvey, M. T., & Michaels, C. F. (2010). Effects of intention and learning on attention to information in dynamic touch. *Attention, Perception, & Psychophysics*, *72*, 721–735. doi:10.3758/APP.72.3.721
- Balcetis, E., & Dunning, D. (2006). See what you want to see: Motivational influences on visual perception. *Journal of Personality and Social Psychology*, *91*, 612–625. doi:10.1037/0022-3514.91.4.612
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Bem, D. J., & Allen, A. (1974). On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, *81*, 506–520. doi:10.1037/h0037130
- Berteletti, I., Lucangeli, D., Piazza, M., & Dehaene, S. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*, 545–551. doi:10.1037/a0017887
- Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1076–1096. doi:10.1037/0096-1523.25.4.1076
- Biesanz, J. C. (2010). *The social accuracy model of interpersonal perception: Assessing individual differences in perceptive and expressive accuracy*. Unpublished manuscript.
- Bond, C. F., Jr., & DePaulo, B. M. (2008). Individual differences in judging deception: Accuracy and bias. *Psychological Bulletin*, *134*, 477–492. doi:10.1037/0033-2909.134.4.477
- Borkenau, P., & Liebler, A. (1992). The cross-modal consistency of personality: Inferring strangers' traits from visual or acoustic information. *Journal of Research in Personality*, *26*, 183–204. doi:10.1016/0092-6566(92)90053-7
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217. doi:10.1037/h0047470
- Castellan, N. J., Jr. (1973). Comments on the "lens model" equation and the analysis of multiple-cue judgment tasks. *Psychometrika*, *38*, 87–100. doi:10.1007/BF02291177
- Cronbach, L. J. (1955). Processes affecting scores on "understanding others" and "assuming similarity." *Psychological Bulletin*, *52*, 177–193. doi:10.1037/h0044919
- Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353–379). Stanford, CA: Stanford University Press.
- DeCarlo, L. T. (1998). Signal detection theory and generalized linear

- models. *Psychological Methods*, 3, 186–205. doi:10.1037/1082-989X.3.2.186
- DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, 74, 63–79. doi:10.1037/0022-3514.74.1.63
- Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods*, 12, 1–22. doi:10.1037/1082-989X.12.1.1
- Fletcher, G. J. O., & Kerr, P. S. G. (2010). Through the eyes of love: Reality and illusion in intimate relationships. *Psychological Bulletin*, 136, 627–658. doi:10.1037/a0019792
- Fletcher, G. J. O., Simpson, J. A., & Boyes, A. D. (2006). Accuracy and bias in romantic relationships: An evolutionary and social psychological analysis. In S. Mark, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolutionary and social psychology* (pp. 189–209). Madison, CT: Psychosocial Press.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, 101, 75–90. doi:10.1037/0033-2909.101.1.75
- Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 652–670. doi:10.1037/0033-295X.102.4.652
- Gagné, F. M., & Lydon, J. E. (2004). Bias and accuracy in close relationships: An integrative review. *Personality and Social Psychology Review*, 8, 322–338. doi:10.1207/s15327957pspr0804_1
- Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston, MA: Houghton Mifflin.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98, 254–267. doi:10.1037/0033-295X.98.2.254
- Gilbert, D. T. (1997). Folk personology. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 89–151). New York, NY: McGraw-Hill.
- Gill, M. J., & Swann, W. B. (2004). On what it means to know someone: A matter of pragmatics. *Journal of Personality and Social Psychology*, 86, 405–418. doi:10.1037/0022-3514.86.3.405
- Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, 82, 379–398. doi:10.1037/0022-3514.82.3.379
- Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Oxford, England: Krieger.
- Hall, J. A., & Andrzejewski, S. A. (2008). Who draws accurate first impressions? Personal correlates of sensitivity to nonverbal cues. In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 87–105). New York, NY: Guilford Press.
- Hall, J. A., Andrzejewski, S. A., Murphy, N. A., Mast, M. S., & Feinstein, B. A. (2008). Accuracy of judging others' traits and states: Comparing mean levels across tests. *Journal of Research in Personality*, 42, 1476–1489. doi:10.1016/j.jrp.2008.06.013
- Harkness, A. R., DeBono, K. G., & Borgida, E. (1985). Person involvement and strategies for making contingency judgments: A stage in the dating game makes a difference. *Journal of Personality and Social Psychology*, 49, 22–32.
- Harvey, J. H., Town, J. P., & Yarkin, K. L. (1981). How fundamental is the “fundamental error of attribution”? *Journal of Personality and Social Psychology*, 40, 346–349. doi:10.1037/0022-3514.40.2.346
- Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, 78, 81–91. doi:10.1037/0022-3514.78.1.81
- Hastie, R., & Rasinski, K. A. (1988). The concept of accuracy in social judgment. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 193–208). New York, NY: Cambridge University Press.
- Heuer, K., Lange, W., Isaac, L., Rinck, M., & Becker, E. S. (2010). Morphed emotional faces: Emotion detection and misinterpretation in social anxiety. *Journal of Behavioral Therapy and Experimental Psychiatry*, 41, 418–425. doi:10.1016/j.jbtep.2010.04.005
- Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of projection. *Journal of Personality and Social Psychology*, 53, 221–234. doi:10.1037/0022-3514.53.2.221
- Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114, 733–758. doi:10.1037/0033-295X.114.3.733
- Hirsch, C. J., Hammond, K. R., & Hirsch, J. L. (1964). Some methodological considerations in multiple-probability studies. *Psychological Review*, 71, 42–60. doi:10.1037/h0041729
- Ickes, W. (1997). *Empathic accuracy*. New York, NY: Guilford Press.
- Ijzerman, H., & Semin, G. R. (2010). Temperature perceptions as ground for social proximity. *Journal of Experimental Social Psychology*, 46, 867–873.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 1–24.
- Jussim, L. (1991). Social perception and social reality: A reflection–construction model. *Psychological Review*, 98, 54–73. doi:10.1037/0033-295X.98.1.54
- Karelaia, N., & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134, 404–426. doi:10.1037/0033-2909.134.3.404
- Karney, B. R., & Coombs, R. H. (2000). Memory bias in long-term close relationships: Consistency or improvement? *Personality and Social Psychology Bulletin*, 26, 959–970. doi:10.1177/01461672002610006
- Kelley, H. H., & Thibaut, J. (1978). *Interpersonal relations: A theory of interdependence*. New York, NY: Wiley.
- 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. doi:10.1037/0022-3514.80.3.439
- Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102, 390–402. doi:10.1037/0033-2909.102.3.390
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York, NY: Guilford Press.
- Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142–182). Orlando, FL: Academic Press.
- Kenny, D. A., & West, T. V. (2010). Similarity and agreement in self and other perception: A meta-analysis. *Personality and Social Psychology Review*, 12, 196–213. doi:10.1177/1088868309353414
- Kenny, D. A., West, T. V., Cillessen, A. H. N., Coie, J. D., Dodge, K. A., Hubbard, J. A., & Schwartz, D. (2007). Accuracy in judgments of aggressiveness. *Personality and Social Psychology Bulletin*, 33, 1225–1236.
- Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, 27, 313–327. doi:10.1017/S0140525X04000081
- Kruglanski, A. W. (1989). The psychology of being “right”: The problem of accuracy in social perception and cognition. *Psychological Bulletin*, 106, 395–409. doi:10.1037/0033-2909.106.3.395
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498. doi:10.1037/0033-2909.108.3.480
- Luo, S., & Snider, A. G. (2009). Accuracy and biases in newlyweds'

- perceptions of each other: Not mutually exclusive but mutually beneficial. *Psychological Science*, *20*, 1332–1339. doi:10.1111/j.1467-9280.2009.02449.x
- MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding, and suppression effect. *Prevention Science*, *1*, 173–181. doi:10.1023/A:1026595011371
- Malouff, J., Schutte, N., & Priest, J. (2010). Publication rates of Australian academic psychologists. *Australian Psychologist*, *45*, 78–83.
- McArthur, L. Z., & Baron, R. M. (1983). Toward an ecological theory of social perception. *Psychological Review*, *90*, 215–238. doi:10.1037/0033-295X.90.3.215
- McKay, R., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior*, *31*, 309–319. doi:10.1016/j.evolhumbehav.2010.04.005
- Murray, S. L. (1999). The quest for conviction: Motivated cognition in romantic relationships. *Psychological Inquiry*, *10*, 23–34. doi:10.1207/s15327965pli1001_3
- Murray, S. L., Holmes, J. G., Bellavia, G., Griffin, D. W., & Dolderman, D. (2002). Kindred spirits? The benefits of egocentrism in close relationships. *Journal of Personality and Social Psychology*, *82*, 563–581. doi:10.1037/0022-3514.82.4.563
- Murray, S. L., Holmes, J. G., & Griffin, D. W. (1996). The benefits of positive illusions: Idealization and the construction of satisfaction in close relationships. *Journal of Personality and Social Psychology*, *70*, 79–98. doi:10.1037/0022-3514.70.1.79
- Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology*, *53*, 431–444. doi:10.1037/0022-3514.53.3.431
- Neyer, F. J., Banse, R., & Asendorpf, J. B. (1999). The role of projection and empathic accuracy in dyadic perception between older twins. *Journal of Social and Personal Relationships*, *16*, 419–442. doi:10.1177/0265407599164001
- Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, *14*, 106–112. doi:10.1111/1467-9280.t01-1-01427
- Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review*, *9*, 32–47. doi:10.1207/s15327957pspr0901_3
- Roeyers, H., & Demurie, E. (2010). How impaired is mind-reading in high-functioning adolescents and adults with autism? *European Journal of Developmental Psychology*, *7*, 123–134. doi:10.1080/17405620903425924
- Ross, L., Greene, D., & House, P. (1977). The false consensus phenomenon: An attributional bias in self-perception and social perception processes. *Journal of Experimental Social Psychology*, *13*, 279–301. doi:10.1016/0022-1031(77)90049-X
- Rusbult, C. E., Van Lange, P. A., Wildschut, T., Yovetich, N. A., & Verette, J. (2000). Perceived superiority in close relationships: Why it exists and persists. *Journal of Personality and Social Psychology*, *79*, 521–545. doi:10.1037/0022-3514.79.4.521
- Sasson, N. J., Pinkham, A. E., Richard, J., Huggett, P., Gur, R. E., & Gur, R. C. (2010). Controlling for response biases clarifies sex and age differences in facial affect recognition. *Journal of Nonverbal Behavior*, *34*, 207–221. doi:10.1007/s10919-010-0092-z
- Simpson, J. A., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology*, *69*, 629–641. doi:10.1037/0022-3514.69.4.629
- Swann, W. B. (1984). Quest for accuracy in person perception: A matter of pragmatics. *Psychological Review*, *91*, 457–477. doi:10.1037/0033-295X.91.4.457
- Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hirsch, Hammond, and Hirsch and by Hammond, Hirsch, and Todd. *Psychological Review*, *71*, 528–530. doi:10.1037/h0047061
- Turvey, M. T., Burton, G., Amazeen, E. L., Butwill, M., & Carello, C. (1998). Perceiving the width and height of a hand-held object by dynamic touch. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 35–48. doi:10.1037/0096-1523.24.1.35
- Warren, W. H., & Whang, S. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 371–383. doi:10.1037/0096-1523.13.3.371
- West, T. V., Pearson, A. R., Dovidio, J. F., Shelton, J. N., & Trail, T. E. (2009). Superordinate identity and intergroup roommate friendship development. *Journal of Experimental Social Psychology*, *45*, 1266–1272. doi:10.1016/j.jesp.2009.08.002
- Wilhelm, P., & Perrez, M. (2004). How is my partner feeling in different daily-life settings? Accuracy of spouses' judgments about their partner's feelings at work and at home. *Social Indicators Research*, *67*, 183–246. doi:10.1023/B:SOCI.0000007339.48649.20
- 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. doi:10.1207/s15327957pspr0103_2

Received August 21, 2010

Revision received January 6, 2011

Accepted January 18, 2011 ■