**Dieses Dokument ist eine Zweitveröffentlichung (Verlagsversion) /
This is a self-archiving document (published version):**

Daniel Leising, Anne-Marie B. Gallrein, Michael Dufner

**Judging the Behavior of People We Know: Objective Assessment,
Confirmation of Preexisting Views, or Both?**

Erstveröffentlichung in / First published in:

Personality and Social Psychology Bulletin. 2014, 40(2), S. 153 – 163 [Zugriff am: 31.07.2019].
SAGE journals. ISSN 1552-7433.

DOI: <https://doi.org/10.1177/0146167213507287>

Diese Version ist verfügbar / This version is available on:

<https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa2-353652>

„Dieser Beitrag ist mit Zustimmung des Rechteinhabers aufgrund einer (DFGgeförderten) Allianz- bzw. Nationallizenz frei zugänglich.“

This publication is openly accessible with the permission of the copyright owner. The permission is granted within a nationwide license, supported by the German Research Foundation (abbr. in German DFG).

www.nationallizenzen.de/

Judging the Behavior of People We Know: Objective Assessment, Confirmation of Preexisting Views, or Both?

Personality and Social
Psychology Bulletin
2014, Vol. 40(2) 153–163
© 2013 by the Society for Personality
and Social Psychology, Inc
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146167213507287
pspb.sagepub.com

Daniel Leising¹, Anne-Marie B. Gallrein¹,
and Michael Dufner²

Abstract

The present study investigates the relative extent to which judgments of people's behavior are influenced by "truth" (as measured by averaged observer-judgments) and by systematic bias (i.e., perceivers' preexisting views of target persons). Using data from online questionnaires and laboratory sessions ($N = 155$), we demonstrate that self- and peer-judgments of people's actual behavior in specific situations are somewhat accurate but are also affected by what perceivers thought of the targets *before* observing their behavior. The latter effect comprises a general evaluative component (generally positive or negative views of targets) and a content-specific component (views of targets in terms of specific characteristics, for example, "restrained"). We also found that friends, but not targets themselves, tend to judge targets' behaviors more positively than unacquainted observers do. The relevance of these findings for person perception in everyday life and in research contexts is discussed.

Keywords

behavior, judgment, accuracy, evaluation, acquaintance

Received October 30, 2012; revision accepted September 8, 2013

A real-life example: A board of assessors meets to see a candidate present his research, and then, based on their observations, decide whether the candidate should be granted a "certificate of eligibility for tenure" (*Habilitation*, specific to the academic system in some European countries). The meeting is public, thus many people who are not members of the board—including one of the candidate's close friends—are also part of the audience. After the talk, the candidate's friend approaches some colleagues and praises the candidate's performance: "Now, *this* is how it's done!" and so on. It is obvious that he actually believes the candidate performed really well and must have impressed everybody in the room quite a bit. Praising the candidate's performance for other reasons (e.g., to influence the decision that is about to be made) would make no sense for him because the people he talks to have nothing to do with the formal evaluation of the candidate. Notably, *nobody* agrees with him. Most people just nod politely to what he says and remain silent. Obviously, the other people in the room, who just saw the candidate for the very first time, agree that the candidate's performance was actually quite horrible. How may such discrepant perceptions of the same person's behavior be explained? May the personal relationship between the candidate and the friend have something to do with it?

According to West and Kenny's (2011) Truth and Bias Model, the variables affecting all sorts of judgments, including judgments of persons, can be categorized as either "truth" or "bias" variables. Truth variables represent what is "real" or "objectively correct" about a judgment, whereas bias variables represent any systematic influences apart from that. The extent to which a truth or bias variable affects judgments is called the "force" of that variable. In the present study, we investigate the relative extents to which self- and peer-judgments of people's behavior in specific situations are affected by how the target actually behaved (truth) and by the preexisting views that the perceivers had of the targets before observing their behavior (bias).

Truth in Judgments of Behavior

There are at least two distinct ways in which behavior may be described. On the one hand, one may focus on qualities of

¹Technical University Dresden, Germany

²Humboldt-University Berlin, Germany

Corresponding Author:

Daniel Leising, Department of Psychology, Technical University Dresden,
01062 Dresden, Germany.

Email: Daniel.Leising@tu-dresden.de

behavior that are measurable with great exactness (e.g., whether person A decreases or increases her physical distance from person B). Such qualities of behavior may often be assessed without relying much on any subjective human judgments (e.g., by means of a ruler). The downside of such assessments is that their meaning is often psychologically ambiguous, and thus needs to be interpreted a posteriori by the researcher. In everyday life, on the other hand, people tend to describe their own and others' behaviors by *meaning-laden* terms like, for example, "friendly," "hasty," "hostile," or "relaxed." That is, they focus on the psychological relevance of people's behavior rather than on its physical qualities. This latter kind of behavioral judgment is also the one that we are dealing with in the present article. However, determining whether a perceiver's judgment of a target's behavior by means of a meaning-laden term reflects the "truth" is less straightforward than in the first case.

It may be argued that the best way to assess whether some behavior is "actually" friendly, hasty, hostile, or relaxed is the *average* view that many neutral observers have of that behavior (cf. Kenny, 2004). Averaging different perceivers' judgments of the same behavior reduces the impact of the individual perceivers' idiosyncratic term uses and focuses on the *common* element instead. The existence of the latter constitutes the very reason why person-descriptive language may be effectively used at all. Moreover, averaging judgments of behavior will yield an impression that is more representative of how a behavior will be perceived by the average other person. This in turn is important because how others interpret our behaviors will strongly determine the *consequences* of those behaviors (Leising & Müller-Plath, 2009), so the "social reality" that only exists in other people's minds at first (e.g., "the remark that Pete just made was really offensive") may eventually have very real effects on the target person (e.g., Pete may be asked to leave). In accordance with this view, we define *accuracy* as the agreement between a given perceiver's judgment of a given behavior and the average judgment of the same behavior by several neutral observers. In terms of West and Kenny's (2011) Truth and Bias model, the truth variable in our study is the average behavior judgment by several neutral observers, and the extent to which this variable predicts judgments of the same behavior by any other perceivers (e.g., targets or informants, see below) is the "force" of the truth variable.

Using such a definition, research has shown that individuals are in fact able to judge their own behavior within specific situations with some degree of accuracy (e.g., Leising, 2011; Sadler & Woody, 2003). In these studies, participants were asked to judge their own behavior during laboratory interactions, and these self-ratings did predict judgments of the same interactions by unacquainted observers. Therefore, we expect that in the present study, self-judgments of behavior will converge at least somewhat with judgments by neutral observers. We also address the

previously neglected question of whether the same is true for behavioral judgments that are provided by close acquaintances of the targets.

Preexisting Views as Biases in Judgments of Behavior

The systematic biases that we focus on in the present article concern perceivers' preexisting views of the target persons whose behavior they judge. In terms of the Truth and Bias model, such preexisting views are bias variables that may affect behavioral judgments independent of the truth. There are at least two conceptually distinct ways in which such views may influence subsequent judgments of behavior, one of them being *globally evaluative* and the other being *content-specific*. In the first case, a perceiver may simply have an overall positive (or negative) attitude toward a target, and thus tend to use all sorts of positive (or negative) terms to describe that target's behavior in a given situation, irrespective of how the target "actually" behaved. In the second case, a perceiver may have a content-specific view of a target (e.g., as being arrogant, witty, or aloof), and thus tend to judge that target's behavior in a given situation accordingly, irrespective of how the target "actually" behaved.

Regarding global evaluation, there is a considerable variation in how positively or negatively people tend to view themselves and others (cf. Bono & Judge, 2003; Furr & Funder, 1998; Kim, Schimmack, & Oishi, 2012; Leising, Erbs, & Fritz, 2010; Nisbett & Wilson, 1977; Saucier, 1994; Srivastava, Guglielmo, & Beer, 2010; Wood, Harms, & Vazire, 2010), and such different attitudes strongly predict whether perceivers will use positive or negative terms in characterizing targets (Leising et al., 2010; Leising, Ostrovski, & Zimmermann, 2013). Sometimes, a perceiver's characterization of a target's behavior in a given situation may even depend more on whether the perceiver is fond of the target or not than on how the target "actually" behaved. We are talking about *partisanship* here. Research in social psychology has demonstrated that perceptions of behavior may be strongly affected by a perceiver's loyalty or allegiance regarding the target person: Studies of presidential debates in the United States showed that this factor predicted which candidate viewers would see as the winner (cf. Munro et al., 2002). In the present study, we expect that a perceiver who already holds a generalized positive image of a target person will see the subsequent behaviors of that target person in a more positive light, whereas a perceiver with a more negative attitude toward a target will do the opposite. We apply this logic to self- and peer-judgments of behavior alike.

Apart from global evaluation, a perceiver may think that a person's behavior in a given situation was highly (for example) sophisticated, friendly, or arrogant because he or she tended to think of the target as being sophisticated, friendly, or arrogant even before actually observing the behavior in question. Note that here it is the specific *content* of the word,

not its evaluative connotation, that matters. Let us consider, for example, a perceiver who is supposed to judge a known target's "quickness" at a problem-solving task. Regardless of how positively or negatively the perceiver thinks of the target in general, if she is convinced that the target is usually "quick" at accomplishing things, she may tend to judge the quickness of the target's problem solving in the situation accordingly, regardless of how quick the target actually was.

There is some evidence supporting the notion that people tend to characterize the behaviors of targets in line with preexisting images, and that to some extent, this happens independent of the targets' actual behaviors. For example, in the study by Sadler and Woody (2003), the participants tended to rate their own behavior during an interpersonal interaction in accordance with their general self-views, even if their "actual" behavior during the interaction (as consensually judged by several observers) was controlled for. In other words, the participants overestimated the extent to which their behavior in the situation concurred with what they generally thought of themselves. Sadler and Woody (2003) called this phenomenon a "consistency bias" (cf. Leising, 2011).

In the present article, we will use the term *preexisting views* (PEV) to denote the images of the targets that already existed in the perceivers' minds before they engage in the task of judging the targets' behavior in specific situations. An important limitation of the aforementioned studies is that they did not distinguish between preexisting views that are globally evaluative and preexisting views that are content-specific. Thus, for example, a given perceiver's judging a target's personality and behavior as "witty" may have been due to the specific content of the term or simply due to the fact that the term has a positive evaluative connotation. In the present study, we disentangle these two possibilities from one another by separately investigating the influences of preexisting views that are globally evaluative (PEV-GE) and preexisting views that are content-specific (PEV-CS).

Moreover, we also disentangle two components of behavioral judgments that were not disentangled in previous studies of the "consistency bias": Preexisting views of targets may affect subsequent judgments of those targets' behaviors because (a) perceivers judge *particular* targets in certain ways and/or because (b) perceivers tend to judge *all* targets in certain ways. In Kenny's (1994) terminology, the former would be denoted as "relationship effects," whereas the latter would be denoted as "perceiver effects." We investigate how strongly both components contribute to the consistency between preexisting views and subsequent judgments of behavior.

Furthermore, in the studies described above the interval between the global self-assessment of personality and the momentary self-assessment of situated behavior was very short (less than an hour). It is thus conceivable that the influence of the perceivers' preexisting views was overestimated because the participants' general images of themselves were made salient shortly before they were asked to judge their

own behavior (Leising, 2011), or the personality assessment took place shortly after the participants had judged their own behavior (Sadler & Woody, 2003). In the present study, we use much longer intervals between the two kinds of assessment in order to rule out this alternative explanation. Finally, our study extends the previous work by testing whether the preexisting views affect not only self- but also peer-judgments of behavior.

Possible Mechanisms

Even though we do not directly address psychological mechanisms in the present article, we think it should still be asked *why* one may expect effects of perceivers' preexisting views of targets on those perceivers' subsequent judgments of the same targets' behaviors. Would it not be more reasonable for a perceiver to form as accurate an impression of the target's behavior as possible and not rely on any other source of information than "what is actually there" (truth)? A large number of studies have demonstrated that people tend to interpret new information in accordance with their preexisting beliefs or expectations (see Nickerson, 1998, for a review). This phenomenon has often been attributed to *motivational* factors such as a basic need to maintain cognitive congruence (Festinger, 1957; Kunda, 1990; Nickerson, 1998). It may be argued that such congruence is needed because only congruent views permit predictions of future events. With regard to behavior perception, this would imply that once a perceiver has formed a certain view of a target person, he or she may be inclined to actively seek or prioritize novel information that concurs with this view. In fact, research has shown that once individuals have developed some belief about another person (e.g. "the person is friendly"), they selectively seek behavioral evidence that confirms rather than falsifies this belief (Lord, Lepper, & Preston, 1984; Snyder & Campbell, 1980; Snyder & Swann, 1978).

Another possible explanation is more *informational* in nature. For a perceiver who has access to several sources of information about a target (e.g., observations of the target's behavior in several situations), it may be wise to pool that information in order to form the most accurate impression of the target possible. The upside of doing so would be that the resulting image of the person would likely become more representative of how the person is *on average*, possibly permitting more accurate predictions of the target's behavior in the long run. Thus, if the target's behavior in a new situation deviates from what the perceiver has previously learned about the person, the new information may be partially discounted as an "exception," and the interpretation of the new behavior may be "colored" by the perceiver's previous experiences with the target. The downside of doing so, however, would be that the perceiver's ability to judge the target's behavior in *just one* situation objectively may at least partly be compromised. The present study investigates the extent to which this is the case.

Overview of the Present Study

Given the ubiquity and importance of self- and peer-judgments of behavior in everyday life as well as in psychological research, it seems worthwhile to examine the factors that contribute to the formation of such judgments. In the present study, we use a design that allows us to simultaneously investigate the influence of truth and bias (i.e., perceivers' preexisting view of targets) on self- and peer-judgments of behavior. In order to maximize external validity (cf. Baumeister, Vohs, & Funder, 2007), we do not use written vignettes of people's behavior (which was the predominant method in previous studies addressing similar issues) but rather investigate judgments of people's *actual behaviors*. Specifically, we invite target persons and their peers to the laboratory and then videotape the targets' behaviors in a variety of standardized situations. Afterward, the targets and their peers judge the targets' behaviors. In order to obtain an accuracy criterion, the same behaviors are also judged by unacquainted observers. The general images that the targets and their peers have of the targets' personalities are assessed by means of online questionnaires several days prior to the laboratory assessments. The design enables us to compare the relative influence of the truth (= averaged observer-judgments), as compared to PEV-GE and PEV-CS, on the participants' judgments of their own and their peers' behaviors.

Method

Sample

The main sample comprised two groups of participants, the target persons ($N = 155$) and the informants ($N = 155$), who came to the lab together in dyads (see below). Each participant (i.e., target or informant) received 15 Euros as a reward for participating. One hundred and fifty target persons completed the online questionnaire. Of these, 86 were female, 63 were male, and 1 did not report sex. The mean age of the targets was 23.2 years ($SD = 4.06$). Of the informants, 153 completed the online questionnaire. Of these, 93 were female, 59 were male, and 1 did not report sex. The mean age of the informants was 23.8 years ($SD = 4.44$). The majority ($n = 85$) of the informants were friends of the targets'. Another large proportion of the informants were the targets' romantic partners ($n = 53$). Other kinds of relationships between informants and partners (e.g., colleagues, siblings) were less frequent ($n < 10$). The informants reported knowing the targets quite well ($M = 4.17$, $SD = 0.79$, on a scale ranging from 1 ["not at all"] to 5 ["very well"]) and liking them very much ($M = 4.70$, $SD = 0.51$, on a scale ranging from 1 ["not at all"] to 5 ["very much"]). The values for the targets' knowing and liking of the informants were virtually identical (Knowing: $M = 4.21$, $SD = 0.73$; Liking: $M = 4.67$, $SD = 0.55$), which is unsurprising given that the roles of target and informant were assigned at random (see below).

Procedure

Recruitment. We advertised the study on a campus website and in local community centers. The announcement of the study contained an e-mail address to which participants were asked to write in order to register for the study. When prospective participants contacted the research team, they received an e-mail containing some broad information regarding the aims and procedures of the study. The e-mail also asked the participants to recruit a second person who knew them reasonably well and would also be willing to take part in the study. When a prospective participant contacted the research team again and named a second person who had agreed to participate, we randomly determined which of the two persons would serve as the *target* and which would serve as the *informant*. In the following, we will call these target-informant pairs *dyads*.

Personality questionnaires. Both the target and the informant of a dyad received an e-mail containing a link to an online questionnaire. They were asked to log in and describe the target person's personality by means of an adjective list (see below). We emphasized that the two persons should complete the questionnaire independently and that they should not communicate with each other about their ratings. The online questionnaire also contained a number of questions regarding the two participants' personal backgrounds (e.g., socioeconomic status) and their relationship with each other (e.g., knowing and liking).

Laboratory session. A few days later, the target and the informant were invited to come to the lab together for a series of observational assessments. These assessments all took place on the same day and comprised three stages. In the first stage, the target person engaged in four tasks, which were presented in random order. *Task 1* consisted of answering a number of questions pertaining to general knowledge (e.g., "How high is Mount Everest?" "How many people live in the Tokyo area?" "Who wrote the opera 'Fidelio'?"). *Task 2* consisted of a role-play in which the target person was to call a "neighbor" (actually played by a confederate) and to demand that the neighbor turns down the volume on her stereo (cf. Borke-nau, Mauer, Riemann, Spinath, & Angleitner, 2004). In *Task 3*, the target person was asked to spontaneously tell a brief story incorporating the terms "Corkscrew," "Holiday," "Catastrophe," and "Glove Box." In *Task 4*, the target person was asked to (a) sing a song of his or her own choice, (b) tell a joke of his or her own choice, and (c) pantomime the term *party* (which in German does not also mean "political party" but only "festivity"). The four tasks were selected to enable an assessment of diverse qualities of behavior. Completing a task usually took a target less than 90 s. The targets' behavior while engaging in the tasks was videotaped. While the target of a given dyad completed the four tasks, the respective informant was seated in another room and asked to wait.

In the second stage of the laboratory session, the target and the informant each rated four video-clips, by means of the same adjective list that was also used for the personality assessments (see below). These video-ratings took place in separate rooms. One member of the dyad (the target or the informant) rated the video-clips that had just been produced—the ones showing the current target engage in the four tasks. The tasks were presented in the same order in which the target had completed them. Meanwhile, the other member of the dyad watched four video-clips showing the *target of the previous dyad* (i.e., an unknown person) engage in the same tasks. Both the target and the informant of the current dyad were asked to watch one of the respective video-clips that had been assigned to them, then rate the respective target's behavior by means of the adjectives, then watch the next video-clip, and so on. Whether the informant watched the current target and the current target watched the previous target, or vice versa, was also determined at random.

In the third stage of the laboratory session, the previous rating assignments were reversed: If the informant had just rated the current target, and the current target had just rated the previous target, then the informant was now supposed to rate the previous target, whereas the current target was supposed to rate himself or herself. If the rating assignments in the second stage had been the other way round, then the assignments in the third stage were reversed accordingly. This way, we let both the informants and the targets judge (a) targets they knew and (b) targets they did not know.

External observers' ratings. Each of the 620 video-clips (= 155 targets $\times$ 4 tasks) was also independently rated by at least three observers who were unacquainted with the targets (the maximum number of observers per video was six). The overall pool of observers comprised 18 persons from which subgroups were selected randomly to judge the individual video-clips. The ratings were balanced such that each observer would get to see all targets but only see each target in one randomly chosen situation. Hence, the observer-ratings always reflected judgments by persons who saw the respective target in just one situation. The observers did not receive any training (apart from general technical instructions on how to watch the videos and provide their ratings) in order to preserve the representativeness of their intuitive ways of applying the adjectives to the target persons' behaviors.

Measure

The adjective list that we employed for the ratings of the targets' personalities and behavior in the lab was devised by Borkenau and Ostendorf (1998) and has been used in several person perception studies (e.g., Leising et al., 2010; Leising et al., 2013). It consists of 30 terms that assess the Big Five personality factors by six items each. Three terms for each factor have a positive valence and three have a negative valence. We let the participants use the same terms for assessing the targets'

more stable behavioral dispositions (in the online personality questionnaire) and their more transient situation-bound behaviors (in the lab). This conforms to the *personality states* approach, which posits that both traits and states may be assessed by means of the same vocabulary (Fleeson, 2001; Bleidorn, 2009). For example, in the online questionnaire, the targets and their informants rated how "shy" or "egoistic" they considered the targets to be in general, whereas in the lab session, they rated how "shy" or "egoistic" they considered the targets' behavior during the tasks to be. However, because factor structures of measures may differ considerably depending on whether a measure is used for assessing states versus traits (e.g., Leising & Bleidorn, 2011), we only report analyses at the level of individual items in the following. This is also the level of abstraction at which everyday communication about individual differences in personality and behavior takes place. The intervals between the completion of the online questionnaire and the laboratory assessments had an average length of $M = 4.08$ days ($SD = 3.24$) for the targets and of $M = 4.43$ days ($SD = 4.37$) for the informants.

Each of the 30 adjectives was also rated in terms of how positive or negative an impression of a target person it evokes (i.e., social desirability). For this, we used the ratings that were provided by 24 student raters as part of the study by Leising et al. (2010). The scale ranged from 1 (*very negative*) to 7 (*very positive*). Interrater reliability for these ratings was $ICC(3, 24) = .98$. The desirability ratings were needed to compute global evaluation indices by correlating a perceiver's ratings of a target on the 30 items with the profile of item desirabilities (Edwards, 1953). The resulting correlations reflect the extent to which the perceiver describes the target in a more positive or negative manner overall.

Results

Agreement Between the Observers

We used averaged observer-ratings as a measure of the targets' "true" behaviors during the laboratory tasks. The average interrater agreement, $ICC(1, 3)$, between three observers in judging the targets' behavior by means of a single item was .45 ($SD = 0.15$) for the story task, .47 ($SD = 0.19$) for the pantomime task, .44 ($SD = 0.18$) for the role-play task, and .49 ($SD = 0.11$) for the knowledge task. These levels of interrater agreement are comparable to the levels of agreement that were obtained in previous studies (e.g., Borkenau et al., 2004) and can be regarded as quite satisfactory. Note that they reflect *lower-bound* estimates for interrater agreement because the performances of most targets in most situations were judged by more than three observers.

Correlations Among the Predictors

In West and Kenny's (2011) Truth and Bias Model, the truth and bias variables may or may not be correlated with one

Table 1. Prediction of the Targets' and the Informants' Behavior Ratings From the Truth and the Two Bias Variables.

Criterion	Predictor	Betas	Standardized Betas	Without perceiver effects
			Raw	
Behavior rating by target	Truth	.22 (.16, .28)	.14 (.10, .18)	.17 (.13, .21)
	PEV-CS bias	.18 (.14, .22)	.17 (.13, .21)	.13 (.10, .15)
	PEV-GE bias	.28 (.21, .35)	.09 (.06, .11)	.03 (.00, .05)
Behavior rating by informant	Truth	.31 (.23, .38)	.18 (.13, .23)	.18 (.15, .22)
	PEV-CS bias	.21 (.17, .25)	.19 (.15, .23)	.08 (.05, .11)
	PEV-GE bias	.22 (.13, .30)	.08 (.04, .10)	.04 (.02, .06)

Note. Coefficients were first computed for each task and item separately and then averaged (99% bootstrap confidence intervals are in parentheses). Truth = averaged observer-ratings of the targets' behavior, PEV-CS = Perceivers' preexisting views of targets (content-specific) as measured by the perceivers' response to the same item in the personality questionnaire. PEV-GE = Perceivers' preexisting views of targets (global evaluation) as measured by the correlation between the perceiver's description of the target in the personality questionnaire and the social desirabilities of the 30 items. PEV-GE was multiplied with minus 1 when predicting behavior ratings on items with a negative valence (average desirability rating < 4).

another. In our sample, all correlations between the three predictors (averaged across items) were statistically significant as indicated by 99% bootstrap confidence intervals that did not include zero. The targets' self-ratings on an individual questionnaire item (PEV-CS) predicted the observers' ratings of the targets' behavior on the same item (truth) at $r = .14$ on average. The respective correlation for the informants' ratings of the targets' personalities was $r = .11$. Thus, the targets' and the informants' item-specific views of the targets' personalities had some validity, in that they could predict independent ratings of the targets' behaviors in the lab later on.

Before computing the average correlation of PEV-GE with PEV-CS and with the observer-ratings of the targets' behavior, PEV-GE had to be inverted (i.e., multiplied with -1) in all cases where it was correlated with an item that had a negative valence (i.e., an average desirability rating < 4). Otherwise, the correlations of PEV-GE with positive and negative items might have canceled out when averaging across items. After implementing this rule, the average correlation between PEV-GE and PEV-CS was .28 for the targets' self-ratings and .34 for the informant-ratings, suggesting that targets who received more positive overall personality ratings were also judged more positively on individual items. The averaged observer-ratings of the targets' behavior (truth) correlated at .06 (on average) with PEV-GE in the targets' self-ratings of their personalities and also at .06 with PEV-GE in the respective informant-ratings. Thus, targets who had received more positive overall personality ratings were also judged in a slightly more positive manner by the average observer, on individual items.

Relative Influence of Truth and Bias on Judgments of the Behavior of Known Targets

Using multiple regressions, we predicted the participants' (i.e., the targets' and the informants') judgments of the targets' behavior in the lab from (a) averaged observer-judgments of the same behavior (truth), (b) the same perceivers'

preexisting views of the targets in terms of particular items (PEV-CS), and (c) the same perceivers' preexisting global evaluations of the targets (PEV-GE). For these analyses, we used the targets' and the informants' ratings of persons they *knew* (i.e., the current targets) because without prior acquaintance between target and perceiver, behavior judgments could not be affected by the latter two factors. Note that by using multiple regressions, we determined the *independent* contributions of each predictor, controlling for the other two predictors.

Table 1 displays the average Betas that emerged for the three predictors, along with 99% confidence intervals that were determined by bootstrapping with the 30 items as cases. To arrive at these correlations, we first ran multiple regressions for each item and task separately (= 120 regressions), then averaged the resulting coefficients across tasks for each item (distinguishing between tasks did not lead to any different conclusions), and then averaged across the 30 items. In computing average Betas for PEV-GE, we again reversed the direction of the predictor for all items with a negative valence. Otherwise, positive and negative associations might have canceled out. In fact, when *not* inverting PEV-GE in predictions of negative items, the average standardized Betas for the individual items correlated at $r(28) = .94$ (targets) and $r(28) = .87$ (informants) with the items' social desirability ratings. Thus, the more an item entailed a positive or negative evaluation, the better the perceivers' ratings of the targets' behavior by means of that item could be predicted from the same perceivers' preexisting global evaluations of the targets (PEV-GE).

In our discussion of the average contributions of the three predictors, we will concentrate on the standardized Betas that are displayed in the second and third data column of Table 1. As can be seen from the entries in the second data column, all three predictors made significant contributions in predicting the targets' and the informants' judgments of the targets' behavior: The targets' actual behavior (truth)—as judged by unacquainted observers—and the perceivers'

content-specific preexisting views of the targets (PEV-CS) both made significant contributions of approximately the same size. In addition, the perceivers' globally evaluative views of the targets (PEV-GE) made another significant, but somewhat smaller, contribution of their own. The relative strengths of the three effects were approximately the same for self- and informant-ratings of behavior.

We also controlled for perceiver effects in order to determine how much of the influence of the two bias variables was due to general perceptual tendencies as opposed to the perceivers' views of the *particular* targets at hand. In the present study, a perceiver's perceiver-effect was obtained by averaging his or her judgments of the behavior of the known target (i.e., himself or herself or the other person in his or her dyad) and the unknown target (i.e., the target of the previous dyad). Perceiver-effects were calculated separately for each of the 30 items and subtracted from the participants' ratings of the targets' behavior (i.e., the DV) before repeating the item-wise multiple regressions described above. The results are displayed in the third data column of Table 1: Controlling for perceiver effects did not change the effect size for the truth variable (i.e., observer-ratings of the targets' behavior). It did, however, significantly reduce the influence of the perceivers' global evaluations of the targets (PEV-GE) on both self- and informant-ratings of behavior. Thus, a substantial proportion of the above-reported effect of this predictor *was* due to perceiver effects. For self-ratings, the respective confidence interval almost included zero (the lower confidence limit was minimally larger than zero, which is obscured in the table because of rounding). Finally, controlling for perceiver effects also lowered the impact of PEV-CS on behavior judgments, with the decrease being considerably stronger for the informants' (compared to the targets') judgments of the targets' behavior.

Loss of Validity Due to Target-Perceiver Relationship?

Given that the perceiver's preexisting views of the targets were found to affect their subsequent judgments of the targets' behavior, we investigated whether the accuracy of judgments of known targets was lower than the accuracy of judgments of unknown targets. To this aim, we compared the accuracy of four kinds of judgments with each other: (a) The targets' judgments of their own behavior, (b) the informants' judgments of the behavior of "their" targets, (c) the targets' judgments of the behavior of the previous targets, and (d) the informants' judgments of the behavior of the previous targets. Thus, we compared the accuracy of perceivers who judged targets they already knew (a, b) with the accuracy of perceivers who judged targets they did not know (c, d). In all four cases, accuracy was defined as the agreement between the respective kind of judgment (a, b, c, d) and the average observer-judgment of the same behavior. If acquaintance

with a target impairs validity, then accuracy should be lower for "a" and "b" as compared to "c" and "d."

First, we computed the correlations between each of the four perspectives and the averaged observer-ratings separately for each task and item. Then we averaged across tasks. Across the 30 items, the mean correlation was $r = .18$ (99% bootstrap CI = [.13, .22]) for the current targets' ratings of their own behavior (a), $r = .22$ (CI = [.17, .27]) for the current informants' ratings of the current targets' behavior (b), $r = .26$ (CI = [.22, .30]) for the current targets' ratings of the previous targets' behavior (c), and $r = .21$ (CI = [.16, .26]) for the current informants' ratings of the previous targets' behavior (d). Confidence intervals revealed that only the greatest difference (.18 vs. .26) was significant, suggesting that the accuracy of the targets' judgments was greater when they judged someone else's behavior as compared to when they judged their own behavior.

Global Evaluation Across Different Assessments

Finally, we compared how positively or negatively participants' personalities and behaviors were judged by the different types of perceivers. In each case, a perceiver's overall evaluation of a target was measured in terms of the within-person correlation between the perceiver's description of the target across the 30 adjectives and the social desirability ratings for these adjectives. Thus, a positive correlation was indicative of positive global evaluation. Note that global evaluation coefficients were computed for the participants' judgments of the targets' personalities by means of the online questionnaire (= PEV-GE) as well as for their judgments of the targets' behavior in the lab. Each target's behavior in the lab was rated by one previously acquainted target (i.e., the target's self-rating) and one previously acquainted informant, as well as by one unacquainted target and one unacquainted informant. In addition, there were three to six observer-ratings of each target's behavior. In order to ensure comparability between perspectives, we first computed a global evaluation index for each individual observer and then averaged across observers. For behavior ratings, evaluation indices were first computed separately for each task and then averaged across tasks.

The average global evaluation indices were $r = .58$ (range = $-.14$ to $.92$) for the targets' self-descriptions and $r = .68$ (range = $.05$ to $.93$) for the informants' descriptions of "their" targets' personalities in the online questionnaire. The targets' average evaluation was $r = .43$ (range = $-.62$ to $.88$) for their own behavior in the lab and $r = .45$ (range = $-.36$ to $.89$) for other targets' behavior. The informants' average evaluation was $r = .54$ (range = $-.48$ to $.93$) for their "own" targets' behavior in the lab and $r = .40$ (range = $-.41$ to $.88$) for other targets' behavior. Finally, the average evaluation of the targets' behavior in the lab by a single neutral observer was $r = .41$ (range = $-.31$ to $.77$).

We first compared the evaluation coefficients for the five different kinds of judgments of the targets' behavior in the lab with one another, using one-way repeated measures ANOVA. The comparison was statistically significant, $F(4, 147) = 12.274$, $p < .001$. Pairwise comparisons with Bonferroni corrections for multiple testing ($\alpha = .01$) suggested that the informants' judgments of their own targets' behavior were significantly more positive than all other behavior judgments but that none of the other behavior judgments differed from each other in terms of global evaluation. In a second analysis, we compared the evaluation coefficients for the targets' versus the informants' judgments (Factor 1) and for questionnaire versus behavior ratings (Factor 2) with one another, using two-way repeated measures ANOVA. Note that in this analysis, we only used judgments within the same dyads, that is, judgments of known targets. The analysis yielded significant effects for both factors, target-informant, $F(1, 146) = 29.009$, $p < .001$; behavior-personality, $F(1, 146) = 94.784$, $p < .001$; but no significant interaction effect, $F(1, 146) = 0.215$, $p = .643$. Hence, informants viewed targets more positively than targets viewed themselves, and for both targets and observers, ratings of overall personality were more positive than ratings of behavior in the lab.

Discussion

Truth and Bias in Judgments of Behavior

The current study aimed to investigate the relative extents to which self- and informant-ratings of targets' behavior in specific situations reflect "the truth," as opposed to two different kinds of bias (i.e., PEV-GE and PEV-CS). Taken together, we found that the truth and both bias variables had significant and independent effects on self- and informant-ratings of behavior. To some extent, such ratings reflect (a) the views that external observers would have of the same behaviors (truth), (b) the generally positive or negative views that the perceivers had of the targets *before* observing their behaviors (PEV-GE), and (c) the views that the perceivers had of the targets before observing their behaviors in terms of specific judgment dimensions (e.g., "restrained"; PEV-CS). The "forces" of the truth and PEV-CS variables were about equally strong and stronger than the force of the PEV-GE variable.

With regard to truth, we found that self- and informant-ratings of behavior moderately agreed with average judgments of the same behavior by three or more observers. This largely confirms the findings of previous studies (Leising, 2011; Sadler & Woody, 2003) and implies that momentary self- and informant-ratings of behavior do in fact concur to some extent with how the same behavior would be interpreted by persons who see the target person for the first time. If one accepts averaged behavior ratings by several neutral observers as a valid measure of the behavior's "true" meaning, then this finding supports the notion that self- and

informant-ratings of a person's behavior are at least partly accurate or valid.

Apart from that, however, the present article also demonstrates that people's judgments of the behavior of targets they know are affected by two different kinds of bias: evaluative and content-specific preexisting views of a person. The former bias closely resembles the effect of partisanship on judgments of candidates' performances in presidential debates (cf. Munro et al., 2002). However, the findings of the present study pertain more directly to everyday interpersonal relationships: Preexisting positive or negative attitudes do not only affect people's judgments of the debate performances of presidential candidates they (dis-)like, they also affect people's judgments of their own behavior, and of the behavior of their close acquaintances, across diverse situations.

Notably, the strength of the PEV-GE bias is strongly moderated by the evaluative connotation of the item that is used to describe the behavior. The more evaluative the item is, the larger the (positive or negative) impact of the perceiver's global evaluation of the target will be (cf. Leising & Borkenau, 2011). The present study is already the third to empirically corroborate this interaction between perceivers' global evaluations of targets (sometimes assessed as "liking") and the evaluativeness of person descriptors (cf. Leising et al., 2010; Leising et al., 2013). In fact, the correlations that we found between the Betas for PEV-GE and the item desirability ratings were *so* high ($r > .86$) that it may be reasonable to interpret the latter as directly reflecting the extent to which items will respond to a perceiver's more positive or negative attitude toward a target. This renders it likely that controlling for item desirability may be an effective way of addressing *socially (un-) desirable responding* (cf. Paulhus, 2002), that is, individual differences between perceivers in terms of their judging (all or particular) targets too positively or too negatively. Future research needs to address this highly intriguing possibility.

The other kind of bias that independently affects behavior judgments is content-specific: If a perceiver has come to generally attribute some level of a particular characteristic to a target person, then his or her judgments of that target's subsequent behavior in specific situations will partly reflect this general view, even when controlling for the target's "actual" behavior in the situation. This finding clearly stands in line with the previous research (cf. Leising, 2011; Sadler & Woody, 2003). Going beyond previous findings, however, the present study demonstrates that this bias effect still holds if one controls for general evaluation. It is also the first study to show that informant-ratings of behavior are just as susceptible to this bias as are self-ratings. Moreover, our study demonstrates that a substantial proportion of the effect of PEV-CS is target-specific and cannot merely be explained in terms of perceivers' general judgment tendencies (cf. Kenny, 1994). Finally, the present study demonstrates that the effect occurs even if the interval between the assessment of the perceiver's general image of the target and same perceiver's judgment of

the same target's behavior in a particular situation is several days long (about 4 days on average). Therefore, it appears highly unlikely that the effect is merely rooted in the researchers' making the perceivers' general images of the targets salient shortly before the behavioral assessment takes place. An explanation in terms of motivational or informational mechanisms (see above) seems more viable. Future research should compare the validities of such theoretical explanations with one another.

Loss of Validity Due to Acquaintance Between Target and Perceiver?

We also investigated whether judgments of the behavior of known targets are less accurate than judgments of the behavior of unknown targets. Using averaged observer-ratings as an accuracy criterion, we found that the targets were in fact systematically less accurate in judging their own behavior as compared to how accurate they were in judging the behavior of other people they did not know. However, it may not be concluded that self-judgments of behavior are generally less accurate than are other-ratings because we did not find significant differences in accuracy between self-ratings and informant-ratings of behavior. Also, the informants were about equally accurate in judging the behavior of known targets and unknown targets. Therefore, the present study does not suggest that prior acquaintance between perceivers and targets is always detrimental to the perceivers' accuracy in judging the targets' behavior. According to our view, the finding that the targets were more accurate in judging someone else than in judging themselves may be attributed to the operation of self-protective mechanisms in self-judgments (e.g., Tice, 1991) as opposed to the targets' *heightened attention* to someone else's performance in the same situations that they had just undergone themselves.

It should be noted, however, that the targets and the informants in the present study judged the targets' behavior directly after seeing them on video. This probably made it relatively easy for them to be accurate because they received the exact same behavioral information on which the observers also based their judgments. The fact that, even under such optimal conditions for accuracy, we still found significant effects for our bias variables (see above) renders these effects even more impressive. We suspect that, under conditions where the interval between the observation and the judgment of behavior is larger, the impact of the perceivers' preexisting views of the targets on their judgments of the targets' behavior might become even stronger and possibly come to impair accuracy. This issue needs to be addressed by future research.

Global Evaluation Across Perspectives

We also investigated mean differences in how positively or negatively the targets' behaviors and personalities were described across the various assessments. First, we found

that the informants' judgments of "their" targets' behavior in the lab were more positive than any other judgments of that behavior. This clearly supports the notion of a "pal-serving bias" (Leising et al., 2010), that is, partisanship in judgments of the behavior of others we know and like. It should be noted that this finding concerns a main effect (people in general are partisan to their friends), whereas the above-described effects of PEV-GE concern individual differences (people who generally see their targets more positively will describe those targets' subsequent behaviors more positively). It should also be noted that we did not find evidence for a general self-serving bias in judgments of behavior. One possible explanation for this discrepancy may be that praising others is less "taboo" than praising oneself (cf. Gallrein, Carlson, Holstein, & Leising, 2013).

Second, we found that judgments of personality were more positive than judgments of behavior in the lab. This may be explained by our tasks eliciting behaviors that were actually less "positive" than the behaviors that most people tend to show in their everyday lives (the latter constituting the basis for personality judgments). Alternatively, it may be the case that judgments of behavior from video are more data-driven than broad retrospective judgments of personality in a questionnaire. Because the latter kind of judgment is more abstract, it may permit more target-serving, selective recall of relevant behavioral episodes.

Broader Implications and Outlook

Which implications do the present results have for behavioral assessment in general? Most importantly, they suggest that, apart from "truth," momentary self- and other-judgments of behavior reflect a substantial proportion of systematic bias variance. If a target's behavior is judged by someone who is previously acquainted with the target, the resulting judgments will partly reflect the perceiver's preexisting view of the target, both in terms of general evaluation and in terms of the specific judgment dimension at hand. In other words, if we know how a perceiver generally views a target, we will be able to partly predict that perceiver's judgments of that target's subsequent behaviors, irrespective of how the target actually behaves. This might explain some of the stunning discrepancies that sometimes exist between different perceivers' views on the same targets' behaviors (e.g., in our introductory real-life example).

The bias effects that we found may become a problem especially when multiple momentary assessments of behavior are averaged (as is often the case in experience sampling studies). For example, if we assume that people's self-images are relatively stable over time, they might be expected to affect all momentary self-assessments of behavior in a similar manner. In contrast, the targets' *actual* behaviors in the different situations may contribute less to the average of momentary self-assessments, to the extent that they vary over time. Taken to the extreme, it would be conceivable that

an average of many momentary self-assessments may only reflect the target's overall self-images, but none of the target's actual behaviors in specific situations anymore, because the latter have averaged out. In such a case, the momentary assessments would become superfluous because a single overall self-assessment would do the job (of assessing what people think of themselves) just as well. As the current study shows, using informant-instead of self-judgments of behavior would not be a solution to this problem because informant-judgments are just as prone to being biased by preexisting views as are self-judgments. The question of how much *incremental validity* (e.g., in predicting important outcomes) aggregated momentary self- or informant-judgments of behavior actually have beyond the respective perceivers' general images of the targets remains to be answered by future research.

The present study also shows that people tend to judge the behavior of their friends, but not their own behavior, in overly positive ways. It seems appropriate to use the term *overly* here because we could show that these judgments were significantly more positive than judgments of the very same behaviors, by means of the very same terms, by anybody else, including neutral observers. The finding is in line with previous research showing that people tend to view their close acquaintances more positively than they view themselves (Barelds & Dijkstra, 2009; Leising et al., 2010; McNulty, O'Mara, & Karney, 2008; Murray, 1999; Murray, Holmes, & Griffin, 1996; Swami, Stieger, Haubner, Voracek, & Furnham, 2009) but—according to our knowledge—the present study is the first study to actually demonstrate that informant-ratings of behavior deviate positively from independent assessments of “the truth.”

This has several important implications for person assessments in real life: For targets, systematically befriending people who might be asked to judge their behavior later on (e.g., as members of the board during job applications or as reviewers for scientific journals) is likely to pay off in terms of obtaining better evaluations. Institutions (e.g., employers, scientific journals) that are interested in assessing targets' performances objectively, however, should take care that no particularly positive or negative personal relationships exist between targets and judges because otherwise the resulting judgments might reflect the influence of those relationships quite strongly (e.g., the average difference between the informants' global evaluations of the behavior of known vs. unknown targets was $d = .51$ in the present study).

An important limitation of our study is that all informants were relatively fond of the targets. As a consequence, the influence of PEV-GE was probably *underestimated* due to variance restriction (Leising et al., 2010; Peabody & Goldberg, 1989). It seems likely that the force of this variable would become stronger when including perceivers who differ more in how favorable or critical an image they have of their respective targets. With regard to self-ratings, this may be achieved by including, for example, participants with

more negative self-images (e.g., people with elevated levels of depression). With such a sample, we would expect global evaluation to play an even more prominent role in predicting the targets' judgments of their own behaviors. When informant-ratings are used, a more realistic estimate of the global evaluation effect may be obtained by also recruiting informants who have no particular loyalties toward, or who even dislike, their respective target persons. Such recruitment may be challenging, but would clearly be worthwhile at the same time, because the resulting samples of targets and perceivers would reflect the interpersonal world we inhabit (comprising friends and foes, allies and competitors, love interests, loved ones and “exes”) more accurately.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by a grant (LE 2151/4-1) to Daniel Leising provided by the German Research Foundation (DFG).

References

- Barelds, D. P. H., & Dijkstra, P. (2009). Positive illusions about a partner's physical attractiveness and relationship quality. *Personal Relationships, 16*, 263-283. doi:10.1111/j.1475-6811.2009.01222.x
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396-403. doi:10.1111/j.1745-6916.2007.00051.x
- Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *European Journal of Personality, 23*, 509-530. doi:10.1002/per.731
- Bono, J. E., & Judge, T. A. (2003). Core self-evaluations: A review of the trait and its role in job satisfaction and job performance. *European Journal of Personality, 17*, 5-18. doi:10.1002/per.481
- Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology, 86*, 599-614. doi:10.1037/0022-3514.86.4.599
- Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality, 32*, 202-221. doi:10.1006/jrpe.1997.2206
- Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*, 69-106.
- Festinger, L. (1957). *A theory of cognitive dissonance* (Vol. 2). Stanford, CA: Stanford University Press.
- Fleeson, W. (2001). Towards a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology, 80*, 1011-1027. doi:10.1037//0022-3514.80.6.1011

- Furr, R. M., & Funder, D. C. (1998). A multimodal analysis of personal negativity. *Journal of Personality and Social Psychology*, 74, 1580-1591. doi:10.1037/0022-3514.74.6.1580
- Gallrein, A. M. B., Carlson, E. N., Holstein, M., & Leising, D. (2013). You spy with your little eye—People are “blind” to some of the ways in which they are consensually seen by others. *Journal of Research in Personality*, 47, 464-471. doi.org/10.1016/j.jrp.2013.04.001
- Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York, NY: Guilford.
- Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality and Social Psychology Review*, 8, 265-280. doi:10.1207/s15327957pspr0803_3
- Kim, H., Schimmack, U., & Oishi, S. (2012). Cultural differences in self- and other-evaluations and well-being: A study of European and Asian Canadians. *Journal of Personality and Social Psychology*, 102, 856-873. doi:10.1037/a0026803
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498. doi:10.1037/0033-2909.108.3.480
- Leising, D. (2011). The consistency bias in judgments of one's own interpersonal behavior. Two possible sources. *Journal of Individual Differences*, 32, 137-143. doi:10.1027/1614-0001/a000046
- Leising, D., & Bleidorn, W. (2011). Which are the basic meaning dimensions of observable interpersonal behavior? *Personality and Individual Differences*, 51, 986-990. doi.org/10.1016/j.paid.2011.08.003
- Leising, D., & Borkenau, P. (2011). Person perception, dispositional inferences and social judgment. In L. M. Horowitz & S. Strack (Eds.), *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions* (pp. 157-170). Hoboken, NJ: John Wiley & Sons.
- Leising, D., Erbs, J., & Fritz, U. (2010). The letter of recommendation effect in informant ratings of personality. *Journal of Personality and Social Psychology*, 98, 668-682. doi:10.1037/a0018771
- Leising, D., & Müller-Plath, G. (2009). Person-situation integration in research on personality problems. *Journal of Research in Personality*, 43, 218-227. doi:10.1016/j.jrp.2009.01.017
- Leising, D., Ostrovski, O., & Zimmermann, J. (2013). Are we talking about the same person here? Inter-rater agreement in judgments of personality varies dramatically with how much the perceivers like the targets. *Social Psychological and Personality Science*, 4, 468-474.
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47, 1231-1243. doi:10.1037/0022-3514.47.6.1231
- McNulty, J. K., O'Mara, E. M., & Karney, B. R. (2008). Benevolent cognitions as a strategy of relationship maintenance: “Don't sweat the small stuff.” But it is not all small stuff. *Journal of Personality and Social Psychology*, 94, 631-646. doi:10.1037/0022-3514.94.4.631
- Munro, G. D., Ditto, P. H., Lockhart, L. K., Fagerlin, A., Gready, M., & Peterson, E. (2002). Biased assimilation of sociopolitical argument: Evaluating the 1996 U.S. Presidential Debate. *Basic and Applied Psychology*, 24, 15-26.
- 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., & 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
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, 175-220. doi:10.1037/1089-2680.2.2.175
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231-259. doi:10.1037/0033-295X.84.3.231
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67-88). Hillsdale, NJ: Erlbaum.
- Peabody, D., & Goldberg, L. R. (1989). Some determinants of factor structures from personality-trait descriptors. *Journal of Personality and Social Psychology*, 57, 552-567. doi:10.1037/0022-3514.57.3.552
- Sadler, P., & Woody, E. (2003). Is who you are who you're talking to? Interpersonal style and complementarity in mixed-sex interactions. *Journal of Personality and Social Psychology*, 84, 80-96. doi:10.1037/0022-3514.84.1.80
- Saucier, G. (1994). Separating description and evaluation in the structure of personality attributes. *Journal of Personality and Social Psychology*, 66, 141-154. doi:10.1037/0022-3514.66.1.141
- Snyder, M., & Campbell, B. (1980). Testing hypotheses about other people: The role of the hypothesis. *Personality and Social Psychology Bulletin*, 6, 421-426. doi:10.1177/014616728063015
- Snyder, M., & Swann, W. B. (1978). Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, 36, 1202-1212. doi:10.1037/0022-3514.36.11.1202
- Srivastava, S., Guglielmo, S., & Beer, J. S. (2010). Perceiving others' personalities: Examining the dimensionality, assumed similarity to the self, and stability of perceiver effects. *Journal of Personality and Social Psychology*, 98, 520-534. doi:10.1037/a0017057
- Swami, V., Stieger, S., Haubner, T., Voracek, M., & Furnham, A. (2009). Evaluating the physical attractiveness of oneself and one's romantic partner: Individual and relationship correlates of the love-is-blind bias. *Journal of Individual Differences*, 30, 35-43. doi:10.1027/1614-0001.30.1.35
- Tice, D. M. (1991). Esteem protection or enhancement? Self-handicapping motives and attributions differ by trait self-esteem. *Journal of Personality and Social Psychology*, 60, 711-725. Retrieved from <http://dx.doi.org/10.1037/0022-3514.60.5.711>
- West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review*, 118, 357-378. doi:10.1037/a0022936
- Wood, D., Harms, P., & Vazire, S. (2010). Perceiver effects as projective tests: What your perceptions of others say about you. *Journal of Personality and Social psychology*, 99, 174-190. doi:10.1037/a0019390