American Psychological Association

Journal of Personality and Social Psychology:
Interpersonal Relations and Group Processes

0022-3514/20/$12.00

2020, Vol. 118, No. 4, 720–742
http://dx.doi.org/10.1037/pspi0000204

The Overblown Implications Effect

Alice Moon
University of Pennsylvania

Muping Gan and Clayton R. Critcher
University of California, Berkeley

People frequently engage in behaviors that put their competencies on display. However, do such actors understand how others view them in light of these performances? Eight studies support an *overblown implications effect* (OIE): Actors overestimate how much observers think an actor's one-off success or failure offers clear insight about a relevant competency (Study 1). Furthermore, actors overblow performances' implications even in prospect, before there are experienced successes or failures on which to ruminate (Studies 2 and 3). To explain the OIE, we introduce the construct of *working trait definitions*—accessible beliefs about what specific skills define a general trait or competency. When actors try to adopt observers' perspective, the narrow performance domain seems disproportionately important in defining the general trait (Study 4). By manipulating actors' working trait definitions to include other (unobserved) trait-relevant behaviors, we eliminated the OIE (Study 5). The final 3 studies (Studies 6a–6c) more precisely localized the error. Although actors and observers agreed on what a single success or failure (e.g., the quality of a single batch of cookies) could reveal about actors' narrow competence (e.g., skill at baking cookies), actors erred in thinking observers would feel this performance would reveal a considerable amount about the more general skill (e.g., cooking ability) and related specific competencies (e.g., skill at making omelets). Discussion centers on how the present theoretical account differs from previous explanations why metaperceptions err and identifies important open questions for future research.

Keywords: metaperceptions, social judgment, working trait definitions, definitional focalism

Supplemental materials: http://dx.doi.org/10.1037/pspi0000204.supp

"Is that your *final* answer?" In the well-known game show *Who Wants to Be a Millionaire?*, the contestant sits in the hot seat answering trivia questions for a shot at riches. For the contestant, each question is high stakes. Most obviously, the monetary stakes are high: Correct answers are necessary to advance toward the shot at the million dollars. But less focally, the evaluative stakes are high as well: Contestants' every move is being closely watched by a couple hundred studio audience members and thousands more at home.

Social psychologists have long appreciated the importance of others' presence on behavior. Such evaluative pressure can at times facilitate effort and, thus, performance (Harkins, 2006; Zajonc, 1965; see Seitchik, Brown, & Harkins, 2017, for a review), but at other times can lead performers to choke (Baumeister, 1984). Both (contrasting) findings reinforce that evaluative stakes matter, but do actors appreciate the true evaluative stakes when in the spotlight? In this article, we argue that a concern about imminent evaluation leads people to show an *overblown implications effect* (OIE): Actors exaggerate how much their performance will speak to their broader competencies in observers' eyes. We ultimately explain this phenomenon by introducing and demonstrating the role of a novel construct, *working trait definitions*—what behaviors form one's momentary definition of a trait or competency. The OIE arises because observers' actual working trait definitions are broader—encompassing more than the specific performance domain—than actors assume. In introducing this new construct (working trait definitions) and developing our theoretical account, we will give particular attention to how this novel psychological mechanism distinguishes itself from previously identified reasons why metaperceptions—people's guesses of how others view them—err.

This article was published Online First July 15, 2019.

Alice Moon, Department of Operations, Information, and Decisions, The Wharton School, University of Pennsylvania; Muping Gan, Department of Psychology, University of California, Berkeley; Clayton R. Critcher, Haas School of Business, University of California, Berkeley.

The research reported in this article was supported in part by: a National Science Foundation Graduate Research Fellowship to Alice Moon; the Wharton Dean's Research Fund and U.S. National Science Foundation award 1749608 to Clayton R. Critcher. We thank Sara Wingrove for her assistance in developing stimuli and organizing datasets, and the following research assistants for their assistance in collecting data: Rohit Agrawal, Kirsten Craig, Nora Harhen, Noma Kahf, David Lee, Sophia Liu, Maria Monroy, Ashley O'Lear, Brianna Pogue, Tanvi Sheth, Jennifer Soliman, Amanda Sun, Sanjana Surkund, Jordy Tandanu, Isha Thapa, Josh Woznica, Gloria Yen, Shun Ting Yung, and Rowena Zhang. We also thank Tom Gilovich for his comments on a draft of this article.

Correspondence concerning this article should be addressed to Alice Moon, Department of Operations, Information, and Decisions, The Wharton School, University of Pennsylvania, Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104. E-mail: ajmoon@wharton.upenn.edu

Why would actors fail to appreciate how they are judged in light of their performance? To begin answering this question, it is important to consider what is being judged when evaluating a trait or competency. For example, what does it mean to evaluate someone as intelligent? Most obviously, intelligent people do intelligent things. They may use more ornate words, remember obscure facts from childhood, or even know the order of all 118 elements in the Periodic Table.

However, cues to a trait or quality are rarely observed all at once and are not perfectly correlated. Trait-relevant behaviors show variability across situations (Fleeson, 2004; Pervin, 1994; Ross & Nisbett, 1991). In part, different situational factors either overwhelm or selectively activate different aspects of one's personality (Cramer et al., 2012). Furthermore, any single behavior gives only so much information about a particular quality in question. Social perceivers appreciate this notion: When determining another's intelligence, one wants to know more than whether they make it all the way from hydrogen to ununoctium.

From this perspective, one potential pitfall to meta-insight is actors may fail to understand how observers characterize or define the trait or competency in question. Psychologists acknowledge that representations of complex constructs may be based on only a limited amount of information at any one time. Consider the self, a target about which we have an almost overwhelming amount of information. Perhaps because these stores are vast, people's *working self-concepts*—their accessible self-knowledge (Markus & Wurf, 1987)—show moment-to-moment variability (Cervone & Shoda, 1999; DeSteno & Salovey, 1997; Markus & Kunda, 1986; McConnell, 2011) with predictable consequences for judgment and behavior (e.g., Showers & Zeigler-Hill, 2003).

Much as people have working self-concepts, we propose that people hold *working trait definitions*. That is, at any given moment, people define a trait or competency by drawing on only a subset of potentially relevant behavioral dimensions. Previous researchers have recognized that people see evidence of broader competencies by appealing to different specific behaviors—often in a self-serving (Critcher, Helzer, & Dunning, 2011; Dunning, Meyerowitz, & Holzberg, 1989; Greve & Wentura, 2003; Hayes & Dunning, 1997; Kurman, 2003) or group-serving (e.g., Maass, Salvi, Arcuri, & Semin, 1989; Wigboldus, Semin, & Spears, 2000) manner. We consider how a person will show variability in their trait definitions—more specifically, their working trait definitions—depending on their *role* in a performance context.

Recent research has argued that under threat, the working self-concept constricts around the threatened domain, leading this damaged identity to occupy a larger portion of the active self-concept (Critcher & Dunning, 2015). For example, after failing an exam, one's academic self looms large in the working self-concept, thereby exerting a disproportionately large effect on one's feelings of self-worth. By analogy, we suggest that working trait definitions may show similar properties. As actors consider the impressions observers form of them, this evaluative threat may cause the specific performance domain to loom large in metaperceivers' working trait definitions. For example, while parallel parking, drivers may assume that an observer's definition of good driving is heavily dominated by parallel parking ability. As a result, actors see their own performance as likely to exert considerable sway on observers' broader impressions. However, observers' actual impressions do not have the same evaluative stakes for observers themselves. Thus, observers' working trait definitions may not be as dominated by the performance behavior as actors anticipate. That is, for the observer on the sidewalk, driving skills encompass not merely a particular type of parking maneuver, but also awareness of blind spots, attention to road signs, and maintaining safe distances from other drivers. We propose that this asymmetry is a primary cause of the overblown implications effect. We summarize this account in Figure 1.

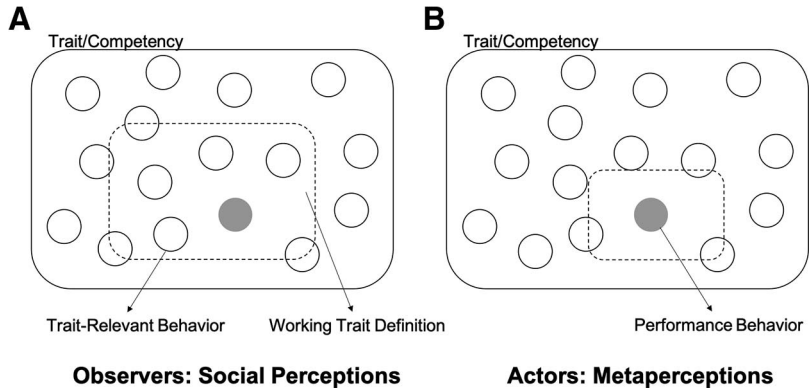

Figure 1. Why social and metaperceptions are hypothesized to diverge and produce the overblown implications effect. At baseline, observers' working trait definitions (the inner, dashed box) may not account for all trait-relevant behaviors (explaining why some trait-relevant behaviors are always outside of the working trait definition rectangle; panel A), but actors' metaperceptions constrict around their performance behavior under evaluative threat (panel B). This predicts that actors' metaperceptions will be more reactive to their own successes or failures than observers' social perceptions, but that expanding actors' working trait definitions should debias them. Note that even when actors and observers agree about the narrow implications of the threat (the performance behavior), differences in their working trait definitions may lead them to draw different inferences about the broader competency—general trait impressions that may later inform judgments about other trait-relevant behaviors.

Empirical and Conceptual Similarities to and Differences From Previous Research

To date, much of the research on meta-insight has examined whether people's metaperceptions correlate with how others actually view them or are merely an egocentric product of self-perceptions (Kenny, 1994; Kenny & DePaulo, 1993). And indeed, correlations between self- and metaperceptions emerge. For example, those who view themselves as sociable believe others are more likely to view them as sociable as well. People believe their self-perceptions are accurate and, thus, leaning on them to form metaperceptions would seem natural (Albright, Forest, & Reiseter, 2001; Albright & Malloy, 1999; Kenny & DePaulo, 1993; Malloy, Albright, Kenny, Agatstein, & Winquist, 1997). But beyond this, people also possess special meta-insight—particularly for observable qualities (Gallrein, Carlson, Holstein, & Leising, 2013)—an understanding of how they are viewed by others that does not merely stem from self-perceptions (Carlson & Kenny, 2012; Carlson, Vazire, & Furr, 2011; Vazire & Carlson, 2010). For example, people understand that they make different impressions on different groups of people, such as on their friends versus their parents (Carlson & Furr, 2009), suggesting that people's metaperceptions do not merely rely on their global, stable self-perceptions.

In this article, we examine *errors* in meta-insight, but depart from much of the just-reviewed approaches in two key ways. First, instead of examining global impressions formed over time, we examine how impressions of competencies are formed or moved by one-off successes or failures. Second, we focus not on correlations between observers' social perceptions and actors' metaperceptions, but instead probe for systematic mean-level differences between them. After all, a driver who completes a simple parallel parking job in three turns clearly has more driving ability than one who takes 23 turns to do the same. In other words, we expect that meta- and social perceptions of drivers would likely correlate. However, by examining systematic, mean-level (i.e., directional) biases in impressions, we can know whether observers' impressions are influenced by such performances to a greater or lesser extent than actors assume (see Hayes & Dunning, 1997, for an eloquent exposition of this point).

To our knowledge, we are the first to posit the existence of working trait definitions and use them to explain why metaperceptions may err. However, we are of course not the first to explore whether people show systematic biases in understanding the impressions their performances leave on others. The most similar and influential work in this tradition has shown that people fail to recognize how charitably others respond to their blunders (Epley, Savitsky, & Gilovich, 2002; Savitsky, Epley, & Gilovich, 2001). Just as decades of research have identified many reasons why self and social perceptions display biases (e.g., Critcher, Dunning, & Rom, 2015; Dunning, 2005; Kruger & Dunning, 1999; Ross & Sicoly, 1979), we have little doubt that failures of meta-insight are themselves multiply determined. That said, we focus next on what most differentiates the present account, mechanistically, from that of previous research. In so doing, we will identify a number of unique implications of the present account.

First, we distinguish ourselves in our claim of what actors are fixated on. We argue that metaperceivers assume observers *define* traits or competencies more narrowly than they do. In contrast, previous work has attributed errors in meta-insight to a tendency for actors to narrowly focus on certain aspects of the performance situation. Savitsky et al. (2001) suggest actors become mentally stuck on their blunders, prompting those unrepresentative moments to loom large in mind. For example, participants (misleadingly and sometimes inaccurately) outed as bedwetters failed to appreciate how additional information would water down observers' negative impressions (Savitsky et al., 2001). As Epley et al. (2002) noted, "musicians who miss one key note in a concert still hit countless others" (p. 310). Whereas these accounts focus on *performance focalism* (disproportionate attention to a blunder while ignoring the largely competent moments), we focus on what we call *definitional focalism*. This means our own account applies even in contexts in which a blunder (or a success) occurs in isolation—that is, without the context of opposing information (e.g., the rest of the concert). In such cases, observers' judgments are tempered because they recognize that they have relatively little information about actors' broader competencies ("I only know he can bake cookies, not whether he is a good cook"), not because they focus on the actor's blunder at the expense of information (given such information is unknown).

Second, our account has unique implications for the *nature* of the error metaperceivers make. By our definitional focalism account, meta-insight fails when translating an evaluation of a specific competency (e.g., parallel parking) into an evaluation of a broader competency (e.g., driving). This contrasts with previous work, which has focused on an earlier step: deciding what a single performance, especially when undertaken under suboptimal conditions, reflects about the narrow competence in question. That musician—by focusing on his one sour note—may mistakenly think his audience has also encoded his performance as a fiasco. Or consider Epley et al. (2002), who had some participants sing the "Star Spangled Banner" while chewing gum. In part because people overestimate how much observers display the correspondence bias (Van Boven, Kamada, & Gilovich, 1999), metaperceivers fail to appreciate how much observers actually do realize the gum plays an interfering role. In both examples, metaperceivers are confused about how observers evaluate the specific performance. We instead argue that there is only so much observers feel they learn about another's musical talents by hearing the performance of one song. This means that we expect the OIE to hold even when there are no situational explanations to appeal to (i.e., the chewing of gum while singing). For example, we predict that metaperceivers will be quite accurate in guessing how much diagnosticity an observer will see in a one-off (unimpeded) performance of the "Star Spangled Banner" for determining skill at singing the national anthem. However, despite anticipating such accuracy, our account predicts metaperceivers will fail to realize that observers will see little information about their musical ability more generally based on this one-off performance.

Third, this reasoning leads us to hypothesize that actors will overblow the implications not merely of performance failures but of successes as well. In contrast, previous work has examined actors' fear of negative evaluations following failure exclusively (Epley et al., 2002; Savitsky et al., 2001). Fourth, previous research has emphasized that observers are in privileged positions that help them recognize the difficulty of the performance context (Epley et al., 2002). Observers witnessing failure often know that *they too* would have failed in a context. This allows observers to understand, for example, that a trivia question is especially diffi-

cult, even though actors lack the same privileged perspective. This too has been identified as a form of focalism: Actors focus on the superficial characterization of their performance as a "failure," not appreciating how observers will be able to empathize with that failure as understandable. This argues not that observers will see less diagnosticity in failures and successes (as we argue), but that their own privileged perspective allows them to recognize a performance context as objectively difficult or easy. This alternative has empirically distinct implications: Observers' personal insight that a task is actually quite difficult or easy means they should judge targets—regardless of whether they succeed or fail—more positively or negatively, respectively, than actors would expect.

Overview of the Present Research

The present research examined if actors overweight how much a performance factors into observers' perceptions of them. Study 1 tests the overblown implications effect in the context of social exclusion and inclusion—that is, a performance failure *and* a success. Studies 2–4 provide initial tests of our working trait definitions account by examining whether observers see less diagnosticity in a future performance than actors would expect (Studies 2 and 3), whether this OIE reflects *metaperceivers'* constricted working trait definitions as opposed to a more general social judgment error (Study 2), and whether performance behaviors—compared with other behaviors that could speak to the same broader competency—loom large in metaperceivers' working trait definitions (Study 4). Study 5 directly manipulated actors' and observers' working trait definitions to determine whether expanding them would debias metaperceptions while leaving observers' social judgments unaffected. Studies 6a–6c tried to localize the OIE to the translation of evaluations of specific competencies (about which we expected actors and observers to agree) into the assessment of what these specific competencies say about general competencies and related behaviors (where constricted working trait definitions should lead metaperceptions and social perceptions to diverge).

Compliant with Simmons, Nelson, and Simonsohn (2011), we report how we determined our sample size as well as all data exclusions (if any), manipulations, and measures (for more information, see https://osf.io/4vzy3/?view_only=c303412bdab842c493958c069aeac7ef). Given we did not have a priori knowledge of the hypothesized effects' true effect sizes—a common problem that limits the usefulness of power analyses—we predetermined a stopping rule for data collection. For each of the multistage lab studies (Studies 1 and 5), we collected as many participants as we could in a year from the undergraduate subject pools we used. For the first two survey studies we conducted (Studies 2 and 3), we aimed for at least 100 participants per cell. Because we predicted a fully attenuated interaction in Study 6a, we attempted to double this sample size to 200 participants per cell. And for the final set of studies we conducted (Studies 4, 6b, and 6c)—some of which included more subtle hypotheses (e.g., partially attenuated interactions) or complex analyses—we doubled this standard again and aimed to recruit 400 participants per cell.

Study 1

Study 1 tested the *overblown implications effect* in a social domain. We placed some participants (actors) in a social context in which they were ostensibly socially accepted or rejected. Yoked observers watched the social dynamic unfold. We then examined how this social success or failure changed perceptions of actors' likability. According to the *overblown implications effect*, actors' metaperceptions of how generally likable they seem to observers should be more reactive to whether they socially succeeded or failed than observers' perceptions actually are.

Study 1 had two primary goals. First, we wanted to test whether the overblown implications effect would occur for social successes, not merely social failures. After all, Savitsky et al. (2001) expressed pessimism that such a result would emerge. They posited two mechanisms that would lead people to overblow the implications of failures but not successes: (a) people hold a naively cynical belief that others are even more self-serving and other-disparaging than they actually are (Kruger & Gilovich, 1999), and (b) people display empathy neglect (Epley et al., 2002)—a tendency for actors to overlook observers' generally charitable orientation. These forces would push actors to assume observers would give them less, not more, credit for their successes. Our account instead suggests that actors will overblow the implications of successes as well.

Second, we offered an initial attempt to disentangle our overblown implications account from an alternative performance focalism account. By the alternative, actors fail to understand observers not because they misunderstand the implications of being accepted or rejected, but because they forget about all of the other cues they have offered about their own sociability. That is, actors focus on their own sour note (or impressive high note) and forget about the rest of the song they performed. Instead, our working trait definition account suggests actors fail to appreciate how little observers feel like they can even learn from one particular social performance context. Observers' broader working trait definitions of likability encompass more specific competencies than what can be gleaned from the present context.

We disentangle these possibilities by separating actors' performance into two halves. Actors' success or failure was based only on the second half. This allowed us to see if actors' metaperceptions erred because they gave too much weight to the half on which they performed well or poorly (consistent with the performance focalism alternative). We instead predicted actors would simply fail to understand the (limited) implications of their success or failure, even when everyone's attention is focused narrowly on the part on which one succeeded or failed.

Method

Participants. Two hundred forty-nine undergraduates completed a lab session for course credit. All actors were randomly assigned to the social *success* or *failure* performance condition. Observers were yoked to a randomly selected actor.

Procedure and materials. We describe actors' experience first. Observers learned the instructions actors received and read all of their specific actor's responses. Figure 2 summarizes the general procedure for actors and observers.

Actors. Actors were seated in a private room in front of a laptop. They were told that they would be completing a study examining: (a) whether people work better with those they like, and (b) whether outside observers can anticipate which groups will work best. Actors were led to believe that there were three partic-

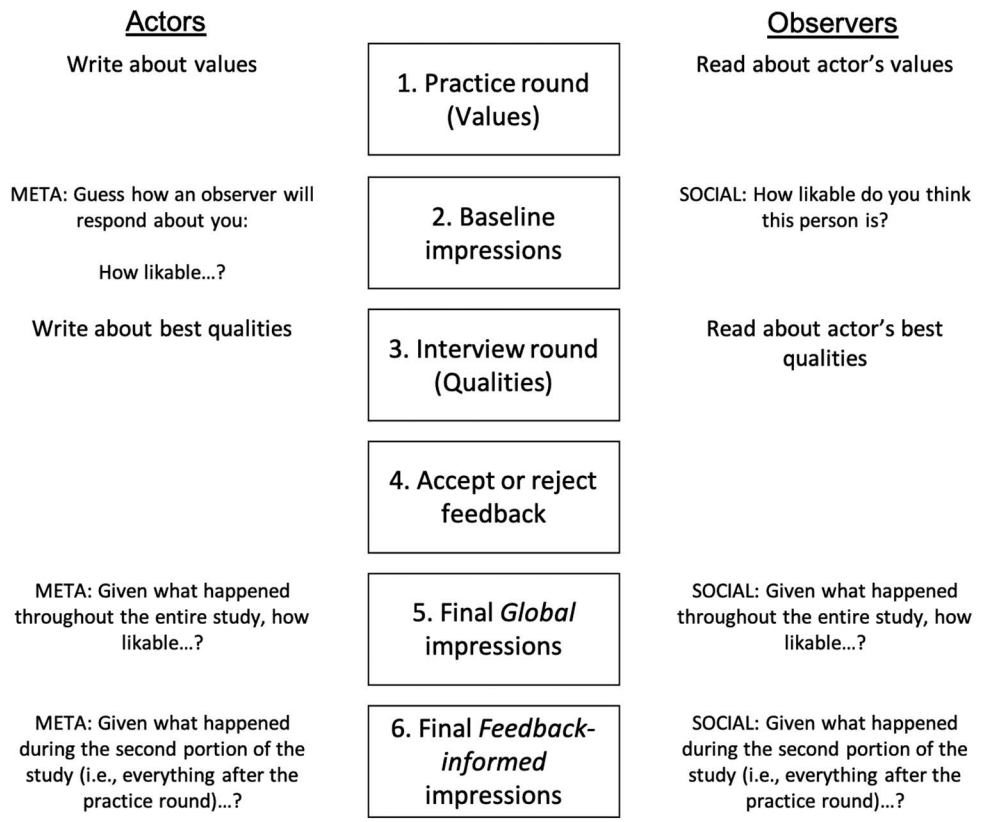

Figure 2. Summary of Study 1's six-stage procedure (from 1 to 6) for actors (left) and observers (right). In addition to metaperception ratings, actors also completed self-perception ratings.

ipants present: two who would be interviewees and one who would be the interviewer. Actors were also informed (truthfully) that in a later session, an observer would watch the situation unfold from the vantage point of the actor, but would not participate in the groups themselves. A graphic was shown along with these instructions to more clearly illustrate the different roles involved.

Actors were told that they would be randomly assigned via a card-choosing task to either the interviewer role or one of two interviewee roles. They were asked to select one of three cards. The procedure was rigged: All actors saw their card assigned them to be an "Interviewee." The interviewer and second interviewee were actually preprogrammed, fictitious participants.

Actors—as one of the Interviewees—were told that they would answer two rounds of questions: (a) a practice round question, not shown to the interviewer, to familiarize them with the question format; and (b) an interview round question, based on which the interviewer would choose whether they would like to work with one, both, or neither of the interviewees on the final "fun" task. A diagram (see Figure 3A) reiterated the role the actor would take as well as the roles of the other participants. To enhance the believability that there were other participants in the current session (playing the role of the interviewer and the other interviewee), the next screen displayed a spinning "loading" icon for 5 s and explained that the survey would automatically continue once everyone had finished learning about their roles.

First, actors completed the practice round of questions, which only the observer would see. This permitted observers to have grounds on which to offer baseline perceptions. In the practice round, actors listed three values that were important to them, then described why each of the values they listed had such importance. They wrote about each value for 1 min. At this point, actors completed the baseline perception measures of likability. Actors provided self-perceptions (evaluations of their own performance) followed by metaperceptions (guesses of how the observers would rate them).

Next, actors completed the interview round of questions, which would be seen by both the interviewer and the observer. In the interview round, actors listed three of their best qualities, then described an instance in which they exhibited each. They described each quality for 90 s, sequentially, until they described all three qualities. After completing the interview round, actors waited for the interviewer to make a decision. Actors saw a spinning loading icon for 25 s and were informed that the interviewer was reading their responses (as well as those of the other interviewee) to decide whether they would like to work with one, both, or neither of the interviewees.

Actors were randomly assigned to learn that the interviewer had chosen to work with them but not the other interviewee (social *success*) or that the interviewer had chosen to work with the other interviewee but not with them (social *failure*). At this point, actors completed the *final* metaperception and self-perception measures of likability—both *global* (based on the entire study) and *feedback-informed* (focusing only on the portion involving social acceptance or rejection). Finally, actors were debriefed and apol-

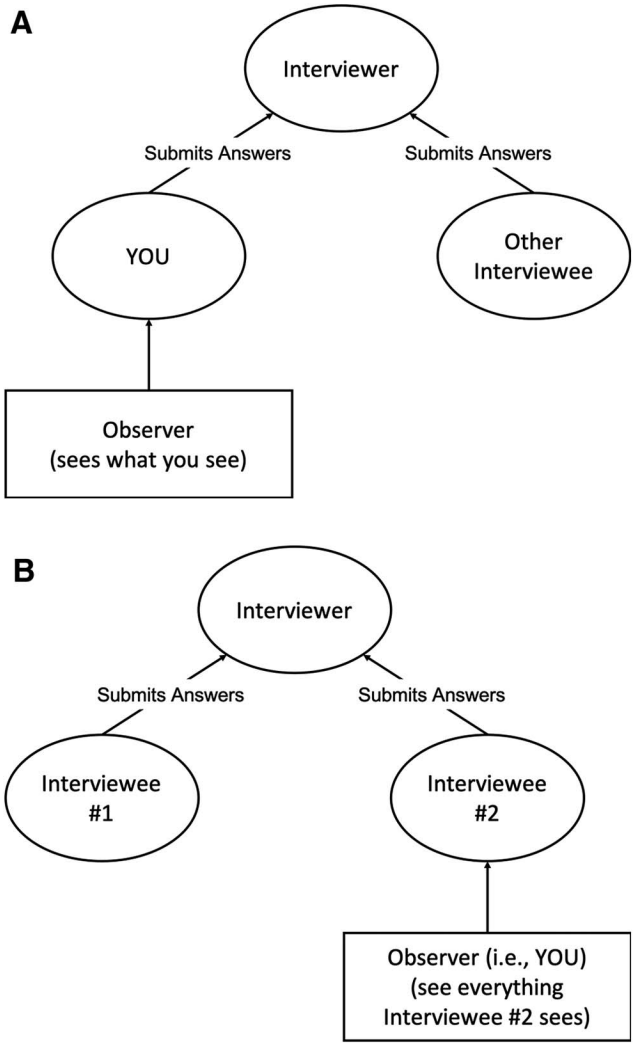

Figure 3. Diagrams explaining to actors (A) and observers (B) their role in Study 1. Actors knew all of their behaviors and experience would be observed by an outside observer. Observers saw the experiment through the perspective of an actor (whom we called "Interviewee #2").

ogized to for the mild deception. No actor (or observer) expressed suspicion about the legitimacy of the interaction or the feedback's credibility.

Observers. Each observer was yoked to one actor. All actors were yoked to at least one observer. We averaged the responses of observers who were yoked to the same actor.

Like actors, observers were also seated in private rooms. However, in observers' cases, they were told they would be *observing* a previous participant who completed a study on whether people work better with those they like and whether outside observers can predict how well a group can work together. Observers saw everything that actors saw, but read actors' responses instead of providing responses of their own.

Observers were shown a diagram (see Figure 3B) that summarized the role they and others were playing. Observers completed *social* perception measures that were parallel to—in timing and in form—the actors' metaperception and self-perception measures. (Observers did not see actors' self- or metaperception responses.) More specifically, observers completed the baseline measures of likability after reading the actor's reply to the practice round prompt. They completed the final social-perception measures after reading the actor's responses to the interview round then witnessing the social acceptance or rejection.

Trait perceptions. The perception measure comprised six items that assessed the actor's likability. The *metaperception* items asked actors to guess how observers would judge them in light of their social performance during the study. Actors answered these questions with the understanding that they were guessing observers' responses to those exact items. The *social perceptions* asked observers to judge the actors in light of their social performance. Similarly, the *self-perceptions* asked actors to judge themselves in light of their performance as well. The first four questions asked participants to rate the actor as "engaging," "likable," "warm," and "charming" on 9-point scales anchored at 1 (*not at all*) and 9 (*extremely*). The final two questions asked participants how much the actor would "make a good impression" and be "able to get along with others" on 9-point scales anchored at 1 (*not at all*) and 9 (*extremely*). We averaged these items to create likability perception composites.

Participants completed two sets of final ratings. The first set asked participants to rate the actors' performance in light of the entire study—that is, based on the *global* set of information to which participants had been exposed. The second set (the *feedback-informed* final ratings) asked participants to rate the actor only based on the performance task in particular. The self-perception ($\alpha = .96$), metaperception ($\alpha = .96$), and social perception ($\alpha = .97$) likability composites all had high reliability.

Results and Discussion

The two competing accounts—*overblown implications* versus *performance focalism*—make the same prediction about how metaperceptions will err regarding global impressions. The two accounts differ in whether metaperceptions will err on the more specific feedback-informed impressions. We submitted both impression ratings to 2 (performance: success or failure) × 3 (perception: meta, social, or self) × 2 (time: baseline or final) mixed-model analysis of variances (ANOVAs). The online supplemental materials present more complete descriptive statistics that separate baseline and final measures for both this study as well as Study 5.

Global final impression. The predicted 2 (performance: success or failure) × 3 (perception: meta, social, or self) × 2 (time: baseline or final) interaction emerged on the global impressions, $F(2, 224) = 6.89$, $p = .001$, $\eta_p^2 = .06$. To determine whether this interaction reflected the predicted pattern, we proceeded to test all three 2 (performance) × 2 (perception) × 2 (time) interactions.

First, we tested whether social perceivers were less reactive to the actors' social success or failure than actors assumed they would be. A 2 (performance) × 2 (perception: meta or social) × 2 (time) interaction suggested that this was the case, $F(1, 112) = 10.96$, $p = .001$, $\eta_p^2 = .09$. As depicted in Figure 4A, although observers did shift their global social impressions in response to actors' success versus failure, $F(1, 112) = 4.09$, $p = .05$, $\eta_p^2 = .04$, this shift was less pronounced than actors thought it would be, $F(1, 112) = 58.88$, $p < .001$, $\eta_p^2 = .34$.

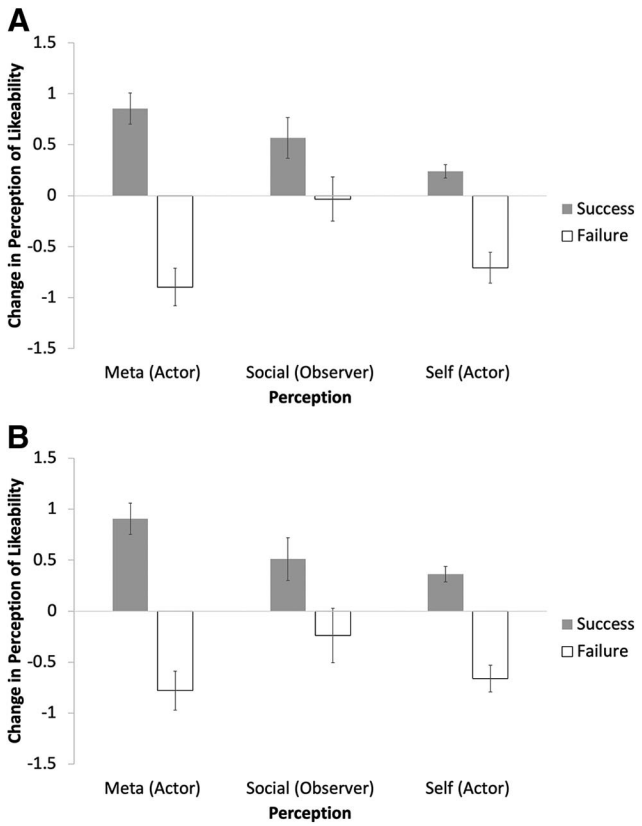

Figure 4. The change in perception (final–baseline) of actors' likability by performance condition and perception type (Study 1). Panel A depicts the global final perceptions; Panel B, the feedback-informed. For each panel, the overblown implications effect is reflected by the larger gap between the two metaperception bars compared with the two social perception bars. Which participant offered each perception is in parentheses. Error bars display ±1 *SE*.

Second, we found that metaperceivers were not merely using their own personal reactions to their performance when making metajudgments about others. That is, we also found a significant 2 (performance) × 2 (perception: meta or self) × 2 (time) interaction, $F(1, 112) = 10.21$, $p = .002$, $\eta_p^2 = .08$. Actors' self-perceptions were not as reactive to their own success or failure, $F(1, 112) = 33.08$, $p < .001$, $\eta_p^2 = .23$, as they assumed observers' perceptions would be.

Third, we observed that actors would have been more accurate in forecasting observers' shift if they had just leaned on their own shift in self-perception. That is, the 2 (performance) × 2 (perception: social or self) × 2 (time) interaction failed to reach significance, $F(1, 112) = 1.02$, $p = .32$.

Feedback-informed final impression. We tested whether perceivers differed in their interpretation of the final performance portion—the one that produced the social acceptance or rejection—and what it signified about the actors' likability. Suggesting perceptions did differ, we found a significant 2 (performance) × 3 (perception: meta, social, or self) × 2 (time) interaction, $F(2, 224) = 3.81$, $p = .02$, $\eta_p^2 = .03$. As before, we decompose this interaction by comparing each pair of perceptions using 2 (performance) × 2 (perception) × 2 (time) mixed-model ANOVAs.

First, we tested whether actors assumed that observers would see actors' social success or failure as more informative about actors' likability than observers themselves actually did. As expected, the 2 (performance) × 2 (perception: meta or social) × 2 (time) interaction was significant, $F(1, 112) = 5.69$, $p = .02$, $\eta_p^2 = .05$. As depicted in Figure 4B, observers did shift their social impressions in response to actors' success versus failure, $F(1, 112) = 4.80$, $p = .03$, $\eta_p^2 = .04$, but actors thought this shift would be more pronounced, $F(1, 112) = 51.64$, $p < .001$, $\eta_p^2 = .32$.

Second, we tested whether actors merely leaned on their own personal reactions to the social acceptance or rejection in guessing observers' reactions, or whether actors thought observers would be particularly reactive. The latter prediction was supported: We observed a 2 (performance) × 2 (perception: meta or self) × 2 (time) interaction, $F(1, 112) = 7.70$, $p = .006$, $\eta_p^2 = .06$. Actors saw fewer implications in their own social acceptance or failure, $F(1, 112) = 44.71$, $p < .001$, $\eta_p^2 = .29$, than they thought that observers would. This once again suggests that actors' failure of meta-insight does not merely reflect a failure of self-insight that is then projected onto observers (see Gallrein et al., 2013; Gallrein, Weßels, Carlson, & Leising, 2016).

Third, we asked whether actors would have been better off leaning on their own self-perceptions in judging how observers viewed them. Consistent with this possibility, we failed to observe a significant 2 (performance) × 2 (perception: social or self) × 2 (time) interaction, $F < 1$. That is, social observers and actors themselves saw similar, and relatively small, implications in the episode that produced social acceptance or rejection.

In short, not only did actors fail to understand how few implications observers would see in observers' final global impressions of them, but they were similarly wrong even when everyone was more narrowly focused on the social interaction that got them included or excluded. (See supplemental Study A for a conceptual replication of this latter finding in the domain of dateability.) Although we found that actors and observers disagreed on the implications of the portion of the performance upon which the social acceptance or rejection was based, it is possible that metaperceivers still focused on an even narrower portion of their performance than observers did. In part, this was the tradeoff that came from using a behaviorally rich context; such a context provides so much information that even by constraining the targets of judgments, we cannot be completely confident that actors and observers are not focusing on somewhat different performance information. Study 2 moves away from a live performance context to address this and another issue more conclusively.

Study 2

If metaperceivers overblow performance implications because their working trait definitions are constricted around the performance domain (what we call *definitional focalism*) rather than because actors become fixated on their failures or successes while neglecting other relevant behavioral or contextual information (*performance focalism*), then we should continue to see evidence of the overblown implications effect under narrower conditions. In Study 2, we asked participants to simulate the perspective of either an actor or an observer to an upcoming performance context. Crucially, there was no additional information that metaperceivers, but not observers, might neglect. Furthermore, participants did not

learn that the performance was a success or a failure. Instead, participants made a pair of judgments: how the actor would be judged if the actor (hypothetically) were to perform well, and then again if the actor (hypothetically) were to perform poorly. When behaviors are seen to have implications for broader traits, they should prompt more divergent judgments when considering a success versus a failure.

By our reasoning, when actors adopt the perspective of an observer, they become focused on an evaluative threat. That is, the observer is someone who is, or who will be, watching and judging actors based on their performance. This leads the specific performance domain to loom disproportionately large in actors' working trait definitions of a more general skill or quality. However, because social perceptions do not have personal stakes for observers, such constriction does not actually occur. This psychology should apply even as one considers a prospective performance.

In contrast, performance focalism depends on mitigating contextual details to temper observers' perceptions: "Embarrassing blunders, after all, do not take place in isolation but instead are part of an ongoing social context . . . a speaker who blows an opening joke often has 49 minutes to recover" (Epley et al., 2002, p. 310). Furthermore, it leans on actors' tendency to ruminate on their performance's extremes: "People's mishaps are often highly salient to them" (Savitsky et al., 2001, p. 49). With both of these features (context and a past performance) removed, it becomes unclear what performance or contextual details metaperceivers would be neglecting in our paradigm. What remains is actors' potentially distorted sense of the performance context's broader implications.

By a second alternative hypothesis, the OIE might not reflect the narrowed working trait definitions that come from the threat of being evaluated, but may instead reflect a general property of how people attempt to read the mind of another person. That is, people may simply guess that others see more diagnosticity in anyone's actions, not just the self's own actions in particular. This alternative is supported by research showing that people think others make more extreme dispositional attributions—not just about the self, but anyone—than they actually do (Pronin, Lin, & Ross, 2002; Van Boven, White, Kamada, & Gilovich, 2003). To address this alternative explanation, we added a third perspective condition. Instead of considering interactions from the perspective of an actor or observer, uninvolved *bystanders* considered the situation from afar but guessed observers' responses. In this way, bystanders estimated the perceptions of someone else (just like actors), but not someone else who was judging the self (unlike actors; see Epley et al., 2002, Study 2, for similar experimental reasoning). The alternative hypothesis predicts that bystanders will also overestimate the implications that observers will see in actors' upcoming performance.

Method

Participants. Three hundred one Americans recruited from Amazon's Mechanical Turk (MTurk) who completed the study for payment. Participants were randomly assigned to one of three perspective conditions: actor, observer, or bystander.

Procedure. Participants considered being in 10 different situations. We told participants that they should fully throw themselves into every simulation, to visualize the scene unfolding and to be attuned to what they would be thinking or feeling. Each scenario described an interaction in which an actor's skills or abilities would be on display to an observer (see Table 1 for a brief description of the scenarios). The wording varied such that the exact same situation was described from the vantage point of the actor, the observer, or an uninvolved bystander. Accompanying each scenario was a picture that was meant to assist with visualization.

As an example, one scenario described a person who had baked and brought cookies to a party. Actors were told, "You baked cookies to take to a party. You overhear someone mention to a friend that you baked the cookies. You watch as the person picks up a cookie to try one."

Observers learned this same information, but from the vantage point of the person about to try the cookies: "A person baked cookies to take to a party. Someone mentions to you which person baked the cookies. You pick up a cookie to try one." Bystanders instead learned, "A person (Person X) baked cookies to take to a party. Someone mentions to another person (Person Y) which person baked the cookies. Person Y picks up a cookie to try one."

At that point, all participants indicated how the observer would view the actor if the actor were successful as well as if the actor were unsuccessful. For the baking scenario, those in the observer

Table 1
Performance Behavior Scenarios (Studies 2 and 3)

Behavior	Broader competency	Observer
Baking cookies	Cooking ability	Person sampling actor's cookie
Answering a trivia question	Intelligence	Person reading actor the question
Remembering to pay back money you borrowed	Exploitativeness	Person who lent actor the money
Introducing a new person to your friend	Inconsiderateness	New person actor is introducing, whom actor just met and had a conversation with
Accepting/rejecting a fork for dessert	Self-control	Fellow diner who knows actor is on a diet and asks if actor wants a fork
Parallel parking	Driving ability	Person waiting for actor to finish parking
Playing chess on your computer	Analytical thinking ability	Person sitting next to actor on a flight
Splitting a restaurant bill	Mathematical ability	Fellow diner looking over actor's shoulder
Conversing with stranger	Social skills	Person overhearing actor's conversation
Leaving work at an unusual time	Work ethic	Coworker asking actor for a ride home at usual time

Note. The behaviors are described as if described to actors.

condition made two judgments in a counterbalanced order: "After sampling [the actor's] cookies, if you thought the cookies tasted *good [bad]*, how good of a cook would you think the other person is?" Those taking the actor's and bystander's perspective tried to guess how observers would respond to this question. All judgments were made on 11-point scales, anchored at 0 (*not at all . . .*) and 10 (*extremely . . .*). We reverse-scored responses to two of these scenarios so higher [lower] responses would always reflect greater generalization from a success [failure].

Results and Discussion

For each scenario, we took the trait judgment for a successful performance and subtracted off the trait judgment for a failed performance. Greater numbers imply greater perceived diagnosticity of the behavior for the trait.[1] We submitted these inferred diagnosticity scores to a 3 (perspective: actor, observer, or bystander) × 10 (scenario) mixed-model ANOVA. Only the first factor was varied between-subjects. The predicted main effect of perspective was significant, $F(2, 298) = 7.09$, $p < .001$, $\eta_p^2 = .05$. (See Table 2 for results by scenario).

We conducted a series of 2 (perspective) × 10 (scenario) repeated-measures ANOVAs to better understand the main effect of perspective. Providing evidence of the overblown implications effect, actors guessed that observers would be more reactive to performance events ($M = 4.04$, $SD = 1.95$) than those in the observer perspective condition were ($M = 3.07$, $SD = 1.88$), $F(1, 298) = 13.27$, $p < .001$, $\eta_p^2 = .04$. In other words, metaperceivers saw more diagnosticity in an upcoming performance, even though they had no contextual details to neglect (and no actual high-point or low-point on which to focus).

Did actors' metaperceptions identify greater diagnosticity in these behaviors because actors were considering being personally evaluated (as we have argued), or merely because they were making judgments about someone else's inferences? Providing support for the predicted account, bystanders did not think that observers would be particularly reactive ($M = 3.33$, $SD = 1.87$). That is, their own guesses about observers' inferences showed less evidence of an overblown implications effect than did actors', $F(1, 298) = 6.96$, $p = .009$, $\eta_p^2 = .02$. Instead, bystanders' guesses were fairly accurate, statistically indistinguishable from the observers', $F(1, 298) = 1.23$, $p = .27$, $\eta_p^2 < .01$.

Study 3

Although Study 2 showed that the OIE emerges in prospect—even when there is not a success or failure to ruminate upon—we did still ask people to imagine the actor (both) succeeding and failing at each performance. Study 3 offers a more conservative test by simply asking participants directly how diagnostic they would find each performance to be of the broader competency. If the prospect of being evaluated is sufficient to constrict metaperceivers' working trait definitions, then we should find metaperceivers estimate observers will find the performance more diagnostic than those considering the situations as observers would report.

Method

Participants. Two hundred fifteen undergraduates from the University of California, Berkeley, completed a lab session for course credit. Participants were randomly assigned to one of two perspective conditions: *actor* or *observer*.

Procedure. Participants considered the same scenarios used in Study 2. And like in Study 2, we told participants they should fully throw themselves into every simulation, to visualize the scene unfolding and to be attuned to what they would be thinking or feeling. However, unlike in Study 2, participants were not prompted to consider a success and a failure. Instead, they merely offered prospective diagnosticity judgments.

Consider the baking cookies situation, in which participants envision trying someone else's cookies or someone else trying participants' own cookies. For this situation, observers were asked, "After sampling their cookies, how much do you feel like you would have learned about whether or not the other person is a good cook?" Actors were asked, "After sampling your cookies, how much do you think the person would feel like they have learned about whether or not you are a good cook?" Each judgment was made on 11-point scales anchored at 0 (*not at all*) and 10 (*a great deal*). The 10 scenarios were presented in a random order.

Results and Discussion

To determine whether those simulating the perspective of observers would see less diagnosticity in actors' upcoming behavior than those considering the situations as actors guessed, we submitted participants' diagnosticity ratings to a 2 (perspective: actor or observer) × 10 (scenario) mixed-model ANOVA. Only the second factor was measured within-subjects. As hypothesized, there was a strong main effect of perspective, $F(1, 213) = 13.29$, $p < .001$, $\eta_p^2 = .06$. Those adopting the perspective of observers saw significantly less diagnosticity in actors' upcoming behavior ($M = 4.93$, $SD = 1.42$) than actors thought observers would ($M = 5.61$, $SD = 1.33$). Table 3 presents these results by scenario.

That actors and observers have prospective disagreement about behaviors' implications cannot be explained by actors' ruminating or focusing on their own recent performance, nor can it be explained by actors ignoring or discounting other relevant performances (again, because there were none to consider). Instead, these findings are consistent with our account that considering how others are evaluating the self causes working trait definitions to constrict around a considered source of evaluative apprehension. Study 4 probes this possibility more directly.

Study 4

We have argued that when actors consider being evaluated, their working trait definitions constrict. More specifically, the specific competency reflected in the performance context (e.g., skill at baking cookies) will begin to loom large in their sense of what observers will see as diagnostic of a broader competency (e.g., cooking ability). In Study 4, we once again had participants

[1] This analysis is equivalent to one in which we include feedback (success or failure) as a third within-subjects factor, but the present approach simplifies the description of the results.

Table 2
Mean Trait Inferences by Perspective for Prospectively Considered Successes or Failures on Each Trait-Relevant Behavior (Study 2)

Trait	Diagnosticity (success–failure)			Successful outcome			Failed outcome		
	Actor	Observer	Bystander	Actor	Observer	Bystander	Actor	Observer	Bystander
Mathematical ability	4.22 (2.52)	3.78 (2.85)	4.05 (2.60)	7.08 (1.58)	7.08 (1.61)	7.12 (1.43)	2.86 (1.89)	3.31 (2.03)	3.06 (1.85)
Social skills	3.51 (2.65)	3.90 (2.78)	3.45 (3.07)	6.84 (1.54)$_a$	7.32 (1.61)$_b$	6.84 (1.63)$_a$	3.33 (1.70)	3.43 (1.90)	3.39 (1.83)
Intelligence	3.18 (2.18)$_a$	1.82 (2.15)$_b$	2.36 (2.43)$_b$	7.37 (1.20)$_a$	6.83 (1.51)$_b$	6.96 (1.59)$_b$	4.19 (1.52)$_a$	5.01 (1.40)$_b$	4.60 (1.62)$_{a,b}$
Analytical thinking ability	3.30 (2.41)$_a$	2.30 (2.64)$_b$	2.77 (2.07)$_{a,b}$	7.42 (1.51)	7.21 (1.61)	7.24 (1.64)	4.12 (1.62)$_a$	4.92 (1.70)$_b$	4.48 (1.51)$_{a,b}$
Exploitativeness	4.89 (4.04)	3.80 (3.94)	4.07 (3.73)	8.00 (2.80)	7.75 (2.99)	7.89 (2.76)	3.11 (2.33)$_a$	3.95 (2.23)$_b$	3.82 (2.35)$_a$
Cooking ability	5.17 (2.67)	4.79 (3.24)	5.13 (2.72)	7.67 (1.16)	7.66 (1.44)	7.64 (1.31)	2.49 (2.19)	2.87 (2.42)	2.51 (1.81)
Driving ability	4.43 (2.85)	3.80 (2.72)	3.95 (2.67)	8.01 (1.61)	7.96 (1.53)	7.87 (1.44)	3.58 (1.92)$_a$	4.17 (2.04)$_b$	3.93 (1.83)$_{a,b}$
Self-control	3.89 (3.24)$_a$	3.03 (3.02)$_b$	2.09 (3.14)$_c$	7.70 (1.84)$_a$	7.56 (1.76)$_a$	6.32 (1.85)$_b$	3.81 (2.23)$_a$	4.53 (1.86)$_b$	4.23 (1.71)$_{a,b}$
Inconsiderateness	3.05 (4.14)$_a$	.09 (3.32)$_b$	1.64 (3.92)$_c$	7.43 (2.94)	6.69 (2.65)	6.95 (3.16)	4.38 (2.40)$_a$	6.59 (2.98)$_b$	5.31 (2.26)$_c$
Work ethic	4.72 (2.80)$_a$	3.37 (2.93)$_b$	3.76 (3.13)$_b$	8.34 (1.49)$_a$	7.67 (1.71)$_b$	7.69 (1.77)$_b$	3.63 (2.03)$_a$	4.30 (1.91)$_b$	3.94 (1.94)$_{a,b}$
Overall	4.04 (1.95)$_a$	3.07 (1.88)$_b$	3.33 (1.87)$_b$	7.59 (1.08)$_a$	7.37 (1.10)$_{a,b}$	7.25 (1.12)$_b$	3.55 (1.20)$_a$	4.31 (1.10)$_b$	3.93 (1.05)$_c$

Note. *SD*s are in parentheses. Means in the same row that are beneath the same subheading (diagnosticity, successful outcome, or failed outcome) that have different subscripted letters differ at the $p < .05$ level.

simulate performance contexts from the perspective of an actor or an observer.

We then aimed to measure whether actors could anticipate how much the displayed competency (compared with other specific, but nondisplayed competencies; e.g., making an omelet) would speak to the broader competency in the eyes of observers.

According to our theoretical account, metaperceivers should overestimate the relative importance of the displayed competency in speaking to the more general competency. An alternative hypothesis—one that could account for the results of our previous studies—is that actors simply think observers will draw broader implications from *all* performance behaviors, not merely the one they are considering being evaluated on. This is why it is important that the actors rank the importance of the displayed behavior compared with other nondisplayed behaviors. If the displayed competency looms large in metaperceivers' working trait definitions, it should "pop out" to actors, giving it a boost in the rankings. We conducted two pilot studies, described more fully in the online supplemental materials, to help with generating stimulus materials (Pilot Study 4a) and provide assurance that baseline perceptions of such materials make them appropriate for our main study (Pilot Study 4b):

Table 3
Mean Diagnosticity by Perspective for Each Trait-Relevant Behavior (Study 3)

Trait	Actor	Observer	*t*
Mathematical ability	5.17 (2.28)	5.96 (2.23)	−2.56*
Social skills	5.74 (2.20)	6.08 (2.16)	−1.13
Intelligence	5.02 (2.45)	4.69 (2.31)	.99
Analytical thinking ability	5.22 (2.24)	4.68 (2.25)	1.75†
Exploitativeness	5.60 (2.62)	4.66 (2.55)	2.64**
Cooking ability	6.43 (2.01)	5.58 (2.27)	2.90**
Driving ability	6.85 (2.36)	5.86 (2.58)	2.96**
Self-control	5.75 (2.64)	4.54 (2.30)	3.55***
Inconsiderateness	4.87 (2.68)	3.41 (2.54)	4.09***
Work ethic	5.48 (2.50)	3.82 (2.67)	4.71***
Overall	5.61 (1.33)	4.93 (1.42)	3.65***

Note. *SD*s are in parentheses. The *t* statistics come from independent-samples *t* tests with 213 *df* assessing whether the actor rating is greater than the observer rating.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Pilot Study 4a

We showed 151 Americans on MTurk specific competencies that directly related to the focal performance behaviors in our main study (e.g., baking cookies). Participants generated two specific performance behaviors that were "pretty different" from the specific competency in our main study but would "give just as much information" about an identified general competency (e.g., being a good cook). For each general competency, we chose four behaviors from the most frequently generated responses that seemed to best meet the above criteria (see Table 4 for a list of the related competencies by scenario).

Pilot Study 4b

It would be problematic if we unintentionally chose related (filler) behaviors that the self saw as more diagnostic of the general competency than hey believed someone else would. A new group of participants ($N = 182$, MTurk) first identified an acquaintance, just as participants in the main study would. Participants estimated whether that acquaintance would see each behavior (the four filler behaviors and one focal performance behavior) as more or less diagnostic of the relevant general competency than would the self. If anything, participants thought their acquaintance (compared with the self) would see *more* diagnosticity in the new filler behaviors than the focal performance behaviors. Given in the main study, we predict the opposite pattern (that estimates of an observing acquaintance's working trait definition will prioritize the focal performance behavior more than would the self's), this pilot study's results suggest that our main study's analyses will be especially conservative.

Method

Participants. Eight hundred seventeen Americans recruited from MTurk completed an online survey for payment. Participants

Table 4
Mean Diagnosticity Rankings by Perspective for Each Trait-Relevant Behavior (Study 4)

General competency	Specific competency	Related competencies	Specific competency ranking		
			Actor	Observer	*t*
Cooking ability	Baking cookies	Cooking a steak to appropriate tenderness, knowing how to use a range of cooking equipment, making an omelet, seasoning food correctly	2.33 (1.37)	2.22 (1.23)	1.14
Self-control	Refusing dessert when on a diet	Choosing to take stairs over the elevator, sticking to a set budget, studying for a test rather than going out, turning down a drink when must drive	3.02 (1.36)	2.86 (1.29)	1.61
Intelligence	Trivial Pursuit	Being good at mental math, being a good speller, having a broad vocabulary, having a good memory for distant events	2.22 (1.36)	2.05 (1.21)	1.76†
Exploitativeness	Forgetting to pay someone back	Asking someone for a lot of rides, eating at a potluck without bring anything, leaving before one's turn to buy a round of drinks, taking advantage of a store's return policy	3.03 (1.45)	2.79 (1.44)	2.19*
Driving ability	Parallel parking	Driving safely in rain or snow, keeping a good following distance, remaining calm during traffic, using turn signals properly	2.48 (1.51)	2.06 (1.43)	3.79***
Inconsiderateness	Forgetting another person's name	Being late to a meeting, forgetting to RSVP to an event, interrupt someone while they are talking, not holding the door open for someone behind you	2.39 (1.31)	1.92 (1.10)	5.25***
Overall			2.58 (.67)	2.32 (.59)	5.55***

Note. *SD*s are in parentheses. Ranks were reverse-coded so that higher rankings indicate greater relative diagnosticity. The *t* statistics come from independent-samples *t* tests with 717 *df* assessing whether the actor ranking is greater than the observer ranking.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

were randomly assigned to one of two perspective conditions: *actor* or *observer*. Ninety-eight participants failed an attention check. This left 719 participants for all analyses. The significance of the results reported below does not depend on excluding these participants.

Procedure. Participants considered six of the scenarios used in Studies 2 and 3. As in Studies 2 and 3, participants were asked to fully place themselves in each situation. Before these simulations began, participants thought of an acquaintance—someone they knew (each knew the other's name), but not particularly well ("an acquaintance instead of a friend"). Observers then considered the acquaintance engaging in the various performance behaviors. Actors instead considered the acquaintance observing them engage in those same performance behaviors. Following each scenario, observers were asked to rank five behaviors—the performance behavior just considered and four newly generated nonfocal behaviors—for how diagnostic they are of whether someone has the broader competency. More specifically, participants dragged and dropped the behaviors in order from least to most diagnostic. Actors were instead asked to guess how the observer they identified would answer the same question.

Results and Discussion

Did actors exaggerate the prominence of the displayed competency in observers' working trait definitions of the general competency? We first reverse-coded the rankings, so that higher numbers reflected behaviors that were seen as more central to the definition of the general competency (e.g., most important behavior was coded as 5, least important behavior was coded as 1). We submitted the performed behaviors' rankings to a 2 (perspective: actor or observer) × 6 (scenario) repeated-measures ANOVA, with only the second factor measured within-subjects. As expected, those considering their own upcoming performance behavior assumed that observers would have different working trait definitions than they actually did, $F(1, 717) = 30.78$, $p < .001$, $\eta_p^2 = .04$. More specifically, observers ranked the performance behavior as less important to the working trait definition ($M = 2.32$, $SD = .59$) than actors thought they would ($M = 2.58$, $SD = .67$). Table 4 presents these results by scenario.

These findings provide more direct evidence that metaperceivers' working trait definitions differ from what observers' work trait definitions actually are. Whereas participants in Studies 2 and 3 made judgments that were consistent with this conclusion, Study 4 showed this more directly by measuring the perceived centrality of the performance behavior compared with other behaviors that could also speak to the same general competency. This provides stronger evidence that metaperceivers' working trait definitions seem to constrict around the specific performance on which they consider being evaluated.

Study 5

In Study 5, we aimed to further test our account by *manipulating* actors' and observers' working trait definitions. We returned to the

lab to examine differences between how metaperceivers and observers made sense of an actual performance. However, in this case, we gave some actors and observers a manipulation designed to make salient the breadth of behaviors that could define a broader competency. Before asking people to consider the performance and its possible implications, we first asked some participants to list other behaviors—beyond those demonstrated in the performance context (cf. Savitsky et al., 2001, Study 3)—that could speak to the broader trait in question.

If actors' metaperceptions, compared with observers' social perceptions, operate under a constricted working trait definition, then this *broadening* manipulation should debias actors' subsequent metaperceptions. That is, by encouraging them to appreciate that the focal performance behaviors in this context constitute only a narrow sliver of what defines the broader trait, metaperceivers may come to appreciate what observers do spontaneously. If the intervention corrects for a mismatch in working trait definitions, then it should attenuate the overblown implications effect.

This intervention should work if actors' distortion involves constricted working trait definitions, but not if it merely involves performance focalism or actors' initial failure to realize the mitigating circumstances that would explain away their poor performance. By way of contrast, consider a "defocusing" manipulation used by Savitsky et al. (2001, Study 3). That intervention asked actors to consider what other factors could influence observers' judgments. Actors came to realize that observers would be less harsh on them once actors began to focus on factors like "the difficulty of the questions." In our intervention, we neither focus participants on other aspects of their performance nor on situational factors. Instead, we keep people from adopting a narrow definition of the broader quality they are judging by: (a) having participants themselves generate these additional behaviors (thereby maximizing the chance participants buy into these other behaviors' relevance), and (b) making sure participants complete this exercise before they even see our final impression measure (to make sure participants would approach the key measure under the potential influence of the manipulation).

Study 5 investigated the overblown implications effect in the domain of intelligence. Actors (contestants) took part in a mock game show and guessed how observers (audience members) would judge them. To provide information upon which baseline impressions could be based, actors began by answering a set of trivia questions. Based on this information, participants offered their baseline impressions of actors' intelligence: Observers offered social judgments of contestants, and actors offered metajudgments of how observers likely viewed them. We then engineered a success or failure. Those randomly assigned to complete the broadening manipulation listed other ways that people like the actors could display whether or not they were intelligent.

Readers may recognize that the present paradigm shares certain similarities with Ross, Amabile, and Steinmetz's (1977) classic quiz show paradigm, but it intentionally differs in two ways. First, Ross et al. (1977) had observers themselves generate the trivia questions. Whereas Ross et al. (1977) focused on how observers fail to appreciate their own structural (question-generating) advantage, we wanted to avoid this feature given its relevance to other mechanisms (Van Boven et al., 1999). We instead standardized the performance task (with experimenter-generated questions) and maintained full control over performance feedback. Second, we wanted to remove the observer from the live context. That is, when observers are focused on their own performance and the impressions they are making, they can fail to notice variability in others' performance (Gilovich, Kruger, & Medvec, 2002). Because actors do not understand that observers are distracted, they experience a failure of meta-insight. But crucially, such differences were found to disappear when observers were completely removed from the performance context. For this reason, we had observers watch videotaped recordings of actors' performance instead of having observers be present (and, thus, less likely to fully encode actors' performance; Gilovich et al., 2002).

Before testing whether expanding working trait definitions might keep actors from displaying the overblown implications effect, we wanted to make sure that the overblown implications effect would appear in this performance context. Toward this end, we conducted supplemental Study B. And indeed, the OIE emerged: Actors overestimated how much observers' impressions would shift in light of actors' performance. Furthermore, like in Study 1, these flawed metaperceptions were more reactive than actors' more inflexible self-perceptions, thus, meeting the stricter criterion for a failure of meta-insight. (See the online supplemental materials for a full description of the methods and results.)

Method

Participants and design. Two hundred ninety-six undergraduates from the University of California, Berkeley, completed a lab session for course credit. Actors were randomly assigned to one of four conditions in a 2 (performance: *success* or *failure*) × 2 (trait definition: *broadened* or *control*) full factorial design. Observers were yoked to a randomly selected actor. As with Study 1, when more than one observer was paired with an actor, we averaged the observers' responses for the purpose of analyses.

Procedure and materials. We describe actors' experience first. Observers observed their yoked actor's complete experience—both by learning the instructions actors received and watching the actors' performance on video.

Actors. Actors took part in the study individually. Upon arrival, actors were seated in front of a laptop. The experimenter informed actors (accurately) that they would be videotaped throughout the entire study so that a future participant could observe their performance. Actors were told that they would be answering a series of multiple-choice trivia questions. Along with providing each answer, actors explained aloud their rationale behind their selection. Actors were told that for each of the 10 questions they answered correctly, they would be given a ticket to enter into a lottery drawing for a $50 Amazon.com gift card. We included this ticket scheme so that—as we explain below—we could create a high-stakes moment for actors during the game.

Actors were presented 10 difficult trivia questions, one at a time, on the laptop. Each trivia question had two answer choices. As an example, one question asked: "Which city has the higher crime rate: Chicago or Detroit?" Actors read each question aloud, indicated their answer, and explained why they chose their answer. Actors completed all three steps out loud, so that their yoked observers would be able to observe the full process. After completing all 10 of the trivia questions and regardless of their actual performance, actors were informed that they had answered 7 out of the 10 questions correctly. Most likely because we did not indicate

which of the questions were supposedly answered correctly or not, no actor (or observer) expressed suspicion during a funnel debriefing about this feedback's credibility.

At this point, the experimenter gave actors seven lottery tickets for the seven questions they had supposedly answered correctly. Following this standardized initial feedback, actors completed the *baseline perception* measures of intelligence. Actors provided metaperceptions (guesses of how the observers would rate them). These measures are described in more detail below.

Next, the experimenter returned to explain the performance task. Actors were told,

> Now, you will answer one Trivial Pursuit question for *double or nothing*. That is, if you answer this question correctly, you will double your chances of winning with a total of 14 lottery tickets for the $50 Amazon.com gift card. But, if you answer incorrectly, you will lose all your tickets and be left with nothing. Again, please explain your reasoning *out loud* for the following question.

We standardized participants' feedback on the first round so that the implications of this final question would be equivalent for all participants. The final question was: "Which novel was published first: *To Kill a Mockingbird* or *The Catcher in the Rye*?"

Our goal was to make this final question feel especially high-stakes as a performance event. In addition to raising the stakes on this question (making it double or nothing), we upped the evaluative stakes by having the experimenter read the question, listen to the actor's reasoning and answer, and then provide verbal feedback. (In the initial round, these steps had been taken by the computer.) We were also careful to choose a question related to a topic with which our participants would be familiar (both books are staples on required reading lists), but not so familiar that they would clearly know the precise detail being asked (their publication dates). This allowed us—unbeknownst to participants—to randomly assign participants to learn they had (supposedly) answered this question correctly or incorrectly. Based on this randomly assigned feedback, participants either received seven more lottery tickets (*success* condition) or had their seven tickets taken away (*failure* condition). Again, no actor (or observer) expressed suspicion about this feedback's credibility.

Before completing the final perception measures, actors assigned to the *broadened* trait definition condition completed the definitional broadening manipulation (see below for additional details). Finally, all actors completed the *final perception* measures. These took the same form as the baseline perception measures—the metaperception measures of intelligence completed before the performance event. After completing these final measures, actors were debriefed and apologized to for the mild deception. They were informed that all participants had an equal chance of receiving the $50 prize.

Observers. Each observer was yoked to one actor. Observers had the same experience as actors, but as onlookers to the situation instead of as active participants. That is, they learned what instructions had been given to actors, but the observers then watched the actors perform on video instead of answering the questions themselves. Before the experimental session, research assistants clipped the full-length footage of the actors into two shorter videos to show observers.

The first video showed the actor answering the first 10 trivia questions and ended with the experimenter coming in to give the actor the seven tickets for the seven trivia questions that they had supposedly answered correctly. After watching this video, observers rated their baseline *social perceptions* of the actor's intelligence. The second video showed the experimenter reading the final question to the actor, the actor answering the question, and the experimenter providing the final feedback. To make sure that observers were not less reactive to the feedback than actors were merely because observers failed to notice these performance details (Gilovich et al., 2002), the computer instructions reiterated the performance outcome to observers before they made their final judgments.

As in the actor condition, some of the observers were assigned to complete the definitional broadening manipulation. In particular, those observers yoked to a *broadened* actor completed the definitional broadening manipulation themselves. Finally, all observers completed their final social perceptions of the actors' intelligence.

Definitional broadening manipulation. The goal of this manipulation was to expand these participants' working trait definition of intelligence to include additional behaviors other than the one that defined the current performance context. The instructions were:

> Before answering the next set of questions, we would like you to think beyond the tasks in this specific experiment and think about other contexts in which a student from University of California, Berkeley, like you could demonstrate that they are or are not intelligent. Please list 5 different ways that a student from University of California, Berkeley, could demonstrate intelligence or lack thereof.

Trait perceptions. The perception measure comprised five items that asked about the actor's intelligence. The *social perceptions* asked observers to judge the actors in light of their performance. The *metaperceptions* instructed actors to guess how observers would judge them in light of their performance. More specifically, actors saw the same prompt given to the observers, and they were asked to guess the observers' responses.

The first three questions asked participants to rate the actor as competent, intelligent, and knowledgeable on 9-point scales anchored at 1 (*not at all*) and 9 (*extremely*). The fourth item asked what score the actor would likely get on an IQ test. Although participants supplied their own numerical score, they were given the following guide in case they were unfamiliar with the standard IQ scale: "80–89 = below average; 90–109 = average; 110–119 = above average; 120–139 = gifted; >140 = genius." Finally, participants were asked into what percentile the actor's IQ fell in comparison with other undergraduates at their university.

Because our five items used different response scales, we first standardized responses. We calculated the grand mean and standard deviation of each item across both baseline and final meta and social perception judgments. By relying on these sample statistics when standardizing each measure, we preserved all effects of perception type (meta and social) and time. The metaperception ($\alpha = .93$) and social perception ($\alpha = .91$) intelligence composites all had good internal reliability.

Results and Discussion

Given our predictions that control actors (i.e., those who did not complete the definitional broadening intervention) would be unique in having a constricted working trait definition, we began

by defining a variable *constricted* that differentiates control actors (3) from those in the other three conditions: broadened actors (−1), control observers (−1), and broadened observers (−1). This distinguished those participants predicted to have a working trait definition of intelligence that focused on the specific type of behavior displayed in the performance context (control actors) from those who—because of the intervention or their baseline expanded perspective—should have a broader definition of the trait (those in the three other conditions). We then submitted the competence perception composite to a 2 (constricted) × 2 (performance: success or failure) × 2 (time: baseline or final) mixed-model ANOVA, with only the final factor measured within-subjects.

The predicted Constricted × Performance × Time interaction emerged, $F(1, 204) = 10.13$, $p = .002$, $\eta_p^2 = .05$ (see Figure 5). As expected, the metaperceptions of those in the control actor condition were more reactive to the performance feedback than were the perceptions of those in the other three conditions. That is, control actors showed more of an overblown implications effect than did those in the broadened actor condition, $F(1, 204) = 6.03$, $p = .01$, $\eta_p^2 = .03$. Furthermore, control actors' metaperceptions were more reactive than the social perceptions of observers, regardless of whether observers were in the control condition, $F(1, 204) = 4.05$, $p = .05$, $\eta_p^2 = .02$, or the broadened condition, $F(1, 204) = 11.54$, $p < .001$, $\eta_p^2 = .05$. The three other comparisons—broadened actor versus control observer, broadened actor versus broadened observer, control observer versus broadened observer—were all nonsignificant, Fs < 1.81, ps $> .18$. (See online supplemental materials for additional descriptive statistics for baseline and final perceptions.)

This shows that asking actors to remind themselves of other behaviors that could display their intelligence—thereby broadening their constricted working trait definitions to include additional behaviors that observers would not see (cf. Savitsky et al., 2001)—was sufficient to eliminate the overblown implications effect. Of course, one concern is that what mattered was not the content of what actors wrote, but the fact that they were momentarily distracted from ruminating about their recent success or failure. This alternative is unlikely for two reasons. First, Studies 2–4 showed that the overblown implications effect does not depend on such rumination (given its effects can be seen in prospect). Thus, an intervention that inadvertently eliminated this post-performance process would seemingly not be sufficient to eliminate the OIE. Second, a posttest conducted by Savitsky et al. (2001) showed that actors remained wrong about how harshly they would be judged for a failure even when actors completed an unrelated, though potentially distracting writing task (i.e., about their favorite restaurant or grocery store—not something that would expand their working trait definitions) in between their performance and their metaperception estimates. That said, Studies 6a–6c adopt a new, convergent approach to test unique predictions of the working trait definition account.

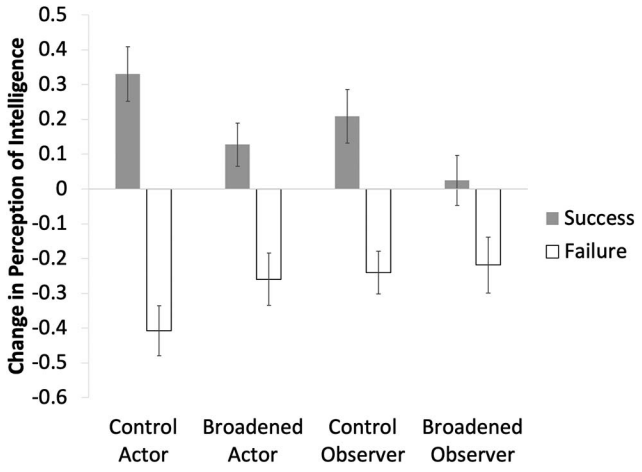

Figure 5. The change in perception (final–baseline) of actors' intelligence by role, performance, and working trait definition intervention (Study 5). The broadening intervention eliminated the overblown implications effect, as seen by the greater gap between the two bars for control actors compared with the gap between the two bars for the other three perspective-intervention combinations.

Study 6a

Our working trait definitions account argues that metaperceivers err in translating impressions of the specific performance domain on display (e.g., skill at answering trivia questions) to impressions of the more global competencies to which those specific skills might speak (e.g., intelligence). This argument differs from past research that identifies various reasons why metaperceivers misassess their performance (e.g., a disproportionate focus on their performance highlight or lowlights, a neglect of situational factors that made the performance very difficult and, thus, worthy of observer charity). Our proposal has caused us to focus on performances that cannot be so easily dismissed. That is, we argue that even when there is agreement in diagnosing the significance of specific performances, metaperceivers' errors should arise in translating impressions of specific skills into impressions on general competencies.

Study 6a examines this proposal directly. Much like in Study 3, those taking the perspective of actors or observers considered what actors would reveal in various prospective performance contexts. For example, actors considered taking homemade cookies to a party, whereas observers considered attending a party where they tried another's homemade cookies.

However, participants either estimated what would be learned about the specific competency in question (e.g., skill at baking cookies) or the general one that this could reflect (e.g., skill at being a cook). We expected to find that those considering these situations as metaperceivers would diverge more from observers when considering the general as opposed to the specific competencies. This would further implicate working trait definitions, which identify how people lean on specific skill perceptions when evaluating general competencies, in the overblown implications effect.

As an additional aim, Study 6a investigated the generality of the OIE. In the typical social judgment study, people judge strangers. For example, when Jones and Harris (1967) examined whether observers drew inferences about actors' true Castro attitudes even when actors had been forced to write a pro- or anti-Castro essay, the experimenters did not conduct this study among good friends. Such decisions tend to be made both to standardize the targets being judged and to make sure that everything else one knows

about a target does not overwhelm the influence of the specific causal effect being studied. Despite this typical rationale, Study 6a probed the robustness of the OIE by having people consider a specific friend as an actor or an observer. If the OIE emerges in this context, we would be more confident that meta-insight emerges not merely during early impressions but even within closer relationships.

Method

Participants and design. Eight hundred one Americans recruited from MTurk completed the study for payment. Participants were randomly assigned to one of four conditions in a 2 (perspective: actor or observer) × 2 (competency: general or specific) full-factorial design.

Procedure. Participants started by first considering a friend. Those in the actor conditions were told, "To start, think of someone specific who would consider you a friend. This person should be someone who thinks of you as more than an acquaintance but not quite a best friend. Please write the initials of the person who thinks of you as a friend." Those in the observer condition were given symmetric instructions: "To start, think of a specific friend. This person should be someone who is more than an acquaintance but not quite a best friend. Please write the initials of the friend below."

Participants then received instructions similar to those used in Study 3. They learned they would consider seven briefly described scenarios. As in Study 3, participants were asked to fully place themselves in those situations. We used the scenarios from Study 3 except with three changes. First, actors considered being observed by the friend they identified, and observers considered observing the friend. Second, we did not include three scenarios, some of which could not be fully naturally adapted to a context involving a good friend: conversing with a stranger, leaving work at an unusual time, and splitting the bill. Third, although the general competencies judged in Study 3 were again used in our general condition, participants in our specific conditions considered narrower competencies that corresponded more closely to the specific behaviors described. Consider the situation in which observers imagine trying a friend's cookies. Those in the observer general condition indicated, "After sampling their cookies, how much do you feel like you would have learned about whether or not [FRIEND'S INITIALS] is a good cook?" Those in the observer specific condition would instead answer, ". . . is good at baking cookies?"

As in Study 3, each judgment was made on 11-point scales anchored at 0 (*not at all*) and 10 (*a great deal*). Each scenario described a different behavior that spoke to a different general and specific trait or competency. The seven scenarios were presented in a random order.

Results and Discussion

To probe the error specificity of the OIE, we submitted participants' diagnosticity ratings to a 2 (perspective: actor or observer) × 2 (competency: general or specific) × 7 (scenario) mixed-model ANOVA. Only the final factor was measured within-subjects. As hypothesized, there was a significant Perspective × Competency interaction, $F(1, 797) = 4.42$, $p = .04$, $\eta_p^2 = .01$. Table 5 summarizes the results by scenario.

When considering the general trait, metaperceivers and observers had different perspectives. Namely, observers saw less diagnosticity in actors' upcoming behavior ($M = 5.37$, $SD = 1.87$) than actors thought observers would ($M = 5.86$, $SD = 1.51$), $F(1, 797) = 10.64$, $p = .001$, $\eta_p^2 = .01$. This replicates Study 3, but in the context of existing relationships.

However, illustrating that the OIE reflects a difference in how metaperceivers and observers translate specific impressions into more general ones, the difference found in the general conditions went away in the specific conditions. More specifically, actors knew how much diagnosticity ($M = 6.87$, $SD = 1.34$) observers would report ($M = 6.84$, $SD = 1.29$) when considering what the one-off performance would reveal about the specific competency, $F < 1$.

Of course, this is not to say that actors and observers will always agree on what a single performance says about the matching, specific competencies. After all, some of the past research on meta-insight has shown that observers are surprisingly (at least in the eyes of metaperceivers) quick to recognize when actors' performance is limited not by actors' own competence but by features

Table 5
Mean Diagnosticity by Perspective and Competency Specificity for Each Trait-Relevant Behavior (Study 6a)

		General competency			Specific competency		
General competency	Specific competency	Actor	Observer	*t*	Actor	Observer	*t*
Intelligence	Trivial Pursuit ability	5.94 (2.34)	5.47 (2.56)	1.96†	6.32 (2.15)	6.32 (2.17)	−.04
Analytical thinking ability	Chess-playing ability	5.64 (2.44)	5.54 (2.61)	.39	6.34 (2.23)	6.69 (2.16)	−1.59
Exploitativeness	Ability to remember to pay people back	5.81 (3.16)	4.88 (3.14)	3.01**	7.65 (2.42)	6.97 (2.38)	2.78**
Cooking ability	Ability to bake cookies	7.00 (1.96)	6.39 (2.40)	2.84**	7.79 (1.84)	7.70 (1.83)	.53
Driving ability	Parallel parking ability	6.38 (2.60)	5.59 (2.87)	2.91**	7.56 (2.27)	7.51 (2.09)	.20
Self-control	Ability to refuse complimentary desserts	6.07 (2.42)	5.40 (2.73)	2.64**	6.24 (2.39)	6.23 (2.61)	.03
Inconsiderateness	Ability to remember new people's names	4.20 (2.78)	4.34 (2.98)	−.52	6.21 (2.23)	6.43 (2.27)	−.96
Overall		5.86 (1.51)	5.37 (1.87)	2.93**	6.87 (1.34)	6.84 (1.29)	.26

Note. *SD*s are in parentheses.
$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$.

of the performance context. The overblown implications effect identifies a different reason why metaperceptions and social perceptions may diverge. Even when actors know how much observers think actions are diagnostic of an actor's skill, actors think those implications are more general than they actually are. That is, everyone can agree that a single instance of a person baking cookies does speak to how well that person bakes cookies. However, actors mistakenly think that observers will go further by taking that single competency and assuming it speaks especially clearly to how good of a cook the actor is. In short, actors overblow the implications of their own performance.

Study 6b

Whereas Study 6a tested for a specific type of error that only the definitional focalism (but not the performance focalism) account would anticipate, Study 6b probed for a more specific prediction that would indicate the role of constricted working trait definitions. This was achieved by using a design that was quite similar to Study 6a, but one in which the same participants indicated how diagnostic the performance behavior would be of the specific and the more general competency. First, we expected to replicate Study 6a's finding that metaperceivers would be wrong about how much diagnosticity observers saw in an upcoming performance for a general competency, but not necessarily for a specific competency. Second, and as a new test, we expected that metaperceivers' diagnosticity judgments of the specific competency would predict diagnosticity judgments of the general competency (more strongly than would observers'). That is, with a constricted working trait definition, metaperceivers' judgments about the specific competency should have more direct implications for the general competency.

Method

Participants and design. Seven hundred seventy-five Americans recruited from MTurk completed the study for payment. All were randomly assigned to one of two conditions in a 2 (perspective: actor or observer) × 2 (competency: general or specific) mixed design. Only the first factor was manipulated between-subjects. Fifty-nine participants failed an attention check. This left 716 participants for all analyses. The significance of the results reported below does not depend on excluding these participants.

Procedure. The procedure was similar to that of Study 6a, except for the following changes. First, instead of having participants think of a friend, we asked them to think of an acquaintance (using the same instructions as in Study 4). Participants supplied this person's initials or name, which was then piped into the relevant measures. Second, participants completed questions about how much they thought they would learn (observer condition) or how much they thought the other person would learn (actor condition) about both the general and specific competencies. Responses were made on an 11-point scale anchored at 0 (*not much at all*) and 10 (*a great deal*). We varied whether participants always completed the general or the specific measures first.

Results

We began by testing whether we could replicate Study 6a's key result in a within-subjects design. We started by using a similar repeated-measures ANOVA to that used before. We submitted the diagnosticity ratings to a 2 (perspective: actor or observer) × 2 (order: general first or specific first) × 2 (competency: general or specific) × 7 (scenario) mixed model. Only the first two factors were manipulated between subjects.

As expected, the Perspective × Competency interaction was significant, $F(1, 712) = 14.13$, $p < .001$, $\eta_p^2 = .02$. Furthermore, this interaction did not depend on the order in which participants completed the general and specific measures, $F(1, 712) = 1.18$, $\eta_p^2 < .01$. When considering the broader competency, metaperceivers demonstrated the overblown implications effect: Observers saw less diagnosticity in the upcoming performance behaviors ($M = 5.61$, $SE = .09$) than actors thought they would ($M = 5.93$, $SE = .09$), $F(1, 712) = 6.02$, $p = .01$, $\eta_p^2 = .01$. This asymmetry disappeared when considering the meaning of the performance behavior for assessment of the more specific competency. In this case, observers saw as much diagnosticity ($M = 7.28$, $SE = .08$) as actors thought they would ($M = 7.18$, $SE = .08$), $F < 1$ (see Figure 6A). Once again, even when metaperceivers understand the narrow implications of an upcoming performance, they overblew its implications.

Next, we proceeded to our more specific test: whether metaperceivers show additional evidence of having constricted working trait definitions. We constructed a random-slope, random-intercept model predicting the diagnosticity ratings. We included three Level-1 fixed effects: perspective (1: actor, −1: observer), order (1: specific first, −1: broad first), and the perceived diagnosticity for the specific competency (standardized within each behavior). These fixed effects were nested within scenario. This permitted the effects of the predictors to vary by scenario (random slopes) while accounting for differences between scenarios in how much the behavior in question spoke to that particular broader competency (random intercept). The model included the interaction terms that could be made from these three predictors (three two-way and one three-way interaction) as well as a random effect of participant (to account for the nonindependence of their multiple responses).

Even with the additional terms, we continued to observe a positive main effect of perspective, $B = .19$, $SE = .07$, $t(2.82)$, $p = .005$. This merely reflects the continued existence of the overblown implications effect, even with the perceived diagnosticity of the performance behavior for the specific competency statistically controlled. More centrally, we observed a significant Perspective × Specific Competency interaction, $B = .08$, $SE = .03$, $t(2048.73) = 2.25$, $p = .024$. Observers' beliefs about what they would learn about the actors' narrow competency translated into what they believed they would learn about the actors' more general competency, $B = 1.09$, $SE = .10$, $t = 10.79$, $p < .001$. This demonstrates that everyone agrees that the specific behaviors do speak to the general competencies, a validation of sorts of our materials.

However, actors assumed that observers would make this leap more strongly, $B = 1.24$, $SE = .10$, $t = 12.47$, $p < .001$. This reflects the constricted nature of metaperceivers' working trait definitions.

Decomposing the interaction the other way gives more insight into how metaperceivers' constricted working trait definitions lead to the overblown implications effect. For participants who didn't see much information in the performance for the specific competency (−1 *SD*), actors knew how much information observers

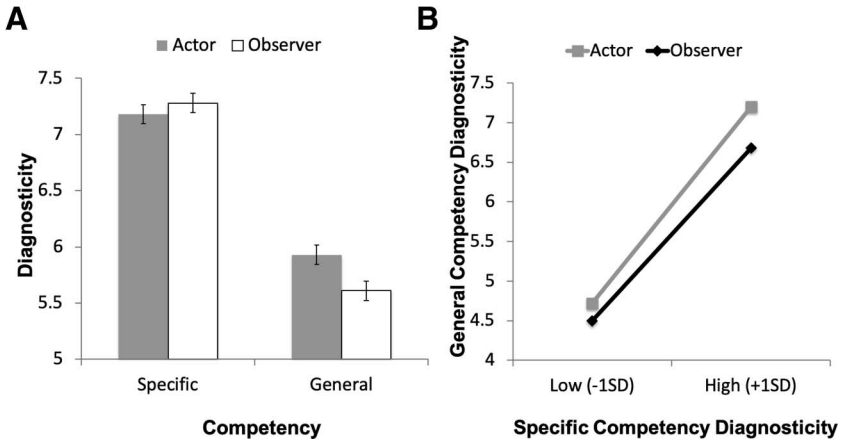

Figure 6. General and specific competency diagnosticity ratings (Study 6b). Panel A displays by perspective condition perceived diagnosticity of actors' performance behavior for the specific, displayed competency and the relevant general competency. Panel B displays predicted perceived diagnosticity of the performance behavior for the relevant general competency by the perspective manipulation and the perceived diagnosticity of the specific, displayed competency (at 1 *SD* above and below the mean).

would see regarding the general competency (*M*s = 5.72 vs. 5.50), $t = 1.44$. In contrast, when participants saw clear implications of the performance for the specific competency (1 *SD*), actors now overestimated how eager observers would be to generalize this to the more general competency (*M*s = 8.20 vs. 7.68), $t = 3.61$ (see Figure 6B). It is only when there is a perception that the performance behavior truly does speak to the specific competency that the conditions should be in place for metaperceivers' constricted working trait definitions to lead them to overgeneralize the implications to the more general competency.

Study 6c

Study 6c was designed both to extend the overblown implications effect and provide another stringent test of the working trait definition account. We have repeatedly shown that metaperceivers assume observers see wider implications for actors' performance than observers actually do. However, what does it actually mean for metaperceivers to be wrong about observers' impressions of actors' general competencies? Presumably this means that observers would be assumed to later use these general impressions to guide new, relevant specific impressions.

Consider someone who blows the game-winning question of Trivial Pursuit. Metaperceivers may be correct in how much this failure reveals in others' eyes the actor's skill at Trivial Pursuit. However, where the player may err is in how much observers now see them as unintelligent. Having decided that their apparent lack of intelligence was on full display, such actors may then think observers will be pessimistic about new indicators of intelligence that they subsequently consider.

In Study 6c, participants considered performance situations from the perspective of actors or observers. We asked observers what they would conclude (and actors, what they thought observers would conclude) if the performance was a success or a failure (much like in Study 2). We were interested in how judgments might differ for the specific competency on display and adjacent competencies that speak to the same general competency. If metaperceivers had already generalized from the specific competency (e.g., Trivial pursuit skill) to the general competency (e.g., intelligence) because of definitional focalism, then they should assume that observers would then draw strong inferences about adjacent competencies as well (e.g., spelling ability, mental math skills, etc.). However, given that observers should not have the same constricted working trait definitions, they should generalize less to these adjacent competencies than metaperceivers would think.

Method

Participants and design. Eight hundred fifty-one Americans recruited from MTurk completed the study for payment. We used a 2 (perspective: actor or observer) × 2 (performance: success or failure) × 2 (competency: displayed or related) mixed design. Only perspective was manipulated between subjects. Performance was randomly assigned at the level of scenario for each participant—that is, whether participants considered the success or failure version of each scenario did not systematically covary with which version they saw on any of the other five scenarios. Competency was measured within-subjects. One hundred participants failed an attention check. This left 751 participants for all analyses. The significance of the results reported below does not depend on excluding these participants.

Procedure. After identifying a specific acquaintance, participants learned they would estimate how that acquaintance would judge them, or estimated how they would judge the acquaintance. Participants considered the six situations used in Study 4, but the procedure differed in two key ways. First, participants considered that a performance was a success or a failure. Actors guessed how the acquaintance would then judge the self. Observers indicated how they would judge the (actor) acquaintance. For four scenarios, the prompt asked "how good" the actor was (or would be judged to be) "at each of the following." For two scenarios, the prompt asked "how likely" the actor was (or would be judged to be) "to do each of the following."

Second, instead of making judgments about both general competencies (e.g., intelligence) and specific competencies (e.g., skill at Trivial Pursuit), participants only considered five specific competencies (one displayed and four related). The *displayed* competency was the specific competency that was just on display (e.g., skill at Trivial Pursuit). The other four competencies were those identified by Pilot Study 4b: specific skills or behaviors that were also related to the same general competency (e.g., mental math, memory for distant events, having a broad vocabulary, and spelling). Participants responded to all five specific competencies on 10-point scales, anchored at 1 (*very bad* or *very unlikely*) and 10 (*very good* or *very likely)*. Participants always rated the displayed competency before the four related competencies. Some ratings were reverse scored, so that higher numbers would always match the direction implied by a performance success. For each scenario, we averaged ratings of these four specific competencies to create a *related* competency composite.

Results and Discussion

We began by testing for evidence of the error specificity predictions of the overblown implications effect—that observers would be more reluctant to generalize the actor's performance as evidence of other related skills and deficiencies than actors would assume. Toward this end, we used a series of random-slope, random-intercept models. In each, we included fixed effects of the perspective manipulation (actor: 1, observer: −1) and performance (success: 1, failure: −1). These effects were nested within scenario. This permitted the effects to vary across scenarios (random-slope) while also accounting for variation in the competency ratings across scenarios (random-intercept). We also included the Perspective × Performance interaction. To account for nonindependence of observations, we included a random effect of participant.

In our first model, we predicted the ratings of displayed competency minus the ratings of the related competencies. In this way, more extreme positive and negative values reflect a *stunted* tendency to overblow the implications of success or failure, respectively, to adjacent competencies. And as expected, we observed a Perspective × Performance interaction, $B = -0.11$, $SE = 0.03$, $t(4487.34) = 3.42$, $p = 001$. The negative beta reflects that metaperceivers thought that observers would extend their evaluations to related competencies more than they did (see Table 6 for results by scenario).

In the next pair of models, we used the same model to predict the displayed competency and the related competencies separately. When trying to explain variation in judgments of the displayed competency, we did not observe a Perspective × Performance interaction, $B = -0.03$, $SE = 0.04$, $t < 1$. Conceptually replicating Study 6b, metaperceivers were accurate in anticipating how observers would judge their displayed competencies following success versus failure. However, when predicting other related competencies, the Perspective × Performance interaction emerged, $B = 0.07$, $SE = 0.03$, $t(4235.35) = 2.62$, $p = .009$. Observers did not generalize the implications of the specific performance to adjacent behaviors as much as metaperceivers thought they would.

Given all participants considered both the displayed and the related competencies, we conduct a more nuanced test of the overblown implications effect that is analogous to that used in Study 6b. For this analysis, first we standardized the specific displayed competency rating separately for each of the six scenarios. We then included this variable as a fixed effect (also nested within scenario), as well as all interactions that could be made with our other two fixed effects (perspective and performance), predicting the related competencies. Of key interest, the Perspective × Displayed Competency interaction was significant, $B = 0.12$, $SE = 0.03$, $t(4303.72) = 4.16$, $p < .001$. Although observers' impressions of actors' specific displayed competencies did generalize to perceptions of their related competencies ($B = 1.15$), metaperceivers overestimated how much observers would engage in such generalization ($B = 1.38$). Even when metaperceivers knew how observers would make narrow sense of an upcoming performance and the specific competencies to which they spoke, their constricted working trait definitions encouraged a direct overgeneralization to subsequently considered related competencies.

The Moderating Role of Public Self-Consciousness

By our theoretical reasoning, metaperceivers' working trait definitions constrict around the performance domain because of the threat of considering how one would be evaluated. This logic suggests that there should be individual variability in the OIE that

Table 6
Mean Inferences About the Specific Competencies (Displayed and Related) by Performance and Perspective Manipulations (Study 6c)

	Displayed competency				Related competencies			
	Success		Failure		Success		Failure	
General competency	Actor	Observer	Actor	Observer	Actor	Observer	Actor	Observer
Cooking ability	8.07 (2.02)	8.07 (2.09)	3.64 (2.79)	3.60 (2.72)	6.82 (1.68)	6.82 (1.78)	$4.34\ (2.27)_a$	$4.91\ (2.15)_b$
Intelligence	7.77 (1.93)	7.68 (1.85)	4.94 (2.13)	5.20 (1.97)	7.04 (1.81)	7.05 (1.58)	5.46 (1.81)	5.75 (1.72)
Self-control	$7.29\ (2.54)_a$	$7.87\ (2.43)_b$	4.29 (2.78)	4.52 (2.52)	6.66 (1.96)	6.85 (1.82)	5.67 (2.04)	5.85 (1.99)
Exploitativeness	7.81 (2.78)	7.78 (2.50)	4.07 (3.08)	4.15 (2.79)	7.70 (2.19)	7.31 (2.10)	5.55 (2.59)	5.73 (2.20)
Driving ability	$7.82\ (2.61)_a$	$8.37\ (1.93)_b$	4.01 (2.72)	3.79 (2.38)	7.22 (1.79)	7.25 (1.88)	6.13 (1.92)	6.11 (1.85)
Inconsiderateness	7.53 (2.35)	7.42 (2.25)	4.22 (2.44)	4.04 (2.38)	$7.09\ (1.51)_a$	$6.74\ (1.57)_b$	6.24 (1.64)	6.29 (1.59)
Overall	7.71 (2.40)	7.86 (2.20)	4.20 (2.69)	4.20 (2.53)	7.08 (1.86)	6.99 (1.80)	5.58 (2.15)	5.77 (1.98)

Note. Ratings indicate means and *SD*s (in parentheses). Adjacent means in the same cluster (e.g., displayed competency successes) that have different subscripts differ at the $p < .05$ level.

is tied to individual variability in how concerned people are with evaluation. Public self-consciousness (PSC) has been shown to predict anxiety around being evaluated (e.g., Hope & Heimberg, 1988; Turner, Carver, Scheier, & Ickes, 1978). Given that the OIE was often measured using higher-order interactions, we knew that we would not have the power to consistently detect the even higher-order interactions with PSC. Instead, we measured PSC—using Fenigstein, Scheier, and Buss's (1975) PSC scale—in every relevant study and conducted a cross-study meta-analysis. Providing additional convergent evidence for our experimental logic, we found that the *overblown implications effect* grew stronger as actors' public self-consciousness increased, $2.57 <$ Stouffer's $Zs < 2.94$, $.003 < ps < .011$ (Study 1: $Z_{Global} = 2.55$, $Z_{Feedback\text{-}Informed} = 2.20$; Study 5: $Z = 1.54$; Study 6c: $Z = .87$; supplemental Study A: $Z_{Global} = .04$, $Z_{Feedback\text{-}Informed} = -.43$; supplemental Study B: $Z = 1.57$; see online supplemental materials for additional details).[2]

General Discussion

People care how others view them. However, without direct access to others' perceptions, understanding how we are perceived entails guesswork. Across eight studies, we provided evidence for an *overblown implications effect*. Actors see their own performance as having more evaluative impact on observers than it actually does. By introducing the construct of working trait definitions, we were able to localize this error to a difference in how metaperceivers and observers were defining the broader competencies (partially) on display.

Contributions of the Present Research

Psychologists have spent decades studying many psychological mechanisms that lead to accuracy and error in self and social perceptions (Dunning, 2005; Funder, 1987; Vazire, 2010). Although metajudgments are a natural extension of this research tradition, research in this area is in an earlier stage of development. Just as there are many reasons why people's self and social judgments are accurate or inaccurate, the same is no doubt true of metajudgment. Instead of focusing on actors' failure to understand how observers make sense of their specific performances (as true successes or failures), we have focused on actors' incorrect beliefs about how strongly observers will generalize these local assessments to broader impressions. We now review our findings through the lens of considering how the overblown implications effect—and our working trait definitions account, in particular—contributes to a more complete understanding of meta-insight.

Definitional focalism (vs. performance focalism). Actors erred because they assumed observers had different working trait definitions than they actually did, not because they assumed that observers focused on a narrower portion of their performance (e.g., one mistake in a sea of triumphs). Consider four ways we made this case. First, metaperceivers were still inaccurate when considering how observers would interpret the narrow portion of their performance on which they did well or poorly (Study 1). Second, even when performance occurred "in a vacuum"—meaning there were not highlights or lowlights to selectively focus on or situational factors to neglect—we still observed the OIE (Studies 2 and 3). Third, when considering how they would be judged on an upcoming performance, actors overestimated the prominence of that performance behavior in observers' working trait definitions of the broader competency (Study 4). Fourth, by experimentally manipulating working trait definitions to broaden them, we debiased metaperceptions while leaving social perceptions untouched (Study 5).

Overblowing (vs. misassessing) performance. The working trait definition account localizes the error that metaperceivers make. More specifically, we argued and demonstrated that actors will often be quite accurate about how observers will interpret the narrow meaning of their behavior (Study 6a–6c). Downtown drivers who gracefully glide into open metered spots are seen to be good parallel parkers, and those drivers likely know it. That is, the OIE need not reflect that observers excuse a one-off performance as nondiagnostic of skill at that specific performance task.

Instead, the OIE is rooted in the distorting influence of working trait definitions that are used to understand the broader implications of narrow competencies. Successful parallel parkers will be mistaken in thinking their full driving skills are on display. This explains why metaperceivers who know how observers view them on specific competencies still exaggerate how much observers will shift their impressions of broader competencies (Studies 6a and 6b) and related skills (Study 6c). More detailed process evidence showed metaperceivers leaned too heavily on their (accurate) impressions of how observers judged specific competencies to draw inferences about those broader competencies (Study 6b) and other specific skills that such competencies imply (Study 6c).

Successes as well as failures. We are not the first to demonstrate that metaperceivers underestimate observers' charitable response to blunders. However, our working trait definition account led us to predict that actors would overestimate how positively observers would respond to successes. This *extremity* bias (overblowing performances' implications) contrasts with a *directional* bias (overestimating observers' harshness) that several previously articulated mechanistic accounts would anticipate. A cross-study meta-analysis provided clear support that actors overestimated the negative implications observers would see in failures, $2.84 <$ Stouffer's $Zs < 3.08$, $.002 < ps < .005$ (Study 1: $Z_{Global} = 3.04$, $Z_{Feedback\text{-}Informed} = 1.67$; Study 2: $Z = 2.94$, Study 5: $Z = 1.94$, supplemental Study A: $Z_{Global} = .01$, $Z_{Feedback\text{-}Informed} = 1.91$; supplemental Study B: $Z = -1.58$). However, they also overblew successes, $4.67 <$ Stouffer's $Zs < 5.70$, $ps < .00001$ (Study 1: $Z_{Global} = 1.31$, $Z_{Feedback\text{-}Informed} = 1.70$; Study 2: $Z = 2.81$, Study 5: $Z = 1.06$, supplemental Study A: $Z_{Global} = .88$, $Z_{Feedback\text{-}Informed} = 2.80$; supplemental Study B: $Z = 4.38$).

Reconciling the Overblown Implications Effect With Other Related Research

Actor-observer effect. At first glance, the OIE might seem inconsistent with the actor-observer effect (Jones & Nisbett, 1971; but see Malle, 2006)—the tendency for observers to make more

[2] Study 1 had both global and feedback-informed final trait measures. This means that for the purpose of any cross-study meta-analysis including this study (or supplemental Study A), there are two ways to select which measure to use from this study (global or feedback-informed). The range of meta-analytic results reflects that the results are robust to all specifications.

dispositional inferences than actors. However, crucially, we do not test whether actors and observers make different attributions for actors' actions. Instead, we examine the accuracy of actors' guesses about how observers view them. But might the OIE reflect actors' sense that observers commit the fundamental attribution error more strongly than they actually do (Van Boven et al., 1999)?

For three reasons, we would not characterize the OIE as a false belief that observers embrace dispositional (instead of situational) explanations for others' behavior. First, actors' misestimates of observers' perceptions stemmed from the narrowness with which metaperceivers thought about the trait category, not actors' explanations for the behavior. Only our preferred account can explain the influence of the definitional broadening intervention (Study 5) or why actors and observers agreed on the narrow implications of a behavior (Studies 6a–6c). Disagreement about whether a behavior actually just reflected situational influence would have produced differences on the narrow impressions as well. Second, and relatedly, in some of our studies (Studies 2–4, 6a–6c), the behavioral contexts were described in a vacuum—that is, without information about how the situation may affect performance success. Thus, it is hard to imagine what situational contexts observers would have been relying upon that actors would have neglected, yet such hypothetical scenarios were sufficient (or perhaps even ideal) for localizing effects to definitional misunderstandings. Third, if the OIE were merely another example of people exaggerating how much others display the fundamental attribution error (Pronin et al., 2002; Van Boven et al., 2003), then participants in the bystander condition (Study 2) should also have overestimated the extent to which observers would draw inferences from actors' behavior, but in fact, such bystanders did not.

Looking-glass self. Although we did not give much attention to actors' self-perceptions in Study 1, some might be surprised that they did not show the same evidence of the OIE that actors' metaperceptions did. This might seem inconsistent with the literature on the looking-glass self, the idea that self-views derive from how (people believe) others view them (Cooley, 1902; Tice, 1992; Tice & Wallace, 2003). Although it is the case that self-perceptions did not move to the same extent as metaperceptions, they still did show sizable shifts in light of recent performance. Furthermore, the correlation between the change in meta- and self-perception was relatively strong ($r = .44$). We, of course, cannot say whether this relationship between meta- and self-perceptions was causal; nonetheless, this finding illustrates how the overblown implications effect should not be interpreted to exist instead of, but rather on top of, that which results from the looking-glass self.

Questions for Future Research

Can observers' reactions mitigate the OIE? In an effort to isolate the hypothesized OIE, we did not have observers directly interact with actors. Previous research has found that participants in an interaction become sufficiently focused on their own behavior that they can fail to notice variability in their interaction partners' performance (Gilovich et al., 2002). For this reason, our in-lab paradigms preserved the subtleties of actors' behavior (by presenting them on video or computer-mediated communication) but did not place the observer into the live context in which self-presentational concerns could distract.

However, had the observer been present, then the actor would have an additional source of information—observers' verbal and nonverbal reactions—when forming metaperceptions. Though even when observers are present, observers are notoriously hard to read. For example, in the classic spotlight effect studies (Gilovich et al., 2002), the live presence of observers did not help actors realize they were not the clear focus of attention. Furthermore, many performance behaviors occur not in dyadic contexts (in which only one observer's reactions could be monitored), but in front of an audience. As the negativity dominance literature and many a professor's experience teaching suggests, that one scornful audience member can loom large in our attentional field (e.g., Hansen & Hansen, 1988; Pinkham, Griffin, Baron, Sasson, & Gur, 2010). Such attentional biases could distort metaperceptions of the audience as a whole.

Furthermore, there is another hint that by leaning on uninvolved observers, our tests may have been especially conservative. Campbell and Fehr (1990) had actors guess how interaction partners and uninvolved observers viewed them. Actors did not distinguish between these two groups in their metaperceptions, but uninvolved observers were actually harsher in their assessments of the actors than were the interaction partners. At least when it comes to anticipating how others will respond to one's poor performance, this suggests that the OIE may be even stronger when live observers are present. Whether the OIE is ultimately mitigated or exaggerated when the observer is present is a worthy question for future research.

Does the OIE extend to naturalistic performances and naturalistic feedback? In some of our studies, the fact that participants had succeeded or failed was made explicit. For example, interviewees learned that their interview responses caused them to be selected or excluded (Study 1), and trivia contestants were told that they had gotten the double-or-nothing question right or wrong (Study 5). On the one hand, this raises a worry that constricted working trait definitions stemmed not from actors considering being personally evaluated but instead from the unusualness of receiving blunt feedback. However, such a characterization would not be consistent with our studies that merely asked people to consider the conclusions that would follow from things going well or poorly (Studies 2, 6c) or that included no feedback whatsoever (Studies 3–4, 6a and 6b).

In our everyday lives, sometimes performance quality can be murkier, in part because the feedback we get is circumspect. That is, when a student makes a comment, and the professor fills the silence in the classroom with a slowly delivered "That's interesting," there can be ambiguity in what was meant. However, the existence of the OIE does not suggest that performance feedback is always clear. It simply means that the perceived implications of the performance—even when those implications are misassessed—get overblown. Such miassessment in itself can be a cause for additional error. If, for example, the student did not realize his observers saw his comment as (mildly) embarrassing, he may assume they took it as strongly revealing of his superior intellect.

Another property of these engineered successes and failures is that they could have created a demand effect of some sort—that is, perhaps metaperceivers felt that this feedback *should* be incorporated into their judgments. However, this feedback-driven demand effect explanation should not apply in Studies 3–4 and 6a and 6b, in which we examine prospective judgments of diagnosticity rather

than judgments that followed the feedback. Furthermore, it is hard to imagine that the higher correlations observed for metaperceivers' (as opposed to observers') displayed competency judgments and their general (Study 6b) or related competency judgments (Study 6c) could be the result of demand. And even for those studies that examine responses to feedback—real (Studies 1, 5) or hypothetical (Studies 2, 6c)—it is unclear whether or why a demand effect would apply only to metaperceivers and not observers asked to respond in light of the *same* performance feedback.

When might others' investment in actors' performance produce or temper the OIE? We have argued that the constricted nature of actors' working trait definitions stems from the evaluative threat posed by the performance situation. However, this also suggests that there may be times in which bystanders show the OIE as well. For example, parents who feel highly invested in their child's soccer game or spelling bee performance may feel empathic evaluative apprehension as their child is on stage. As such, it may feel that their child's image as athletic or intelligent is on clear display to observers.

Though when the observers themselves have a personal stake in actors' performance, this may actually cause them to make the sort of extreme social judgments that actors expect. For instance, research on outcome dependency suggests that observers whose own outcomes (e.g., payments in a research study) depend on actors' performances are likely to judge such actors extremely (Berscheid, Graziano, Monson, & Dermer, 1976). These actors may be right that their performances have big evaluative implications. More generally, how investment in actors' performances would moderate the overblown implication effect is an open question for future research.

How does the OIE change over time? Finally, our studies investigated how actors' and observers' perceptions respond to a single performance event. What would happen if actors' skills—both successes and failures—are on display over multiple rounds? Do actors feel most under evaluative threat when they know observers do not know them well, meaning that the OIE may diminish across time? Or instead will actors' metaperceptions respond to what is evaluatively focal, that which has just occurred (or is about to occur)? We did find that the OIE extends to evaluations of (and estimates of being evaluated by) acquaintances (Studies 4, 6b, and 6c) and even friends (Study 6a). This suggests that the OIE may continue to persist over time.

On the other hand, with repetition, actors may forget just how much their talents will surprise and, thus, impress others. Consider a professional singer who mindlessly sings along to the radio. She may fail to realize just how impressed her taxi driver will be. Or a party guest who brings his tried and true recipe, though one with which he has become somewhat bored, may fail to appreciate how much his cookies will be encoded as a success that reflects his superior cooking abilities. In such contexts, actors' performance may actually loom larger in the eyes of observers than actors realize.

Finally, there are differences among qualities in what Kammrath, Ames, and Scholer (2007) call their *maintenance* levels—how much or how often one must show a relevant behavior to continue to be credited with having that quality. For example, Kammrath et al. (2007) found that observers update their impression of others' agreeableness more quickly than impressions of others' conscientiousness. If actors fail to anticipate that observers are quick versus slow to update their impressions in certain domains, then this itself could reduce or enhance the overblown implications effect. More generally, although we think that working trait definitions are a key but overlooked construct in understanding meta-insight, a more complete understanding will require additional empirical work as well.

Conclusions

As people navigate through their personal and professional lives, they aim not merely to passively estimate but also to actively manage others' impressions (e.g., Jones & Pittman, 1982; Leary & Kowalski, 1990; Schlenker & Weigold, 1992). Thus, metaperceptions are important barometers of whether people (think they) are doing so effectively. When people's metaperceptions are inaccurate, they may make suboptimal decisions about how best to invest in further impression management. Those who make a single inane comment during a work meeting may go to unnecessary lengths to redeem themselves in the eyes of their colleagues, and those who offer a single stroke of genius may be mistaken about how much they can rest on these (thin) laurels (see Anderson, Ames, & Gosling, 2008; Elfenbein, Eisenkraft, & Ding, 2009). We may do well to keep in mind that although our specific competencies are sometimes on full display, our broader abilities almost never are.

References

Albright, L., Forest, C., & Reiseter, K. (2001). Acting, behaving, and the selfless basis of metaperception. *Journal of Personality and Social Psychology, 81,* 910–921. http://dx.doi.org/10.1037/0022-3514.81.5.910

Albright, L., & Malloy, T. E. (1999). Self-observation of social behavior and metaperception. *Journal of Personality and Social Psychology, 77,* 726–734. http://dx.doi.org/10.1037/0022-3514.77.4.726

Anderson, C., Ames, D. R., & Gosling, S. D. (2008). Punishing hubris: The perils of overestimating one's status in a group. *Personality and Social Psychology Bulletin, 34,* 90–101. http://dx.doi.org/10.1177/0146167207307489

Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology, 46,* 610–620. http://dx.doi.org/10.1037/0022-3514.46.3.610

Berscheid, E., Graziano, W., Monson, T., & Dermer, M. (1976). Outcome dependency: Attention, attribution, and attraction. *Journal of Personality and Social Psychology, 34,* 978–989. http://dx.doi.org/10.1037/0022-3514.34.5.978

Campbell, J. D., & Fehr, B. (1990). Self-esteem and perceptions of conveyed impressions: Is negative affectivity associated with greater realism? *Journal of Personality and Social Psychology, 58,* 122–133. http://dx.doi.org/10.1037/0022-3514.58.1.122

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science, 20,* 1033–1039. http://dx.doi.org/10.1111/j.1467-9280.2009.02409.x

Carlson, E. N., & Kenny, D. A. (2012). Meta-accuracy: Do we know how others see us? In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge* (pp. 242–257). New York, NY: Guilford Press.

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology, 101,* 831–846. http://dx.doi.org/10.1037/a0024297

Cervone, D., & Shoda, Y. (1999). Beyond traits in the study of personality coherence. *Current Directions in Psychological Science, 8,* 27–32. http://dx.doi.org/10.1111/1467-8721.00007

Cooley, C. H. (1902). *Human nature and the social order* (Rev. ed.). New York, NY: Scribner's.

Cramer, A., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S., . . . Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality, 26,* 414–431. http://dx.doi.org/10.1002/per.1866

Critcher, C. R., & Dunning, D. (2015). Self-affirmations provide a broader perspective on self-threat. *Personality and Social Psychology Bulletin, 41,* 3–18. http://dx.doi.org/10.1177/0146167214554956

Critcher, C. R., Dunning, D., & Rom, S. C. (2015). Causal trait theories: A new form of person knowledge that explains egocentric pattern projection. *Journal of Personality and Social Psychology, 108,* 400–416. http://dx.doi.org/10.1037/pspa0000019

Critcher, C. R., Helzer, E. G., & Dunning, D. (2011). Self-enhancement via redefinition: Defining social concepts to ensure positive views of self. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 69–91). New York, NY: Guilford Press.

DeSteno, D., & Salovey, P. (1997). Structural dynamism in the concept of self: A flexible model for a malleable concept. *Review of General Psychology, 1,* 389–409. http://dx.doi.org/10.1037/1089-2680.1.4.389

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York, NY: Psychology Press. http://dx.doi.org/10.4324/9780203337998

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57,* 1082–1090. http://dx.doi.org/10.1037/0022-3514.57.6.1082

Elfenbein, H. A., Eisenkraft, N., & Ding, W. W. (2009). Do we know who values us? Dyadic meta-accuracy in the perception of professional relationships. *Psychological Science, 20,* 1081–1083. http://dx.doi.org/10.1111/j.1467-9280.2009.02396.x

Epley, N., Savitsky, K., & Gilovich, T. (2002). Empathy neglect: Reconciling the spotlight effect and the correspondence bias. *Journal of Personality and Social Psychology, 83,* 300–312. http://dx.doi.org/10.1037/0022-3514.83.2.300

Fenigstein, A., Scheier, M., & Buss, A. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43,* 522–527. http://dx.doi.org/10.1037/h0076760

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13,* 83–87. http://dx.doi.org/10.1111/j.0963-7214.2004.00280.x

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90. http://dx.doi.org/10.1037/0033-2909.101.1.75

Gallrein, A. 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. http://dx.doi.org/10.1016/j.jrp.2013.04.001

Gallrein, A. B., Weßels, N. M., Carlson, E. N., & Leising, D. (2016). I still cannot see it—A replication of blind spots in self-perception. *Journal of Research in Personality, 60,* 1–7. http://dx.doi.org/10.1016/j.jrp.2015.10.002

Gilovich, T., Kruger, J., & Medvec, V. H. (2002). The spotlight effect revisited: Overestimating the manifest variability of our actions and appearances. *Journal of Experimental Social Psychology, 38,* 93–99. http://dx.doi.org/10.1006/jesp.2001.1490

Greve, W., & Wentura, D. (2003). Immunizing the self: Self-concept stabilization through reality-adaptive self-definitions. *Personality and Social Psychology Bulletin, 29,* 39–50. http://dx.doi.org/10.1177/0146167202238370

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54,* 917–924. http://dx.doi.org/10.1037/0022-3514.54.6.917

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology, 91,* 436–455. http://dx.doi.org/10.1037/0022-3514.91.3.436

Hayes, A. F., & Dunning, D. (1997). Construal processes and trait ambiguity: Implications for self-peer agreement in personality judgment. *Journal of Personality and Social Psychology, 72,* 664–677. http://dx.doi.org/10.1037/0022-3514.72.3.664

Hope, D. A., & Heimberg, R. G. (1988). Public and private self-consciousness and social phobia. *Journal of Personality Assessment, 52,* 626–639. http://dx.doi.org/10.1207/s15327752jpa5204_3

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3,* 1–24. http://dx.doi.org/10.1016/0022-1031(67)90034-0

Jones, E. E., & Nisbett, R. E. (1971). *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, NJ: General Learning Press.

Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (pp. 231–261). Hillsdale, NJ: Erlbaum.

Kammrath, L. K., Ames, D. R., & Scholer, A. A. (2007). Keeping up impressions: Inferential rules for impression change across the Big Five. *Journal of Experimental Social Psychology, 43,* 450–457. http://dx.doi.org/10.1016/j.jesp.2006.04.006

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

Kenny, D. A., & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114,* 145–161. http://dx.doi.org/10.1037/0033-2909.114.1.145

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77,* 1121–1134. http://dx.doi.org/10.1037/0022-3514.77.6.1121

Kruger, J., & Gilovich, T. (1999). "Naive cynicism" in everyday theories of responsibility assessment: On biased assumptions of bias. *Journal of Personality and Social Psychology, 76,* 743–753. http://dx.doi.org/10.1037/0022-3514.76.5.743

Kurman, J. (2003). The role of perceived specificity level of failure events in self-enhancement and in constructive self-criticism. *Personality and Social Psychology Bulletin, 29,* 285–294. http://dx.doi.org/10.1177/0146167202239053

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107,* 34–47. http://dx.doi.org/10.1037/0033-2909.107.1.34

Maass, A., Salvi, D., Arcuri, L., & Semin, G. (1989). Language use in intergroup contexts: The linguistic intergroup bias. *Journal of Personality and Social Psychology, 57,* 981–993. http://dx.doi.org/10.1037/0022-3514.57.6.981

Malle, B. F. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin, 132,* 895–919. http://dx.doi.org/10.1037/0033-2909.132.6.895

Malloy, T. E., Albright, L., Kenny, D. A., Agatstein, F., & Winquist, L. (1997). Interpersonal perception and metaperception in nonoverlapping social groups. *Journal of Personality and Social Psychology, 72,* 390–398. http://dx.doi.org/10.1037/0022-3514.72.2.390

Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology, 51,* 858–866. http://dx.doi.org/10.1037/0022-3514.51.4.858

Markus, H., & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology, 38,* 299–337. http://dx.doi.org/10.1146/annurev.ps.38.020187.001503

McConnell, A. R. (2011). The multiple self-aspects framework: Self-concept representation and its implications. *Personality and Social Psychology Review, 15,* 3–27. http://dx.doi.org/10.1177/1088868310371101

Pervin, L. (1994). A critical analysis of current trait theory. *Psychological Inquiry, 5,* 103–113. http://dx.doi.org/10.1207/s15327965pli0502_1

Pinkham, A. E., Griffin, M., Baron, R., Sasson, N. J., & Gur, R. C. (2010). The face in the crowd effect: Anger superiority when using real faces and multiple identities. *Emotion, 10,* 141–146. http://dx.doi.org/10.1037/a0017387

Pronin, E., Lin, D., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28,* 369–381. http://dx.doi.org/10.1177/0146167202286008

Ross, L., Amabile, T. M., & Steinmetz, J. L. (1977). Social roles, social control, and biases in social-perception processes. *Journal of Personality and Social Psychology, 35,* 485–494. http://dx.doi.org/10.1037/0022-3514.35.7.485

Ross, L., & Nisbett, R. (1991). *The person and the situation*. Philadelphia, PA: Temple University Press.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology, 37,* 322–336. http://dx.doi.org/10.1037/0022-3514.37.3.322

Savitsky, K., Epley, N., & Gilovich, T. (2001). Do others judge us as harshly as we think? Overestimating the impact of our failures, shortcomings, and mishaps. *Journal of Personality and Social Psychology, 81,* 44–56. http://dx.doi.org/10.1037/0022-3514.81.1.44

Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology, 43,* 133–168. http://dx.doi.org/10.1146/annurev.ps.43.020192.001025

Seitchik, A. E., Brown, A. J., & Harkins, S. G. (2017). Social facilitation: Using the molecular to inform the molar. In S. G. Harkins, K. D. Williams, & J. M. Burger (Eds.), *The Oxford handbook of social influence* (pp. 183–203). New York, NY: Oxford University Press.

Showers, C. J., & Zeigler-Hill, V. (2003). Organization of self-knowledge: Features, functions, and flexibility. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 47–67). New York, NY: Guilford Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22,* 1359–1366. http://dx.doi.org/10.1177/0956797611417632

Tice, D. M. (1992). Self-concept change and self-presentation: The looking glass self is also a magnifying glass. *Journal of Personality and Social Psychology, 63,* 435–451. http://dx.doi.org/10.1037/0022-3514.63.3.435

Tice, D. M., & Wallace, H. M. (2003). The reflected self: Creating yourself as (you think) others see you. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 91–105). New York, NY: Guilford Press.

Turner, R. G., Carver, C. S., Scheier, M. F., & Ickes, W. (1978). Correlates of self-consciousness. *Journal of Personality Assessment, 42,* 285–289. http://dx.doi.org/10.1207/s15327752jpa4203_10

Van Boven, L., Kamada, A., & Gilovich, T. (1999). The perceiver as perceived: Everyday intuitions about the correspondence bias. *Journal of Personality and Social Psychology, 77,* 1188–1199. http://dx.doi.org/10.1037/0022-3514.77.6.1188

Van Boven, L., White, K., Kamada, A., & Gilovich, T. (2003). Intuitions about situational correction in self and others. *Journal of Personality and Social Psychology, 85,* 249–258. http://dx.doi.org/10.1037/0022-3514.85.2.249

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology, 98,* 281–300. http://dx.doi.org/10.1037/a0017908

Vazire, S., & Carlson, E. N. (2010). Self-knowledge of personality: Do people know themselves? *Social and Personality Psychology Compass, 4,* 605–620. http://dx.doi.org/10.1111/j.1751-9004.2010.00280.x

Wigboldus, D. H., Semin, G. R., & Spears, R. (2000). How do we communicate stereotypes? Linguistic bases and inferential consequences. *Journal of Personality and Social Psychology, 78,* 5–18. http://dx.doi.org/10.1037/0022-3514.78.1.5

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269–274. http://dx.doi.org/10.1126/science.149.3681.269

Received November 6, 2017
Revision received May 30, 2019
Accepted June 7, 2019 ■