The Overblown Implications Effect

Date Submitted: 7/2/2016

Abstract

As actors in a social world, people constantly engage in behaviors that put their traits and abilities on stage. But do actors understand the implications of these social performances for how others view them? Seven studies support an *overblown implications effect*: Actors overestimate how much observers think an actor's one-off success or failure offers clear insight about the actor's relevant trait. Actors overestimated how much observers would draw inferences about actors' intelligence (Study 1), dateability (Study 2), or likeability (Study 3) based on actors' performance during a trivia quiz, answers during a dating profile video, or exclusion from a particular group, respectively. The OIE emerged because actors failed to understand how observers construed the traits being judged, not because actors did not understand the strength of observers' priors (Studies 2-3), nor because actors ruminated on their past successes or failures (Study 4). In explaining the OIE, we introduce the construct of *working trait definitions*—accessible beliefs about what behaviors define a trait. When actors try to adopt observers' perspective, actors' attention is drawn to the threat of evaluation, leading the behavior being (or just) performed to seem disproportionately important in defining a trait. This explains why only actors' guesses about observers' judgments, not uninvolved bystanders' guesses, show evidence of the OIE (Study 5), as well as why actors most anxious about social evaluation show the strongest OIE (cross-study meta-analysis). An intervention that broadened actors' working trait definitions to include other trait-relevant behaviors enhanced the accuracy of actors' meta-perceptions (Studies 6-7).

Keywords: meta-perceptions, social judgment, working trait definitions, behavioral diagnosticity

The Overblown Implications Effect

"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.

Psychologists have been long interested in how being observed changes people's behavior. Being observed can facilitate or inhibit performance. When in the presence of evaluators, people can be energized by forces of social facilitation (Zajonc, 1965) or live up to standards that others expect of us (e.g., Rosenthal & Jacobson, 1968). Alternatively people may choke under evaluators' watchful eye (e.g., Baumeister, 1984), or show diminished performance due to fears of confirming stereotypes they may hold (Steele, 1997).

But with all of this concern about how being evaluated can influence actors' behaviors, there has been much less examination of whether actors actually understand how observers view them. Under scrutiny, do actors appreciate the true evaluative stakes of their own behaviors? Or do actors underappreciate their behaviors' evaluative importance? Perhaps game show contestants are so concerned about getting a question right that they don't fully realize just how much they will come off as a genius or an idiot based on their knowledge of esoteric trivia. But as we develop below, we instead argue that the threat that comes with being evaluated is itself the culprit that keeps actors from understanding how observers view them. We argue and demonstrate that actors show an *overblown implications effect:* Actors exaggerate how much their performance will speak to their standing on traits in the eyes of observers.

Thus far, meta-perceptions—people's estimates of how others view them—and their accuracy have mainly been studied in the field of personality psychology. This work primarily focuses on how much actors' meta-perceptions correlate with observers' actual perceptions. And indeed, people possess some awareness of how they are viewed by others (e.g., Carlson, Vazire, & Furr, 2011; Albright, Forest, & Reiseter, 2001). For instance, people are able to correctly pinpoint what traits others deem most characteristic of them (Carlson, Furr, & Vazire, 2010). Furthermore, people recognize that they make different impressions on different groups of people, such as on their friends versus their parents (Carlson & Furr, 2009). Together, these findings suggest that people have some idea of what observers find to be distinctive of them as well as how their behaviors impact observers' perceptions. However, meta-perceptions also possess blindspots. For example, people often fail to realize that they possess certain distinctive traits, even when there is social consensus that one does (Gallerin, Carlson, Holstein, & Leising, 2013). The nature of these blind spots, however, has not been systematically identified, suggesting that there is more to be understood about precisely when and why such inaccuracies emerge.

When personality psychologists study the accuracy of meta-perceptions, they typically examine the self's decontextualized beliefs about how others view the stable, enduring dispositions of the self. Thus, more study into the formation of meta-perceptions—i.e., how actors assume observers form impressions based on actors' behavior (and whether such assumptions are right)—is needed. After actors fail or succeed on a task, they likely have insight into the *direction* that observers' perceptions will shift. After all, answering the game-winning Trivial Pursuit question or turning a simple parallel parking job into a 23-point turn clearly speak

well and poorly to one's intelligence and driving ability, respectively. But are actors good at knowing how much weight observers will put on such focal successes and failures?

In considering why actors may fail to understand how observers judge them, it is useful to consider more carefully what is being judged when making a personality assessment. For example, what does it mean to evaluate someone as intelligent? Most obviously, intelligent people do intelligent things. Those who are intelligent may use more ornate words, remember obscure facts from their childhood, or even know the order of the 118 elements in the Periodic Table. But these different cues to intelligence are not perfectly correlated, and we rarely observe them all at once.

According to latent trait models of personality, personality traits are unobservable qualities that predispose people to different behaviors (Borsboom, 2008; McCrae & Costa, 2008). However, as those familiar with the great person versus situation debate know (e.g., Pervin, 1994; Ross & Nisbett, 1991), trait-relevant behaviors show variability across situations (e.g., Fleeson, 2004). In part, this reflects the influence of different situational factors that overwhelm or even selectively activate different aspects of one's personality (Cramer et al., 2012). But also, this reflects that any one behavior gives only so much information about a person's trait. When determining whether someone else is intelligent, one probably wants to know more than whether they can get all the way from hydrogen to ununoctium.

From this perspective, one may ask whether actors understand how observers think about and define a trait in question. Psychologists have long appreciated that for complex constructs, people's conceptions of those constructs may be based on only a limited amount of information. Consider the self, an object about which we have almost overwhelming amount of information. Despite the vast stores of self-knowledge, only certain aspects of the self are salient at one time.

That is, people's *working self-concepts*—their accessible self-knowledge (Markus & Wurf, 1987)—show variability from moment to moment (Markus & Kunda, 1986; DeSteno & Salovey, 1997; Cervone & Shoda, 1999; 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 lean on only a subset of relevant behavioral cues when considering another's standing on a trait. 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 the working trait definitions may be similarly predictable. As actors consider the impressions observers are forming (or soon will form) of them, this evaluative threat may cause the corresponding behavior to loom large in actors' working trait definitions. As a result, actors see their own performance as likely to exert considerable sway on observers' impressions. But observers' actual impressions do not have the same evaluative stakes for observers themselves. As a result, observers' working trait definitions may not be as dominated by the focal behavior as actors anticipate. We propose that this asymmetry is what gives rise to the overblown implications effect. We summarize this account in Figure 1.

[INSERT FIGURE 1 HERE]

Although we are—to our knowledge—the first to articulate this account of why meta-perceptions may err, it is useful to consider our proposal in light of other proposals for why people fail to appreciate how others are thinking of them. Consider a driver who is attempting to

parallel park. Past theorizing has suggested that actors may: (a) neglect that observers will empathize with them if they perform poorly (*empathy neglect*; Epley, Savitsky, & Gilovich, 2002), (b) incorrectly believe that observers judge them in self-interested and self-aggrandizing ways (*naïve cynicism*; Savitsky, Epley, & Gilovich, 2001), and/or (c) disregard situational constraints (e.g., how tight the space is) that provide alternative explanations for performance quality (Savitsky et al., 2001). The *overblown implications effect* instead hypothesizes that when predicting how observers will view them, actors' (compared to observers') working trait definitions of "driving ability" become hyper-focused on parallel parking, thereby overlooking other behaviors (including behaviors about which observers recognize they have no information) that will moderate observers' impressions.

The overblown implications effect generates a number of testable predicts that are either not anticipated by, or that directly go against, these previously articulated theories. First, instead of predicting that meta-perceptions merely err in believing that observers are harsher than they actually are, the overblown implications effect predicts meta-perceptions will be too extreme. That is, the OIE predicts does not merely predict that actors will see observers as especially harsh in response to their failure, but also as especially laudatory in light of their success. Second, the overblown implications effects predicts that even when performances take place in a vacuum—i.e., when there are no situational cues to explain away behaviors as non-diagnostic—asymmetries between actors and observers should persist. This is especially relevant when considering people's understanding of behaviors that have yet to occur ("Whether or not I deliver a compelling speech will prove to my employer that I am competent or incompetent…"), and thus when situational excuses are not yet knowable. Third, the overblown implications effect—by localizing the constriction of working trait definitions to the fact that they are threatening—

should not be a phenomenon that merely describes people's lay beliefs about the charitableness or cynicism of other people. Instead, those under evaluative threat, especially those most concerned about how they are evaluated, should be those who most show the OIE. Fourth, the overblown implications effect identifies a unique tool for debiasing meta-perceptions: expanding actors' working trait definitions.

Overview of the Present Research

The present research examined if actors overweigh how much a specific behavioral performance factors into observers' perceptions of them. We begin by testing the overblown implications effect in three evaluative domains in which we could (unbeknownst to actors) randomly assign them to fail or succeed. We test whether actors overblow the implications of their own success or failure in predicting how observers will view their intelligence (Study 1), desirability as a dating partner (Study 2), and likeability (Study 3). We test whether the overblown implications effect emerges in prospective judgments of how much diagnosticity one's own behaviors will offer (Study 4), as well as whether its effects are specific to actors (Study 5). After testing (and ruling out) four alternative explanations for the OIE, Studies 6 and 7 determine whether expanding actors' working trait definitions debiases their meta-perceptions.

We report all manipulations below. The only exclusions were participants for whom there was a problem collecting data (e.g., a video of them did not record). For the multi-stage lab studies (Study 1-3, 7), we collected as many participants as we could in a semester for the subject pools that we used. For the survey studies (Studies 4-6), we aimed for 100 participants per cell. We report exploratory or unanalyzed measures in the Supplemental Materials.

Study 1

Study 1 tested 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 meta-judgments of how observers likely viewed them. We then engineered a focal failure of success. We predicted an *overblown implications effect*, that meta-judgments would shift more in light of this focal performance than social judgments actually do.

Method

Participants and design. One hundred ninety-eight undergraduates from an American university completed a lab session for course credit. Actors were randomly assigned to a focal *success* or a focal *failure* condition. Each observer was yoked to a randomly selected actor.

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 in a lab room. 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 their rationale behind their selection aloud. This made sure that actors would offer plenty of behavioral information, beyond their multiple choice response, that could be interpreted to signal their competence and intelligence. Actors were told that for each of the ten 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 focal, 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 the trivia questions and regardless of their actual performance, actors were informed that they had answered 7 out of the 10 questions correctly. No actor or observer expressed suspicion about this feedback's credibility.

At this point, the experiment 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 both meta-perceptions (guesses of how the observers would rate them) and self-perceptions (evaluations of their own performance on the task), in a counterbalanced order. (Because we found that our effect was uninfluenced by order, we did not vary order in subsequent studies). These measures are described in more detail below.

Next, the experimenter returned to explain the focal evaluative 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 focal as a performance event. In addition to raising the stakes on this question (making it double or nothing), we had 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 question being asked (their publication dates). This allowed us—unbeknownst to participants—to randomly assign participants to learn they had (supposedly) answered this focal 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).

Finally, actors completed the *final perception* measures. These took the same form as the baseline perception measure—the meta-perception and self-perception measures of intelligence completed before the focal evaluative 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, seated at individual cubicles, 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. Prior to

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 supposed seven trivia questions that they 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, Medvec, & Savitsky, 2000), we reiterated the performance outcome to observers before they made their final judgments. At that point, observers completed their final social perceptions of the actors' intelligence.

Trait perceptions. The perception measures comprised five items that assessed the actor's intelligence. The *social perceptions* asked observed to judge the actors in light of their performance. The *meta-perceptions* asked 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 *self-perceptions* asked actors to judge themselves in light of their performance.

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 to other undergraduates at their university.

In order to combine our five items, which used different response scales, we standardized responses. We calculated the grand mean and standard deviation of each item across both baseline and final meta, social, and self perception judgments. By relying on these sample statistics when standardizing each measure, we preserved all effects of perception type (meta, social, and self) and time. The self-perception ($\alpha = .77$), meta-perception ($\alpha = .77$), and social perception ($\alpha = .75$) of intelligence composites all had good internal reliability.

Results and Discussion

We wanted to test whether responses to the focal feedback depended on the nature of the perception (self, meta, or social). Toward this end, we submitted the perception composites to a 2 (feedback: success or failure) X 3 (perception: self, meta, or social) X 2 (time: baseline or final) mixed-model ANOVA, with only the first factor manipulated between subjects. We found a significant three-way interaction, $F(2, 194) = 3.26$, $p = .04$, $\eta^2_p = .03$, demonstrating that perceivers responded differently to actors' success versus failure. We proceeded to conduct a series of 2 (feedback) X 2 (perception) X 2 (time) ANOVAs to understand whose perceptions were out of step with whose.

[INSERT FIGURE 2 HERE]

First, did observers respond to the high-stakes final round like actors thought they would? A significant 2 (feedback) X 2 (perception: meta or social) X 2 (time) mixed-model ANOVA

suggested that that they did not, $F(1, 97) = 4.25$, $p = .04$, $\eta^2_p = .04$[1]. Decomposing this interaction provided evidence of the overblown implications effect: Observers responded less extremely to the actors' final performance than actors thought observers would. As depicted in Figure 2, observers did shift their social impressions in response to actors' success versus failure, $F(1, 97) = 17.93$, $p < .001$, $\eta^2_p = .16$. But this shift was less pronounced than actors thought it would be, $F(1, 97) = 60.61$, $p < .001$, $\eta^2_p = .38$.

Second, we examined whether actors' meta-perceptions merely reflected their own self-perceptions of their performance. That is, although we argued that actors overblow the implications of their own performance when estimating how they are being evaluated, perhaps they more generally overblow the meaning of their own behavior in their own mind as well. Contradicting this possibility, we found that actors' self-perceptions and their meta-perceptions diverged. That is, a significant 2 (feedback) X 2 (perception: self or meta) X 2 (time) mixed-model ANOVA showed these judgments diverged, $F(1, 97) = 7.81$, $p = .006$, $\eta^2_p = .07$. Although self-perceptions were reactive to the feedback, $F(1, 97) = 37.81$, $p < .001$, $\eta^2_p = .28$, they were less so than meta-perceptions.

Third, and related to the last analysis, we asked whether actors would have had a more accurate understanding of how observers actually responded to their performance if actors had merely leaned on their own self-perceptions. As foreshadowed by these two previous analyses, a final 2 (feedback) X 2 (perception: self or social) X 2 (time) mixed-model ANOVA returned a non-significant three-way interaction, $F < 1$. Thus, both actors and observers responded to the

[1] We caution readers against comparing effect sizes of the 2(feedback) X 2(perception) X 2(time) ANOVAs because they vary in whether the two perception conditions are measured within-participants or manipulated between-participants.

actors' high-stakes performance in a similar way—i.e,, less reactively than actors thought observers would.

Although the divergence between meta-perceptions and social perceptions provides straightforward support for the overblown implications effect, we did not expect that this effect would be driven by the success condition. Instead of speculating on the reason for this asymmetry, we note that this is not robust: The extent to which the overblown implications effect is driven by a meta-/social perception mismatch in response to success, failure, or some combination of the two varies by study. Whether this variation reflects noise or is itself traceable to different aspects of our different performance contexts is a nuanced question that is beyond the scope of the present article, but may be an interesting question for future research.

At the same time, it is worthwhile to note that this asymmetry is precisely the opposite of what would be predicted by the naïve cynicism and empathy neglect accounts—those believed to underlie Savitsky et al.'s (2001) effects. We don't think Savitsky and colleagues' hypotheses were wrong, but merely reflected a different focus. Although our hypotheses have focused on how different perceptions respond to focal performance outcomes (Feedback X Perception X Time interactions), we did find a main effect of Perception that is reminiscent of Savitsky et al.'s hypothesized mechanisms. That is, we found that actors expected observers to judge actors more harshly than observers actually did. In other words, this naïve cynicism may operate in the background (depressing all meta-perceptions), but the *overblown implications effect* may characterize actors' failure to understand dynamic shifts (or, more accurately, the lack of such shifts) in observers' perceptions.

Study 2

Study 2 aimed to expand on Study 1 in two ways. First, we wanted to test the overblown implications effect in a new context and for a new trait. Participants were told they were taking part in a study on how people present themselves and communicate in dating contexts. For this reason, actors were prompted to record a video dating profile, in which they answered a series of interview questions about their dating style and preferences. After actors recorded this video or after observers had watched it, participants offered baseline impressions of actors' dateability: Observers made social judgments of actors' dateability, and actors made meta-judgments of how they thought observers saw their dateability. Once back on camera, actors were asked a "relationship IQ" question on which they were randomly assigned to succeed or fail. To see evidence of the overblown implications effect, we would expect that actors' meta-perceptions would be more reactive to this focal feedback than would observers' actual social perceptions.

Second, we wanted to disentangle two general explanations for the overblown implications effect. By our reasoning, when actors adopt the evaluative perspective of observers, they have a constricted working trait definition of the quality being judged. Applied to this context, actors imagine observers who are thinking about (and thus ready to make judgments about) dateability solely in terms of the focal performance context (the dateability IQ question). In this sense, the performance behaviors have "overblown implications" in the actors' minds.

However, by an alternative hypothesis, actors may simply fail to understand just how sticky observers' baseline beliefs are (*slow-to-update hypothesis*). That is, observers many actually think that actors' performance on the focal task speaks volumes about actors' dateability. But due to anchoring and insufficient adjustment (Epley & Gilovich, 2001, 2004, 2005, 2006), observers may not shift their overall impressions much. We see this account as less plausible because it is unclear why this mechanism of insufficient adjustment would apply more

to observers' social judgments than actors' meta-judgments. Nonetheless, we included new measures that would allow us to differentiate this *slow-to-update hypothesis* from our preferred explanation.

After the focal success or failure, participants offered impressions of two types. Crucially, participants were asked to make judgments based only on the *specific* event—whether the actor was dateable or not based only on the focal failure or success. According to the *slow-to-update hypothesis*, actors and observers should agree on the meaning of the specific event; they merely incorporate that information into an overall impression differently. But according to the idea that the overblown implications effect stems from constricted working trait definitions, actors should believe that observers will see more meaning in that specific performance event than they actually do.

Method

Participants and design. One hundred eighty-two undergraduates at an American university completed a lab session for course credit. Participants were assigned to be an *actor* or an *observer*. Actors were randomly assigned to a performance *success* or *failure* condition.

Procedure and materials. As in Study 1, each actor was yoked to an observer. We describe actors' experience first. Observers observed actors' complete experience – both by seeing the instructions actors received and watching the actors on video.

Actors. Actors took part in the study individually. Upon arrival, actors were seated in front of a laptop. They learned they would complete a study on "how people present themselves and the impressions they communicate." Actors were told that they would create a video dating profile that consisted of two parts. In the first part, participants answered questions about themselves—the "get to know you" questions. In the second part, actors answered a relationship

IQ question. The experimenter informed actors (accurately) that they would be videotaped throughout the entire study so that a future participant could observe them and their performance.

Actors were given three minutes to prepare before the video dating profile was shot. Actors introduced themselves before reading aloud and answering seven questions. They had received a list of these questions three minutes before filming began. This gave actors some time to consider responses to the prompts. Two of the questions were: "Describe the sorts of activities you would like to do on a first date" and "Think of your ideal dating partner. In what ways would you want the two of you to be similar? In what ways would you want the two of you to be different?" Actors were asked to spend around 30-60 seconds on each answer.

At this point, actors completed the baseline perception measures of their "dateability" (how good of a dating partner they would be). Actors provided self-perceptions (evaluations of their own performance at the task) followed by meta-perceptions (guesses of how the observers would rate them). These measures are described in more detail below.

Next, the experimenter returned to explain the focal task (i.e., the relationship IQ question). Actors were told, "For the last question, we will be testing your relationship IQ. This question has an objectively right or wrong answer as determined by previous research on the psychology of close relationships, and gives a sense of your 'relationship IQ.' Also, a fun fact: the producers of the TV shows *The Bachelor* and *The Bachelorette* use these questions to screen potential candidates for their show!" This final fib was meant to give a sense of legitimacy and interest value to the question. Actors were then presented with what was ostensibly a relationship IQ question: "Research has shown that five key qualities differentiate happy couples from unhappy couples. Which quality is MOST important for a couple's relationship satisfaction? Rank the following from MOST important to LEAST important. (a) communication, (b)

flexibility, (c) emotional closeness, (d) compatibility of personalities, and (e) conflict handling." Experimenters read the question to the actor and provided the actor with a written version of the question to examine.

Experimenters first confirmed verbally with actors that the answer they provided was from most to least important. Then, based purely on random assignment, actors were told that they had answered the question in the exact right order (*success*) or in the exact opposite of the right order (*failure*). In the *success* condition, experimenters said, "Wow, actually that's the exact order!" and repeated the items in the order that the actor had given, followed by, "Good job!" In the *failure* condition, experimenters said, "Wow, that's actually the exact opposite of what it should be," and repeated the items in the opposite order as the actor had given. This entire exchange happened on camera. Finally, actors completed the final measures of dateability: the same meta-perception and self-perception items they completed before the focal evaluative event. After completing these final measures, actors were debriefed and apologized to for the mild deception.

Observers. Each observer was yoked to one actor. We conducted the study until all actors were yoked to at least one observer. We collected additional observer data until the end of the semester. For the sake of both analytic consistency across studies and simplicity of presentation, we averaged the measures for observers who were yoked to the same actor.

Each actor video was clipped into two shorter videos. The first video showed the actor answering the "get to know you" questions. The second video showed the experimenter reading the relationship IQ question to the actor, the actor answering the question, and the experimenter providing the final feedback. A screen with text reiterated that the actor either answered the question correctly (*success* condition) or that the actor answered the question (*failure* condition).

Observers, seated in private rooms, 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 their yoked actor's video dating profile instead of creating one themselves.

Whenever actors completed measures of their meta-perceptions and self-perceptions, observers completed social perception measures. That is, after the first part of the actor's video dating profile but before the relationship IQ question, observers offered their baseline social perceptions. After observing the second part of the video dating profile, observers then offered their final social perceptions of their actor's dateability.

Trait perceptions. The perception measures comprised five items that assessed the actor's dateability. The *meta-perceptions* asked actors to guess how observers would judge them in light of their dating video. Actors knew they were guessing how observers would respond to those exact questions. The *social perceptions* asked observers to judge the actors in light of the dating video. The *self-perceptions* asked actors to judge themselves in light of the dating video they recorded.

The five questions asked participants to rate the actor compared to other undergraduates at the same university: "would make a better date than," "would make a better relationship partner than," "is more knowledgeable about dating than," "is likely to be in a happier relationship than," and "is more desirable as a dating partner than." Each of the questions was rated on a 101-point scale from 0% of fellow undergraduates at their university to 100% of fellow undergraduates at their university. For instance, an answer of 60% would indicate that the participant thought the actor would make a better date than 60% of fellow undergraduates at their university.

Critically, Study 2 differed from Study 1 in that participants completed two sets of final ratings. One set asked participants to rate the actors' performance in light of the total dating video—i.e., based on the *general* set of information to which participants had been exposed. The other set asked participants to rate the actors' performance only based on the focal task, the relationship IQ question—i.e., by responding to the *specific* behaviors for which actors had received success or failure feedback.

We averaged these items to create dateability perception composites. The self-perception ($\alpha = .97$), meta-perception ($\alpha = .97$), and social perception ($\alpha = .97$) of dateability composites all had good reliability.

Results and Discussion

Our two hypotheses—the *overblown implications* account and the *slow-to-update* hypothesis—differ in whether they think that actors will be mistaken about how observers will interpret the specific focal event. Thus, we began by testing whether meta-perceptions and social perceptions diverge in their interpretation of the specific focal event before analyzing the more general perceptions.

Specific final impression. For the specific impression from the focal task, we submitted the perception composites to a 2 (feedback: success or failure) X 3 (perception: self, meta, or social) X 2 (time: baseline or final specific) mixed-model ANOVA, with only the first factor manipulated between subjects. The Feedback X Perception X Time interaction was significant, $F(2, 176) = 7.93$, $p < .001$, $\eta^2_p = .08$, demonstrating that perceivers responded differently to success versus failure outcomes. We proceeded to conduct a series of 2 (feedback) X 2 (perception) X 2 (time) ANOVAs to understand whose perceptions were out of step with whose.

First, a significant 2 (feedback) X 2 (perception: meta or social) X 2 (time) mixed-model ANOVA revealed that actors believed observers would be more reactive to the final relationship IQ question than they actually were, $F(1, 88) = 11.51$, $p = .001$, $\eta^2_p = .12$. As depicted in Figure 3, observers did shift their social impressions in response to actors' success versus failure, $F(1, 88) = 22.41$, $p < .001$, $\eta^2_p = .20$. But this shift was less pronounced than actors thought it would be, $F(1, 88) = 80.20$, $p < .001$, $\eta^2_p = .48$.

Second, also consistent with Study 1, a significant 2 (feedback) X 2 (perception: self or meta) X 2 (time) mixed-model ANOVA suggested that actors' meta-perceptions did not simply reflect their own self-perceptions, $F(1, 88) = 4.23$, $p = .04$, $\eta^2_p = .05$. Although self-perceptions were reactive to the feedback, $F(1, 86) = 65.23$, $p < .001$, $\eta^2_p = .43$, they were less so than meta-perceptions.

Third, unlike Study 1, a significant 2 (feedback) X 2 (perception: self or social) X 2 (time) mixed-model ANOVA demonstrated that actors' self-perceptions were also out of sync with, and more extreme than, observers' social perceptions, $F(1, 88) = 5.37$, $p = .02$, $\eta^2_p = .06$. In short, actors' meta-perceptions were more reactive than were their self-perceptions, though observers' social perceptions were least reactive of all.

[INSERT FIGURE 3 HERE]

General final impression. For the general final impression, we submitted the perception composites to a 2 (feedback: success or failure) X 3 (perception: self, meta, or social) X 2 (time: baseline or final general) mixed-model ANOVA, with only the first factor manipulated between subjects. Surprisingly, this three-way interaction was not significant, $F < 1$.

We found this null effect somewhat baffling, especially given that the three perspectives differed in their interpretation of the focal event—the only event that occurred between the

baseline and final perceptions. Regardless, we present this study and full results both in the interest of disclosure and because the effects on the specific impression measures provide clear support for the *overblown implications* account over the *slow-to-update hypothesis*. But given the surprising null effect on the general perception measure, we felt it best to replicate our findings once more before moving to a more detailed examination of why the overblown implications effect emerges. Study 3 tests our hypotheses once again, but in a new context and with a new focal event.

Study 3

Study 3 tested the overblown implications effect in a new context and on a new trait. More specifically, we placed some participants (actors) in a social context in which they were socially accepted or rejected. Observers watched the social dynamic unfold from afar. We then examined how this social success or failure changed perceptions of actors' likeability. According to the overblown implications effect, actors' meta-perceptions of how observers view them should be more reactive to the focal success or failure than observers' perceptions actually are. By observing the overblown implications effect on both the general (like in Study 1) and specific (like in Study 2) impression measures, we can have more confidence in the robustness of both effects.

Method

Participants. Two hundred forty-eight undergraduates completed a lab session for course credit. All actors were randomly assigned to the focal *success* or *failure* condition, which involved being included or rejected as a sign of like or dislike, respectively. 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.

Actors. Actors were seated in a private room in front of a laptop. Actors were told that they would be completing a study examining whether people work better with those they like and additionally, whether outside observers can predict how well a group can work together. Actors were led to believe that there were three participants total in the current session, two who would be interviewees and one who would be the interviewer. Actors were also informed that in a later session, everyone would be an observer. Those observers would watch the situation unfold from the vantage point of previous actors, but would not participate in the groups themselves. A graphic was shown along with these instructions to more clearly display the different roles involved.

Actors were told that they would be randomly assigned to either the interviewer role or one of two interviewee roles. The instructions read,

> In order to determine whether you will be the interviewer or one of the two interviewees, each of you will be asked to choose one of the three cards on the next page. Based on the card you choose, you will be randomly assigned to one of the three roles.

Although actors completed a card-choosing task, in actuality, all actors were preprogrammed to choose a card that indicated that they would be one of the interviewees, The interviewer and second interviewee were fictitious, pre-programmed participants.

At this point, participants learned the structure of the study from their vantage point. Actors were told that they would answer a practice round question, not shown to the interviewer, to familiarize them with the format of the question, and 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. Actors were also informed, truthfully, that an observer would later be able to see everything they saw (but nothing more than they saw) throughout the entire study. A diagram (see Figure 4A) 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 session (playing the role of the interviewer and the other interviewee), the next screen, presented for five seconds, displayed a spinning "loading" icon and explained that the survey would automatically continue once everyone had finished learning about their roles.

[INSERT FIGURE 4 HERE]

Actors first completed a practice round of questions about themselves that only the observer, but not the other two (actually fictitious) participants, would see. We included this step so that observers would have some grounds on which to offer baseline perceptions. First, actors listed three values they had and were informed that they would be asked to elaborate on each of the values they chose. The instructions read,

> Please take a moment to think about and name the three values that are important to you in your life. These values could be things like artistic or musical skills or appreciation, financial security, sense of humor, relations with friends/family, spontaneity/living in the moment, social skills, educational accomplishment, creativity, organizational/managerial skills, or physical health.

After listing the three values, the survey asked about each of the values sequentially. The instructions read, "Please spend about a minute describing why the first value you picked has importance to you. Please describe why the value has importance to you as if you were describing it to another person." Their verbatim response for the value they chose was reiterated (e.g., "You listed your first value as:"). Actors were also informed that after a minute had

elapsed, the computer would automatically advance them to writing about the next value. The process repeated until the actor wrote about all three of the values they had originally listed.

At this point, actors completed the baseline perception measures of likeability. Actors provided self-perceptions (evaluations of their own performance at the task) followed by meta-perceptions (guesses of how the observers would rate them). These measures are described in more detail below.

Next, actors completed the interview round of questions about themselves that both the interviewer and the observer would see. First, actors were asked to list their three best qualities and were informed that they would be asked to elaborate on each of the qualities they chose. The instructions read, "Name three of your best qualities, that is, three aspects about yourself that you like the most." After listing the three qualities, the survey asked about each of the qualities sequentially. The instructions read,

> Please spend about a minute and a half giving an example of an instance in which you exhibited the first characteristic you indicated that you like about yourself. Please describe this example as if you were describing it to another person.

Their verbatim response for the quality they chose was reiterated (e.g., "You listed your first quality as:"). Actors were also informed that after 90 seconds had elapsed, the computer would automatically advance them to writing about the next quality. As with the practice questions, this process repeated until the actors had written about all three of the qualities they had listed.

After completing the interview round, actors ostensibly waited for the interviewer to make a decision. Actors were shown a screen with a spinning "loading" icon for 25 seconds and informed,

> Since both you and the other interviewee submitted answers in three parts, the interviewer has been reading both of your responses as the two of you were answering them. Now, please wait for the interviewer to decide whether or not they would like to work with one, both, or neither of you.

Actors were randomly assigned to learn that the interviewer had chosen to work with them but not the other interviewee (*success*), or that the interviewer had chosen to work with the other interviewee but not with the participant (*failure*). At this point, actors completed the *final* meta-perception and self-perception measures of likeability—both specific (focusing specifically on the portion involving the acceptance or rejection) and general. After completing the final measures, actors were debriefed and apologized to for the mild deception.

Observers. Each observer was yoked to one actor. As with Study 2, we conducted the study until all actors were yoked to at least one observer, but we continued to collect additional observers until an a priori identified stopping point. Once again, we averaged the measures for observers who were yoked to the same actor.

Each actor's responses were split into two transcripts, each on a single sheet of paper. The first transcript included the actor's responses to the practice round questions and responses, those about the actor's values. Observers offered their baseline perceptions after reading that transcript. The second participant transcript provided actors' responses to the focal interview round, when the actors discussed their best qualities.

Like actors, observers were also seated in private rooms. But in observers' case, they were told they would be *observing* a previous participant who completed a study on whether people work better with those they like. They were also told the study was examining whether outside observers can predict how well a group can work together. Observers saw everything that

actors saw, but read actors' responses to the three practice questions and the three interview questions instead of providing responses of their own. Given the moderate complexity of the situation, a diagram (see Figure 4B) reiterated the role the observer would take as well as the roles of the other participants. Observers completed *social* perception measures that were parallel to—in timing and in form—the actors' meta-perception and self-perception measures. (As before, observers did not see actors' social or meta-perception responses). More specifically, observers completed the baseline measures of likeability after the practice round and the final social perceptions measures after witnessing the social acceptance or rejection.

Trait perceptions. The perception measures comprised six items that assessed the actor's likeability. The *meta-perception* 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*) to 9 (*extremely*). The final two questions asked participants how much the actor would "make a good impression" and would be "able to get along with others" on 9-points scales anchored at 1 (*not at all*) to 9 (*extremely*).

As in Study 2, participants completed the final perception ratings twice. One set of items asked participants to offer their *general* impressions of the actor. The other *specific* items asked participants to respond to what was conveyed by the focal task in particular (i.e., the second round of questions that was the basis of social inclusion or rejection). We averaged these items to

create likeability perception composites. The self-perception (α = .96), meta-perception (α = .96), and social perception (α = .97) of likeability composites all had good reliability.

Results and Discussion

We proceeded to test whether different types of perceptions were differentially sensitive to the actor's social failure or success. We follow the same analytic approach used in Study 2. That is, we conducted 2 (feedback: success or failure) X 3 (perception: self, meta, or social) X 2 (time: baseline or final) mixed-model ANOVA. For our first set of analyses, the final measure was the impression based on the specific focal event. For the second set of analyses, the final measure was the general impression.

Specific final impression. First, we tested whether perceptions differed about what the final episode (i.e., the interview that led to social acceptance or rejection) signified about the actor's likeability. Suggesting they did, we found a significant Feedback X Perception X Time interaction, $F(2, 226) = 3.70$, $p = .03$, $\eta^2_p = .03$. As before we decompose this interaction by comparing each type of perception using Feedback X 2 (Perception) X Time mixed-model ANOVAs.

First, we tested whether actors assumed that observers would see actors' social success or failure as more informative about actors' likeability than observers themselves believed. As expected, the Feedback X Perception X Time interaction was significant, $F(1, 113) = 5.46$, $p = .02$, $\eta^2_p = .05$. As depicted in Figure 5A, observers did shift their social impressions in response to actors' success versus failure, $F(1, 113) = 4.90$, $p = .03$, $\eta^2_p = .04$, but actors thought this shift would be more pronounced, $F(1, 113) = 46.86$, $p < .001$, $\eta^2_p = .29$.

[INSERT FIGURE 5 HERE]

Second, we tested whether actors merely leaned on their own self-perceptions of what the social acceptance or rejection implied in guessing observers' reactions, or whether actors thought observers would be particularly reactive. Supporting the latter interpretation, we observed a Feedback X Perception (meta or self) X Time interaction, $F(1, 113) = 7.21$, $p = .01$, $\eta^2_p = .06$. Actors saw fewer implications in their own social acceptance or failure, $F(1, 113) = 33.03$, $p < .001$, $\eta^2_p = .23$, than they thought that observers would.

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 (and with Study 1), we failed to observe a significant Feedback X Perception (self or social) X Time interaction, $F < 1$. That is, social observers and actors themselves saw similar, but relatively smaller, implications in the episode that produced social acceptance or rejection.

General final impression. Conceptually replicating Study 1, we found the same pattern of results on the general impressions. That is, the predicted 2 (feedback: success or failure) X 3 (perception: self, meta, or social) X 2 (time: baseline or final) interaction emerged, $F(2, 226) = 6.67$, $p = .002$, $\eta^2_p = .06$. To determine whether this interaction reflected the predicted pattern, we proceeded to test all three 2 (feedback) X 2 (perception) X 2 (time) interactions.

First, we tested whether social perceivers were less reactive to the actors' social success or failure than observers assumed they would be. Consistent with Study 1, a Feedback X Perception (meta or social) X Time interaction suggested that this was the case, $F(1, 113) = 10.50$, $p = .002$, $\eta^2_p = .08$. As depicted in Figure 5B, although observers did shift their general social impressions in response to actors' success versus failure, $F(1, 113) = 4.23$, $p = .04$, $\eta^2_p = .04$, this shift was less pronounced than actors thought it would be, $F(1, 113) = 46.86$, $p < .001$,

η^2_p = .29. This finding reinforces our suspicion that this null result in Study 2 may have simply been a fluke.

Second, we once again found that meta-perceivers were not merely using their own personal interpretations of their performance when making meta-judgments about others. That is, we also found a significant Feedback X Perception (meta or self) X Time interaction, $F(1, 113) = 9.71$, $p = .002$, $\eta^2_p = .08$. That is, actors' self-perceptions were not as reactive to their own success or failure, $F(1, 113) = 33.03$, $p < .001$, $\eta^2_p = .23$, 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 Feedback X Perception (self or social) X Time interaction failed to reach significance, $F < 1$. In short, not only did actors fail to understand how few implications observers would see in the social interaction that got them included or excluded, but they were similarly wrong about observers' final general impressions of them.

Study 4

We have repeatedly demonstrated—across contexts and for different traits—that actors' meta-perceptions are misguided. Actors assume social perceivers will draw stronger trait inferences from their performance than they actually do. By showing that actors and observers disagreed on the meaning of the performance in question (Studies 2 and 3), we provided support for our overblown implications account. But why exactly do actors and observers have these conflicting perspectives?

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 focal behavioral context to loom

disproportionately large in actors' meta-perceptions. But because social perceptions do not have personal stakes for observers, such constriction does not actually occur.

By an alternative *self-rumination* hypothesis, actors overblow the implications of their performance because they ruminate on their recent failures and successes. That is, actors may replay in their minds the specifics of their embarrassing, demotivating defeats or the glorious details of their energizing victories. If social observers do not have the same penchant for ruminating on others' performance, then this could offer an alternative explanation for meta-perceivers' relatively constricted working trait definitions. One difficulty for this account is that it does not easily account for our findings in Studies 1-3 why actors' self-perceptions did not show the same reactivity to focal successes or failures as actors' meta-perceptions. Of course, the self has much more information about itself than do social perceivers (perhaps explaining the more muted reactions of self-perceptions), so we hesitate to draw strong conclusions about the relatively muted responses of self-perceptions.

Study 4 aimed to provide a more definitive test. We asked participants to simulate the perspective of either an actor or an observer to an upcoming situation. By asking people to comment on prospective behavioral episodes, we eliminated the possibility that any results could be driven by rumination on one's recent performance. Observers indicated how much they would learn about another person's trait based on their performance on a trait-relevant behavior. Actors were asked to consider that they were about to perform those behaviors, but guessed how those who would be observing (and judging) them would answer the same questions.

By our preferred account, those taking actors' perspective should guess that their evaluators would see more diagnosticity in upcoming behaviors than those taking observers' perspective actually would. That is, merely considering how they themselves would be judged

should be sufficient to constrict the working trait definitions that underlie their meta-perceptions. By the alternative *self-rumination* hypothesis, given there was no past performance to actually ruminate about (because actors were merely considering an upcoming evaluation), there should be no systematic mismatch between actors' meta-perceptions and observers' actual perceptions of diagnosticity.

Method

Participants. Two hundred fifteen undergraduates from an American university completed a lab session for course credit. Participants were randomly assigned to one of two perspective conditions: *actor* or *observer*.

Procedure. Seated at a laptop in a private room, participants were asked to simulate being in different situations. We emphasized that they should try to fully place themselves in the context, to vividly visualize being there, and to be attuned to what they would be thinking and feeling. Each scenario described an interaction in which an actor's skills or abilities would be on display to an observer. The wording was varied such that the exact same situation was described from the vantage point of the actor or the observer. Crucially, all scenarios asked participants to consider these behavioral contexts in prospect. That is, no information was provided about whether the behavior was performed successfully or not. A picture also accompanied each scenario to reinforce the perspective manipulation (see Figure 6).

[INSERT FIGURE 6 HERE]

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."

For this item, 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 to predict observers' responses: "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*). Each scenario described a different behavior that spoke to a different trait (see Table 1 for a summary of the behaviors and traits associated with each scenario). 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) X 10 (scenario) repeated-measures ANOVA. As hypothesized, there was a strong main effect of Perspective, $F(1, 213) = 13.29$, $p < .001$, $\eta^2_p = .06$. Those adopting the perspective of observers saw significantly less diagnosticity in actors' upcoming behavior ($M = 5.93$, $SD = 1.42$) than actors thought observers would ($M = 6.61$, $SD = 1.33$). Table 1 presents these results by scenario.

That actors and observers have prospective disagreement about behaviors' implications cannot be explained by actors' ruminating on their recent performance. Instead, these findings are consistent with our account that imagining how others are evaluating the self causes working trait definitions to constrict around the source of evaluative apprehension. That said, this story emphasizes that constricted working trait definitions stem not merely from imagining how

another evaluates just anyone, but from considering how another evaluates the self. Study 5 tests this boundary condition.

Study 5

In Study 5, we aimed to test whether the overblown implications effect stems from people estimating how they themselves will be evaluated. By an alternative account, the OIE does not reflect the narrowed working trait definitions that come from considering how others view the self, but may instead merely be a property of how people attempt to read the minds of another person. That is, do we simply guess that others see more diagnosticity in anyone's actions, not our own actions in particular? At least superficially, this alternative possibility is supported by research showing that people think others make more extreme dispositional attributions—not just about the self, but about anyone—than they actually do (Van Boven, White, Kamada, & Gilovich, 2003; Pronin, Lin, & Ross, 2002).

To address this alternative explanation, Study 5 added a third perspective condition. We retained our actor and observer perspective conditions, but added a third group who simulated the perspective of uninvolved bystanders. Bystanders considered the interactions of actors and observers from afar, but had to estimate observers' perceptions of actors. In this way, bystanders estimated the perceptions of someone else (just like actors), but not someone else who was judging the self (unlike actors).

Comparing actors' and bystanders' estimates of observers' perceptions is particularly informative. By our reasoning, the constricted working trait definitions come from actors considering being evaluated. Imagining how another will judge the self focuses actors on that performance judgment in observers' minds. But if such constriction emerges merely from trying

to adopt someone else's perspective (instead of the perspective of someone else who is evaluating the self), actors and bystanders should agree on their guesses of observers' judgments.

As in Study 4, we measured the perceived diagnosticity of behaviors. But unlike Study 4, we did not probe it directly. Instead, we allowed diagnosticity to be revealed through a pair of judgments. That is, we asked participants to report how they would judge the actor twice—if (hypothetically) the actor were to perform well, and then again if the actor were to perform poorly. When behaviors are particularly diagnostic of traits, they should prompt more divergent judgments in these two cases (i.e., the two judgments will have a greater difference).

Method

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

Procedure. Like in Study 4, we told participants they should fully throw themselves into every simulation—to visualize the scene unfolding, to be attuned to what they would be thinking or feeling. The scenarios were the same as those used in Study 4. The instructions for actors and observers were nearly equivalent to those used in Study 4. Participants in the new bystander condition were informed that they would consider each situation as an outside onlooker. Their task would be to predict how a person (the observer) would form judgments about another (the actor). As before, an image accompanied every scenario to reinforce the perspective manipulation.

Next, participants indicated how the observer would view the actor if the actor were successful as well as if the actor were unsuccessful. As one example, consider again the scenario in which an actor brings cookies to a party. Those in the observer condition made two

judgments: "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 perspective tried to guess how observers would respond to this question. Bystanders made similar judgments, but they did not consider that they themselves had baked the cookies; they considered how another person (the observer; Person Y) judged someone else (the actor; Person X). For the cookie scenario, bystanders saw questions of this form: "After sampling their cookies, if Person Y thinks the cookies taste *good* [*bad*], how good of a cook do you think Person Y would think Person X is?" All judgments were made on 11-point scales, anchored at 0 (*not at all…*) and 10 (*extremely…*).

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. We submitted these inferred diagnosticity scores to a 3 (perspective: actor, observer, bystander) X 10 (scenario) repeated-measures ANOVA. The predicted main effect of perspective was significant, $F(2, 298) = 7.31$, $p < .001$, $\eta^2_p = .05$. (See Table 2 for results by scenario).

We conducted a series of 2 (perspective) X 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 focal events ($M = 4.07$, $SD = 1.93$) than those in the observer perspective condition were ($M = 3.09$, $SD = 1.86$), $F(1, 299) = 13.76$, $p < .001$, $\eta^2_p = .04$. Did actors' meta-judgments see greater diagnosticity in these behaviors because actors were imagining being personally evaluated (as we have argued), or merely because they were making judgments about someone else's inferences? Providing

support for the preferred account, bystanders did not think that observers would be particularly reactive ($M = 3.36$, $SD = 1.87$). That is, their own guesses about observers' inferences showed less evidence of an overblown implications effect than did actors', $F(1, 299) = 6.88$, $p = .009$, $\eta^2_p = .02$. Instead, bystanders' guesses were fairly accurate, statistically indistinguishable from the observers', $F(1, 299) = 1.30$, $p = .26$, $\eta^2_p < .01$.

Study 6

Studies 4 and 5 helped to pinpoint what it is about actors' perspective and experience that is (i.e., being the object of evaluation) and is not (i.e., having already performed) necessary for *the overblown implications effect* to emerge. Although we framed these studies as investigations of what is responsible for constricted working trait definitions, we have yet to study the contents of these working trait definitions. In Study 6, we aimed to manipulate actors' and observers' working trait definitions. More specifically, we asked some actors and observers to list other behaviors—beyond the focal performance behavior—that could speak to the trait in question.

If actors' meta-perceptions, compared to observers' social perceptions, operate under a constricted working trait definition, then this *broadening* manipulation should debias actors' subsequent meta-perceptions. That is, by encouraging them to adopt a broader working trait definition that matches observers' baseline conceptions of the trait, this intervention should help actors avoid the overblown implications effect. Given that Studies 4 and 5 found that merely simulating the role of actor or observer was sufficient to produce the overblown implications effect, we begin by testing this intervention in a simulation scenario. The final study will return to the lab to test whether the intervention can debias judgment following actual performance behavior.

Method

Participants and design. Eight hundred seventeen participants took part in the study. Participants were randomly assigned to one of 8 conditions in a 2 (perspective: actor or observer) X 2 (feedback: success or failure) X 2 (trait definition: broadened or original) full-factorial design. In order to achieve a large sample size, we recruited participants from two sources simultaneously: an American university subject pool and Amazon's Mechanical Turk.

Procedure. Participants were asked to read a short passage about a work meeting. For those in the *actor* condition, the passage read as follows:

> Imagine that you are in a meeting at work with several co-workers. You have previously worked with each of these co-workers, and you often have meetings during which you generate ideas for new projects. You are in your second month at this job. As with most people, some of the ideas you have come up with have gone on to be successful projects and some have never taken off the ground. At today's meeting, everyone is brainstorming how best to recruit a new client who could potentially bring in a lot of work. You feel like you have identified an area that this new client could work on to improve their business, and you are very excited about making a pitch with this in mind.

At this point, it was said that the idea was either accepted (*success* condition) or rejected (*failure* condition). In both conditions, the description began, "After you have shared your idea with your team of co-workers, there is…" At that point, what happened varied by condition. The success condition continued, "…excited chatter among everyone, and it is clear that everyone is interested in following up on your idea. The rest of the meeting focuses on how to best execute your idea." The failure condition concluded with, "…a long moment of silence among everyone, and it is clear that no one is interested in following up on your idea. The rest of the meeting shifts to other people's suggestions."

The *observer* condition was parallel to the actor condition, but all descriptions were written from the perspective of another co-worker who was present at the meeting. The actor—i.e., the co-worker who shared the idea—was referred to as "Person X." All participants—both those in the actor and observer condition—were provided with a visualization that. All participants were provided with a picture to help them to both: (1) visualize the scenario, and (2) understand the perspective that they were supposed to take in the scenario (see Figure 7).

[INSERT FIGURE 7 HERE]

Broadening manipulation. At this point, participants assigned to the broadened trait definition condition completed the broadening manipulation. It made certain that participants had an expanded working trait definition of competence—calling their attention to the fact that competence is defined by more behaviors than offering ideas in a meeting. The wording for the actor [observer] condition was as follows:

> Before answering questions about this situation, we would like you to think about things that could happen outside of this situation. In a workplace, like the one described, there are many ways that employees could display their competence or incompetence. Please think of and list 5 ways that an employee like you ["Person X"] in this scenario could display competence or incompetence.

Trait perceptions. The perception measures comprised three items that assessed the actor's competence. *Observers* were asked to judge Person X in light of the meeting. *Actors* were asked to guess how another co-worker would judge them (i.e., "Person X") in light of the meeting. Three items measured perceptions of workplace competence: *competent*, *intelligent*, *creative*. Responses were made on 9-point scales anchored at 1 (*not at all*) and 9 (*extremely*). We averaged these measures to create a competence perception composite ($\alpha = .92$).

Results and Discussion

Given our predictions that original actors (i.e., those who did not complete the broadening intervention) would be unique in having a constricted working trait definition, we began by defining a *contrast* that differentiates original actors (+3) from those in the other three conditions: broadened actors (-1), original observers (-1), broadened observers (-1). This differentiated those participants who should have a working trait definition of competence that focused on the specific type of behavior described in the scenario (original actors) from those who—due to 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 two-way 2 (contrast) X 2 (feedback: success or failure) between-subjects ANOVA.

As predicted, there was a significant Contrast X Feedback interaction, such that participants in the *original actor* condition were more reactive to the focal evaluative event than were those in the other three conditions, $F(1, 813) = 8.43$, $p = .004$, $\eta^2_p = .01$ (see Figure 8). To make certain that the predicted contrast fit the data well, we conducted a series of six pairwise comparisons to understand whether each group of the four groups of participants that had been assigned the contrast codes was more or less reactive to the feedback than the other groups. Every pattern emerged as expected.

Original actors were significantly more reactive to the focal evaluative event compared to participants in the broadened actor condition, $F(1, 813) = 5.40$, $p = .02$, $\eta^2_p = .01$; participants in the *original observer* condition, $F(1, 813) = 4.82$, $p = .03$, $\eta^2_p = .01$; and participants in the *broadened observer* condition, $F(1, 813) = 7.56$, $p = .01$, $\eta^2_p = .01$. The three other

comparisons—broadened actor vs. original observer, broadened actor vs. broadened observer, original observer vs. broadened observer—were all non-significant, Fs < 1, ps > .57.

[INSERT FIGURE 8 HERE]

Study 7

Although Studies 4-6 showed that merely simulating the role of actors (asking how others *would* judge the self) is sufficient to produce the overblown implications effect, Study 7 built on the previous study by testing our broadening intervention in an actual behavioral context. On the one hand, Study 6 might be considered a particularly conservative test of our hypotheses. After all, when under actual (instead of merely simulated) evaluative threat, there may be more of a constricted working trait definition for the broadening manipulation to undo. On the other hand, one might worry that Study 6 was instead a liberal test of our hypotheses. That is, under actual evaluative threat, perhaps a minimal intervention like ours will not be sufficient to expand a more rigidly constricted working trait definition. Given our interest in examining a practical debiasing intervention in addition to testing our theoretical account, it is important to determine whether the broadening task has an influence in this situation as well.

We returned to the paradigm used in Study 1, the trivia contest. As before, audience members (observers) provided ratings of the contestants' (actors') intelligence both before and after a focal 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. We expected to conceptually replicate the effects of Study 6, that actors in the original condition would show evidence of the overblown implications effect. That is, original actors' meta-perceptions of how observers would view them should be more reactive to their focal success or failure than actors who completed the broadening manipulation. Broadened

actors should be relatively accurate compared to observers (regardless of whether such observers completed the broadening manipulation).

Method

Participants and design. Two hundred thirty-six undergraduates from an American university completed a lab session for course credit. Actors were randomly assigned to the *success* or *failure* condition. Observers were yoked to a randomly selected actor. As with Studies 2 and 3, when more than one observer was paired with an actor, we averaged the observers' responses for the purpose of analyses.

Procedure. The procedure and measures were similar to those used in Study 1. Actors began by answering ten trivia questions. Regardless of their actual performance, they were told (and their yoked observers also learned) that they answered seven correctly. Actors and observers then completed baseline perceptions (actors offered meta-perceptions; observers provided social perceptions). We used the same intelligence perception measures as in Study 1: meta-perception ($\alpha = .93$) and social perception ($\alpha = .91$). On a subsequent high-stakes double-or-nothing round, actors received success or failure feedback. Unbeknownst to them (or to the observers who saw this), this feedback was randomly determined.

Before completing the final perception measures, some participants completed the broadening manipulation. The goal was to expand these participants' working trait definition of intelligence to include additional behaviors than the one that defined the current performance context. The instructions were similar to those used in Study 6, but modified for the current context:

> 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

[masked university name] like you could demonstrate that they are or are not intelligent. Please list 5 different ways that a student from [masked university name] could demonstrate intelligence or lack thereof.

Results and Discussion

We used a nearly identical analytic approach to that used in Study 6. That is, we first defined a variable *contrast* to differentiate those participants whose perceptions were expected to be more reactive to the focal performance feedback (original actors: +3) from those hypothesized to be less reactive (broadened actors, original observers, broadened observers: -1). We then submitted the intelligence perception measures to a 2 (contrast) X 2 (feedback: success or failure) X 2 (time: baseline or final) mixed-model ANOVA, with only the final factor measured within-subjects.

The predicted Contrast X Feedback X Time interaction emerged, $F(1, 210) = 11.36$, $p < .001$, $\eta^2_p = .05$ (see Figure 9). As expected, the meta-perceptions of those in the original actor condition were more reactive to the focal evaluative feedback than were the perceptions of those in the other three conditions. That is, original actors showed more of an overblown implications effect than did those in the broadened actor condition, $F(1, 210) = 6.31$, $p = .01$, $\eta^2_p = .03$. Furthermore, their meta-perceptions were more reactive than the social perceptions of observers, regardless of whether they were in the original condition, $F(1, 210) = 4.66$, $p = .03$, $\eta^2_p = .02$, or the broadened condition, $F(1, 210) = 12.81$, $p < .001$, $\eta^2_p = .06$. The three other comparisons—broadened actor vs. original observer, broadened actor vs. broadened observer, original observer vs. broadened observer—were all non-significant, Fs < 1.61, ps > .20.

[INSERT FIGURE 9 HERE]

General Discussion

People care how others view them. But without direct access to others' perceptions, understanding how we are perceived entails guesswork. Across seven studies, we documented how and why meta-perceptions systematically err. More specifically, actors show evidence of an *overblown implications effect*—seeing their own performance as having more evaluative impact in observers' eyes than it actually does.

The first three studies used multi-stage lab paradigms to document the overblown implications effect. Each study tested perceptions of a different trait in a different context. When observers watched actors win or lose at a trivia contest (Study 1), succeed or fail during a video dating profile (Study 2), or be embraced or rejected in a social situation (Study 3), observers drew more moderate inferences abut the actors' intelligence, dateability, and likeability, respectively, than actors thought they would. Beyond demonstrating the overblown implications effect in three different performance contexts, we disentangled two reasons for why this effect emerged. It was not the case that actors merely assumed that observers possess sticky priors that fail to fully incorporate new information. Instead, actors assumed that observers would interpret the meaning of actors' focal behavior differently than observers actually did, that observers would see clear diagnosticity where they did not. In actors' minds, the implications of their own performance was overblown.

The final four studies ruled out various accounts of the overblown implications effect and provided converging evidence for our favored mechanism. We reasoned that actors—whether contemplating a future performance or engaging with one in the moment—are under evaluative threat. As actors contemplate their evaluators, they imagine observers with relatively narrow working trait definitions. This account successfully predicts several empirical patterns we observed. First, it correctly anticipates that the overblown implications effect will be seen in

prospect, even when there are no actual failures or successes to ruminate about (Study 4). Second, it correctly predicts that the overblown implications effect will describe actors' meta-perceptions about themselves (given the potential for being evaluated), but not bystanders' guesses of how observers will view actors (Study 5). Third, the account led us to a successful debiasing intervention: Broadening actors' constricted working trait definitions brought their meta-perceptions in line with social reality (Studies 6 and 7).

Although we have focused on the role that encouraged the overblown implications effect (i.e., actors being evaluated), our logic suggests that not all such actors should show equally strong evidence of the OIE. If it is indeed actors' evaluative anxiety that encourages constricted working trait definitions, then those most concerned with how they are evaluated should be those who show the strongest OIE. Public self-consciousness (PSC) has been shown to predict anxiety around being evaluated (e.g., Hope & Heimberg, 1988; Turner, Scheier, Carver, & Ickes, 1978). Given that the *overblown implications effect* was typically instantiated as a three-way interaction (Feedback X Perception X Time), we knew that we would not have the power to consistently detect four-way interactions showing that actors' public self-consciousness moderated the *overblown implications effect*. Instead, we measured PSC—using the public self-consciousness scale (Fenigstein, Scheier, & Buss, 1975)—in the five studies in which we examined whether actors or observers drew stronger inferences about observers after learning of how actors had behaved. As expected, we found that the *overblown implications effect* grew stronger as actors' public self-consciousness increased, Stouffer's Zs between 2.22 and 2.64, *p*s between .03 and .008[2] (Study 1: $Z = 1.57$; Study 2: $Z_{General} = .25$, $Z_{Specific} = -.35$; Study 3: $Z_{General} = 2.63$, $Z_{Specific}$ =

[2] Studies 2 and 3 had both general and specific final trait measures. This means that for the purpose of the cross-study meta-analysis, there are four ways to select which measure to use

2.29; Study 6: Z= -0.02, Study 7: Z = 1.47). This result provides a fourth piece of convergent evidence for our mechanistic logic.

Reconciling the Overblown Implications Effect with Previous Research

At first glance, the overblown implications effect might seem inconsistent with the actor-observer effect (Jones & Nisbett, 1971; but see Malle, 2006)—the tendency for observers to make more dispositional inferences than actors. It is important to note that it does not actually conflict with the overblown implications effect. 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., 2003)?

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 meta-perceivers thought about the trait category, not actors' explanations for the behavior. The latter, alternative explanation has difficulty explain why our broadening intervention (Studies 6-7) would debias actors. Second, and relatedly, in some of our studies—particularly our scenario studies (Studies 4-6)—the behavioral contexts were described in a vacuum—i.e., without information about how the situation may affect performance success. It is thus hard to imagine what situational contexts actors would have been relying upon that observers would have neglected. Instead, we argue it is the mere consideration of being evaluated that narrows meta-

---

from each study. The range of meta-analytic results reflects that regardless of which measure is chosen, the meta-analytic result is statistically significant.

perceivers' working trait definitions. Third, if the overblown implications effect 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 5) should also have overestimated the extent to which observers would draw inferences from actors' behavior. Such bystanders did not. Instead, the overblown implications effect seems to be operating through a novel mechanism—the constricted working trait definitions that characterized the meta-perceptions of those under evaluation.

Although we did not give much attention to actors' self-perceptions in Studies 1-3, some might be surprised that they did not show the same evidence of the overblown implications effect that actors' meta-perceptions did. This might seem inconsistent with the literature on the looking glass self, the idea that self-views derive from how (they 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 meta-perceptions, they still did show sizable shifts in light of recent performance. Furthermore, the correlations between the change in meta- and self-perception were strong (r = .63 in Study 1, r = .73 in Study 2, and r = .44 in Study 3). We, of course, cannot say whether this relationship between meta- and self-perceptions was causal; nonetheless, these findings do illustrate 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

In an effort to provide strict experimental control and reduce variability in actors' experience in our multi-stage lab studies, observers were not physically present during actors' performance. In this way, observers were merely that, observers who were unable to influence the actors' behavior. Had observers been physically present, actors could have leaned on

observers' reactions—verbal and non-verbal—when forming meta-perceptions. Though even when observers are present, observers are notoriously hard to read. For example, in the classic spotlight effect studies (Gilovich et al., 2000), 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 can attest, 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 disort meta-perceptions of the audience as a whole. Future research would be necessary to determine how the live presence of an observer or observers could either help or further hinder the accuracy of meta-perceptions.

Another constant feature of our studies was that we examined performances that were clearly successes or failures. That is, in our lab studies, we explicitly referred to the performance as a win or a loss (Studies 1 and 7), had the experimenter express surprise at the participant's exceptionally good or bad performance (Study 2), or called special attention to a dichotomous acceptance or rejection (Study 3). Moreover, in the scenario study, the explicit description necessarily called attention to and provided a characterization of an event as positive or negative (Study 6). It is possible, however, that in our daily lives, some performances may loom larger in the eyes of observer than actors. That is, a professional singer who mindlessly sings along to a radio 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. If this reasoning is correct, this may be a context in which observers' live feedback ("These cookies are fantastic!") could push actors back in the direction of the OIE.

We argued that the constricted nature of actors' working trait definitions stems from the evaluative threat posed by the performance situation. Bolstering this logic, our cross-study meta-analysis found that those most disposed to performance anxiety showed the strongest overblown implications effect. But this also suggests that there may be times in which bystanders show the overblown implications effect 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. Future research should examine whether such invested bystanders exhibit the OIE as well.

Finally, our studies investigated how actors' and observers' perceptions respond to a single focal performance event. What would happen if actors' skills—both successes and failures—are on display over multiple rounds? Do actors feel most under evaluation when they know observers do not know them well, meaning that the OIE may diminish across time? Or instead will actors' meta-perceptions respond to what is evaluatively focal, that which has just occurred (or is about to occur)? These questions are ripe for future research 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). This means meta-perceptions are important barometers of whether people (think they) are doing so effectively. And thus, when people's meta-perceptions 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). But fortunately, there may be a simple remedy for actors' overblown implication effects: calling to mind other behaviors that also define a trait. The challenge in implementing this intervention in everyday life may be the same force that produces the bias in the first place—how to get people to take their attention off the performance at hand to instead see that the present moment only captures a sliver of how others view us.

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.

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.

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.

Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, *64*, 1089-1108.

Carlson, E. N., & Furr, R. M. (2009). Evidence of differential meta-accuracy: People understand the different impressions they make. *Psychological Science*, *20*, 1033–1039.

Carlson, E. N., Furr, R. M., & Vazire, S. (2010). Do we know the first impressions we make? Evidence for idiographic meta-accuracy and calibration of first impressions. *Social Psychological and Personality Science, 1,* 94-98.

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.

Cervone, D. & Shoda, Y. (1999). Beyond traits in the study of personality coherence. *Current Directions in Psychological Science*, *8*, 27-32.

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

Cramer, A., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., & Aggen, S. et al. (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.

Critcher, C. & Dunning, D. (2015). Self-affirmations provide a broader perspective on self-threat. *Personality and Social Psychology Bulletin*, *41*, 3-18.

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.

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.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12,* 391-396

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin*, *30,* 447–460.

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, *18,* 199–212.

Epley, N., & Gilovich, T. (2006). The anchoring and adjustment heuristic: Why adjustments are insufficient. *Psychological Science*, *17,* 311–318.

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.

Fenigstein, A., Scheier, M., & Buss, A. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, *43*, 522-527.

Gallrein, A.-M. B., Carlson, E. N., Holstein, M., & Leising, D. (2013). You spy with your little eye: People are "blind" to some of the ways in which they are consensually seen by others. *Journal of Research in Personality*, *47*, 464–471.

George, L. & Stopa, L. (2008). Private and public self-awareness in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*, 57-72.

Gilovich, T., Medvec, V., & Savitsky, K. (2000). The spotlight effect in social judgment: An egocentric bias in estimates of the salience of one's own actions and appearance. *Journal of Personality and Social Psychology*, *78*, 211-222.

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.

Hope, D. & Heimberg, R. (1988). Public and private self-consciousness and social phobia. *Journal of Personality Assessment*, *52*, 626-639.

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: Lawrence Erlbaum.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, *107*, 34–47.

Malle, B. (2006). The actor-observer asymmetry in attribution: A (surprising) meta-analysis. *Psychological Bulletin*, *132*, 895-919.

Markus, H. & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, *51*, 858-866.

Markus, H. & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, *38*, 299-337.

McConnell, A. (2011). The multiple self-aspects framework: Self-concept representation and its implications. *Personality And Social Psychology Review*, *15*, 3-27.

McCrae, R. R., & Costa, P. T. J. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. Boyle, G. Matthews, & D. Saklofske (Eds.), Sage handbook of personality theory and assessment (Vol. 1, pp. 273–294). Los Angeles: Sage.

Pervin, L. (1994). A critical analysis of current trait theory. *Psychological Inquiry*, *5*, 103-113.

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.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891.

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.

Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom.* New York: Holt, Rinehart, & Winston.

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

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.

Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology*, *43*, 133–168.

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: 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.

Steele C.M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52,* 613–629.

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.

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: Guilford Press.

Turner, R., Carver, C., Scheier, M., & Ickes, W. (1978). Correlates of self-consciousness. *Journal of Personality Assessment*, *42*, 285-289.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

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.

Zajonc, R.B. (1965). Social facilitation. *Science, 146,* 269-274.

Table 1
Diagnosticity Ratings by Perspective for Each Trait-Relevant Behavior (Study 4).

Context	Behavior	Trait	Observer	Actor Rating	Observer Rating	*t*
Restaurant with a group	Splitting the bill	Mathematical ability	Person looking over actor's shoulder	6.17 (2.28)	6.96 (2.23)	-2.56*
Cocktail party	Conversing with stranger	Social skills	Person overhearing actor's conversation	6.74 (2.20)	7.08 (2.16)	-1.13
Game night at a friend's house	Answering a trivia question	Intelligence	Person reading actor the question	6.02 (2.45)	5.69 (2.31)	0.99
On a flight	Playing chess on your computer	Analytical thinking ability	Person next to actor on the flight	6.22 (2.24)	5.68 (2.25)	1.75†
Cash-only restaurant	Remembering to pay back the $20 you borrowed	Exploitativeness	Acquaintance that lent actor the money	6.60 (2.62)	5.66 (2.55)	2.64**
Party	Baking cookies	Cooking ability	Person sampling actor cookie	7.43 (2.01)	6.58 (2.27)	2.90**
Near work right before a meeting	Parallel parking	Driving ability	Coworker waiting for actor to walk in together	7.85 (2.36)	6.86 (2.58)	2.96**
Restaurant with a group	Accepting/ rejecting a fork for dessert	Self-control	Person that knows actor is on a diet and asks if actor wants a fork	6.75 (2.64)	5.54 (2.30)	3.54***
Party	Introducing a new person to your friend	Inconsiderateness	New person actor is introducing, whom actor just met and had a conversation with	5.87 (2.68)	4.41 (2.54)	4.09***
Office	Leaving work at an unusual time	Work ethic	Coworker asking actor for a ride home at usual time	6.48 (2.50)	4.82 (2.67)	4.71***
Total:				6.61 (1.33)	5.93 (1.42)	13.29***

Note. Ratings indicate means and standard deviations (in parentheses); † $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$,

Table 2
Trait Inferences by Perspective For Hypothetical Success or Failure on Each Trait-Relevant Behavior (Study 5).

	Diagnosticity (Positive – Negative)			Successful Outcome			Failed Outcome		
Trait	Actor	Observer	Bystander	Actor	Observer	Bystander	Actor	Observer	Bystander
Mathematical ability	4.22 (2.52)	3.78 (2.85)	4.05 (2.60)	8.08 (1.58)	8.08 (1.61)	8.12 (1.43)	3.57 (2.94)	4.31 (2.65)	4.05 (3.16)
Social skills	3.51 (2.65)	3.90 (2.78)	3.45 (3.07)	7.84_{a} (1.54)	8.32_{b} (1.61)	7.84_{a} (1.63)	4.33 (1.70)	4.43 (1.90)	4.39 (1.83)
Intelligence	3.18_{a} (2.18)	1.82_{b} (2.15)	2.36_{b} (2.43)	8.37_{a} (1.20)	7.83_{b} (1.51)	7.96_{b} (1.59)	5.19_{a} (1.52)	6.01_{b} (1.40)	$5.60_{a,b}$ (1.62)
Analytical thinking ability	3.30_{a} (2.41)	2.30_{b} (2.64)	$2.77_{a,b}$ (2.07)	8.42 (1.51)	8.21 (1.61)	8.24 (1.64)	5.12_{a} (1.62)	5.92_{b} (1.70)	$5.48_{a,b}$ (1.51)
Exploitativeness	5.21_{a} (3.82)	4.02_{b} (3.72)	$4.41_{a,b}$ (3.46)	7.89_{a} (2.33)	7.05_{b} (2.23)	7.18_{b} (2.35)	2.68 (2.40)	3.03 (2.76)	2.77 (2.35)
Cooking ability	5.17 (2.67)	4.79 (3.24)	5.13 (2.72)	8.67 (1.16)	8.66 (1.44)	8.64 (1.31)	3.49 (2.19)	3.87 (2.42)	3.51 (1.81)
Driving ability	4.43 (2.85)	3.80 (2.72)	3.95 (2.67)	9.01 (1.61)	8.96 (1.53)	8.87 (1.44)	4.58_{a} (1.92)	5.17_{b} (2.04)	$4.93_{a,b}$ (1.83)
Self-control	3.89_{a} (3.24)	3.03_{b} (3.02)	2.09_{c} (3.14)	8.70_{a} (1.84)	8.56_{a} (1.76)	7.32_{b} (1.85)	4.81_{a} (2.23)	5.53_{b} (1.86)	$5.23_{a,b}$ (1.71)
Inconsiderateness	3.05_{a} (4.14)	$.09_{b}$ (3.32)	1.64_{c} (3.92)	6.62_{a} (2.41)	4.92_{b} (2.38)	5.69_{c} (2.26)	3.57 (2.94)	4.31 (2.65)	4.05 (3.16)
Work ethic	4.72_{a} (2.80)	3.37_{b} (2.93)	$3.76_{a,b}$ (3.13)	9.34_{a} (1.49)	8.67_{b} (1.71)	8.69_{b} (1.77)	4.63 (2.03)	5.30_{b} (1.91)	$4.94_{a,b}$ (1.94)
Total	4.07_{a} (1.93)	3.09_{b} (1.86)	3.36_{b} (1.87)	8.29_{a} (.95)	7.93 (.92)	7.86 (.93)	4.23_{a} (1.29)	4.79_{b} (1.41)	$4.50_{a,b}$ (1.25)

Note. Ratings indicate means and standard deviations (in parentheses). Means in the same row that reflect the same type of score (diagnosticity, successful outcome, failed outcome) but that do not share the same subscripted letter differ at the $p < .05$ level.

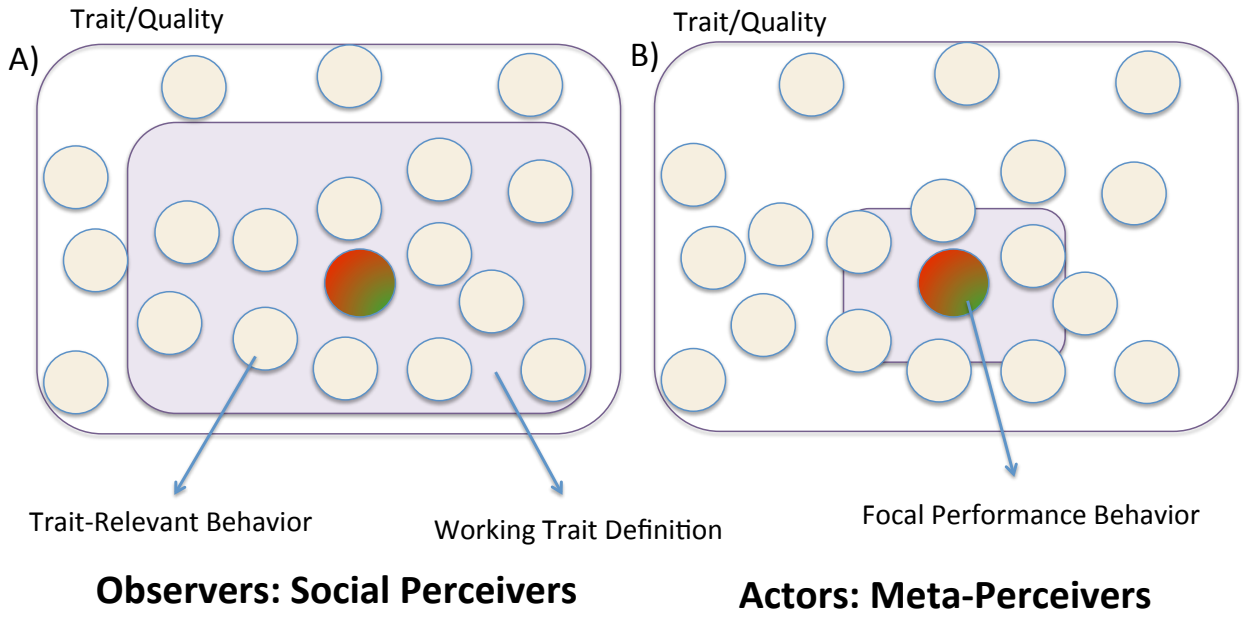


Figure 1. Why social and meta-perceptions are hypothesized to diverge and produce the overblown implications effect. At baseline, observers' working trait definitions may not account for all trait-relevant behavior (explaining why some trait-relevant behaviors are always outside of the working trait definition; panel A), but actors' meta-perceptions constrict around their focal performance behavior (panel B). This predicts that actors' meta-perceptions will be more reactive to their own focal successes or failures than observers' social perceptions, but that expanding actors' working trait definitions should debias them.

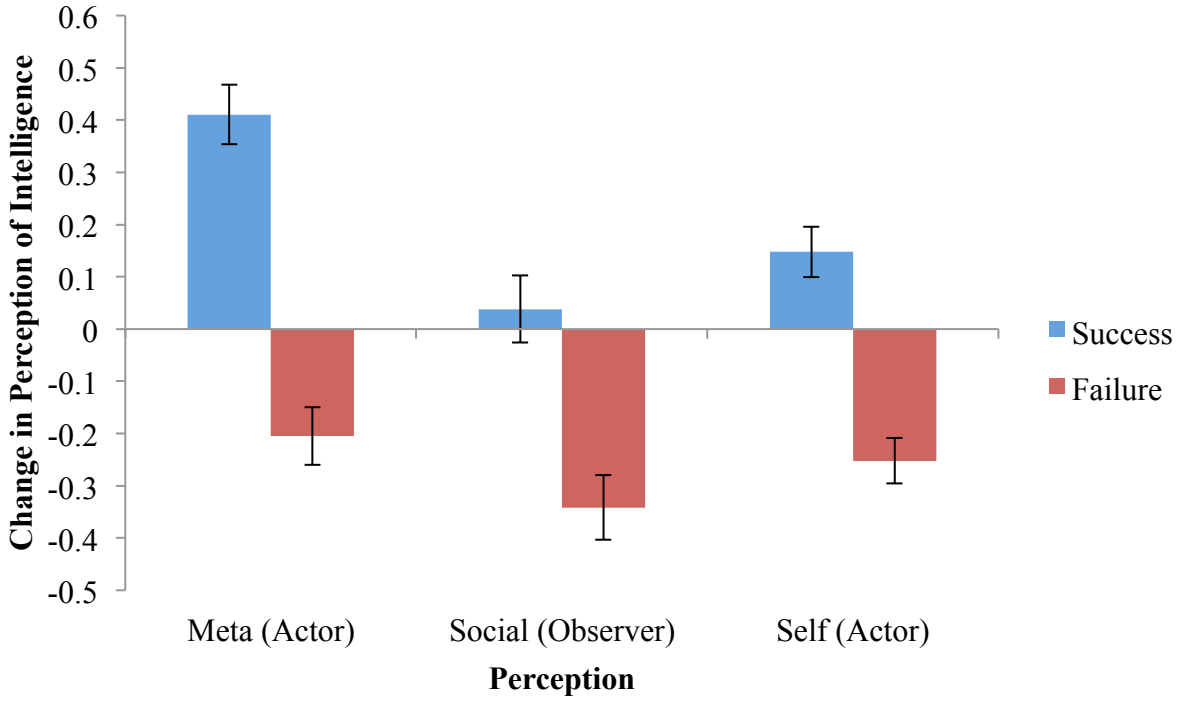


Figure 2. The change in perception (final – baseline) of actors' intelligence by feedback condition and type of perception (Study 1). Which participant offered each perception is in parentheses. The overblown implications effect is reflected by the larger gap between the two meta-perception bars compared to the two social perception bars.

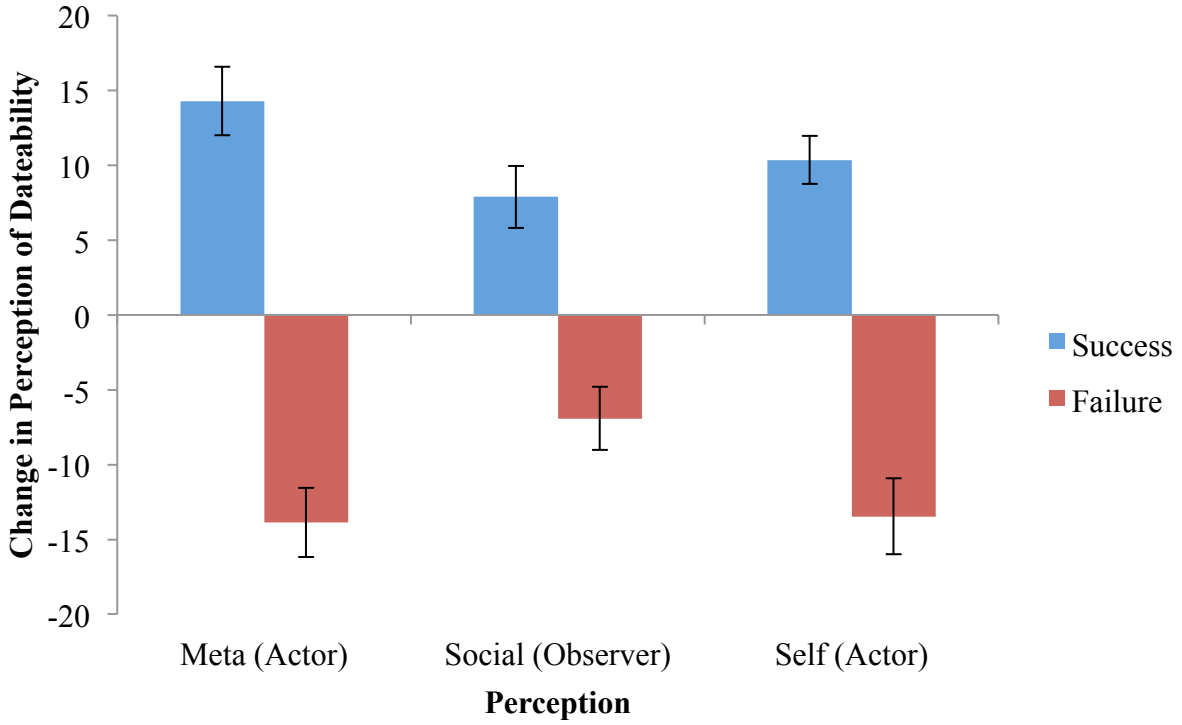
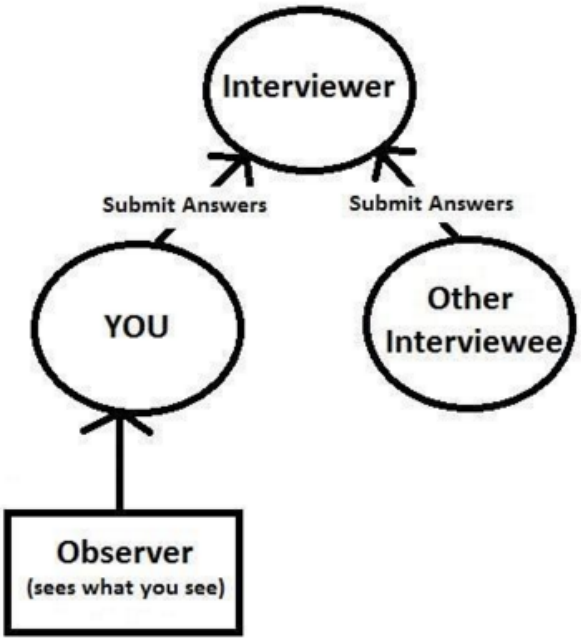
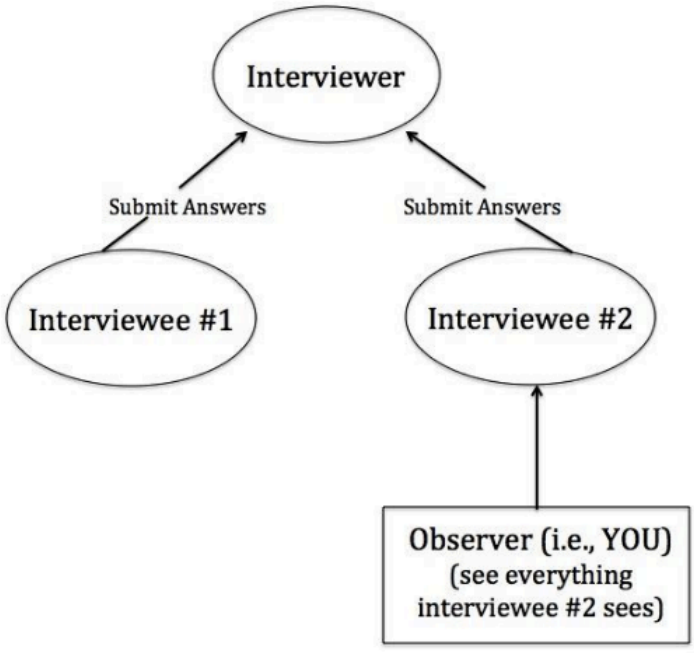


Figure 3. The change in perception (final – baseline) of actors' dateability by feedback condition and type of perception (Study 2). Which participant offered each perception is in parentheses. The overblown implications effect is reflected by the larger gap between the two meta-perception bars compared to the two social perception bars.



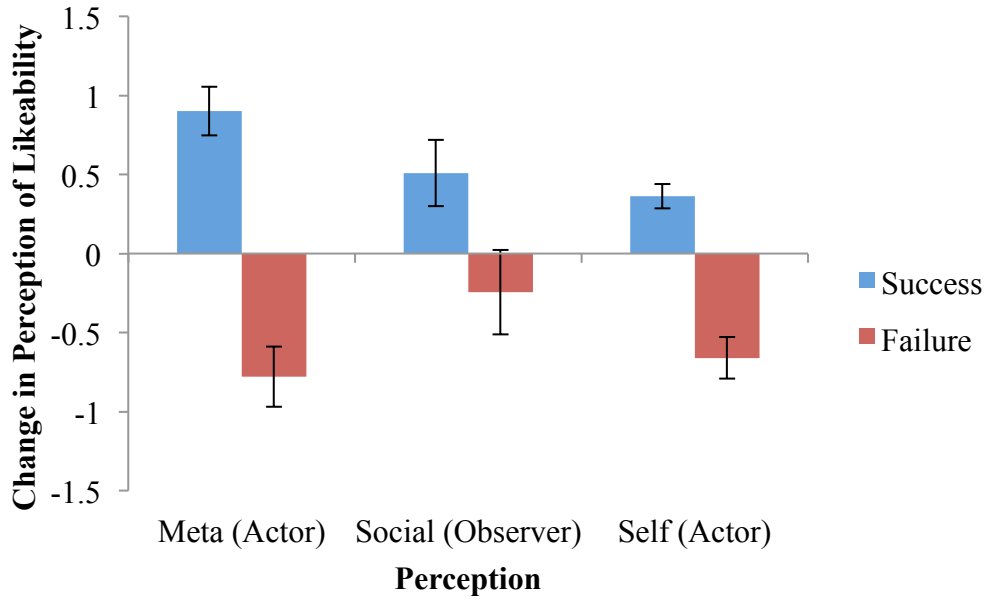
A)



B)

Figure 4. Diagrams explaining to actors (A) and observers (B) their role. 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”).

A)



B)

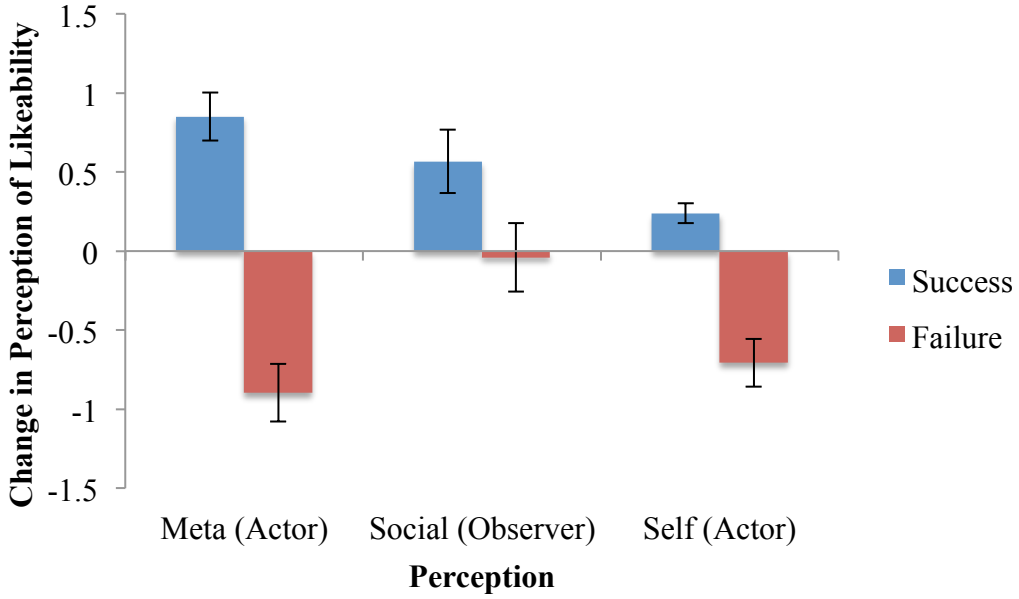


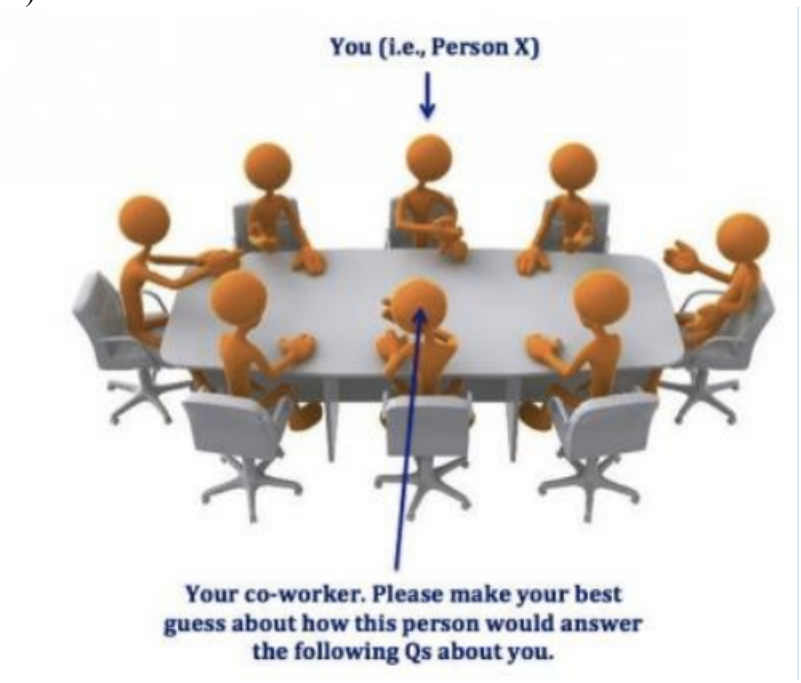
Figure 5. The change in perception (final – baseline) of actors' likeability by feedback condition and type of perception (Study 3). Panel A uses the specific final perceptions. Panel B uses the general final perceptions. Within each panel, the overblown implications effect is reflected by the larger gap between the two meta-perception bars compared to the two social perception bars. Which participant offered each perception is in parentheses.

A)

B)

Figure 6. Picture accompanying the baking cookies scenario (Studies 4-5). Actors saw Panel A. Observers and bystanders (Study 5 only) saw Panel B.

A)



B)

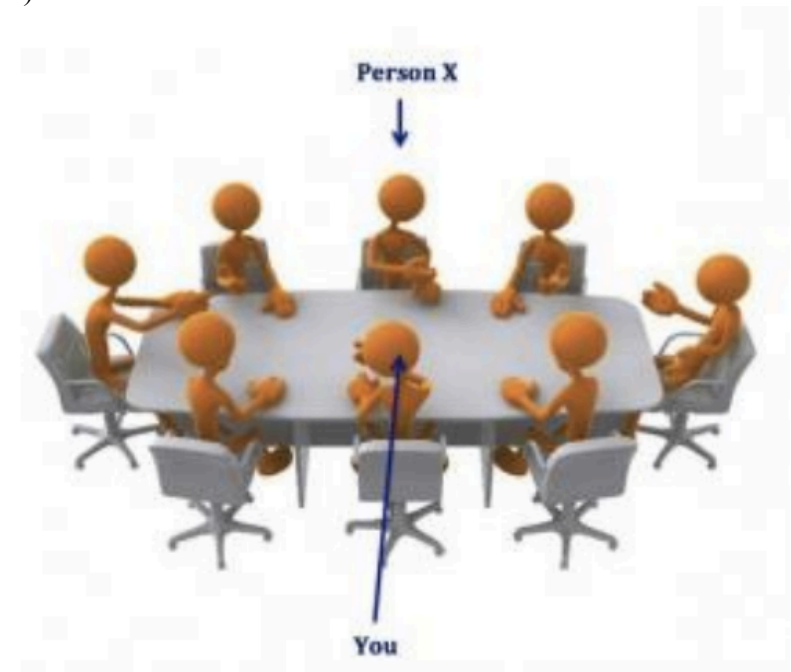


Figure 7. Graphic seen by observers (A) and actors (B) in Study 6 to clarify the trait perception task. Observers made judgments of the described actors (Person X). Actors estimated how an observer would view them.

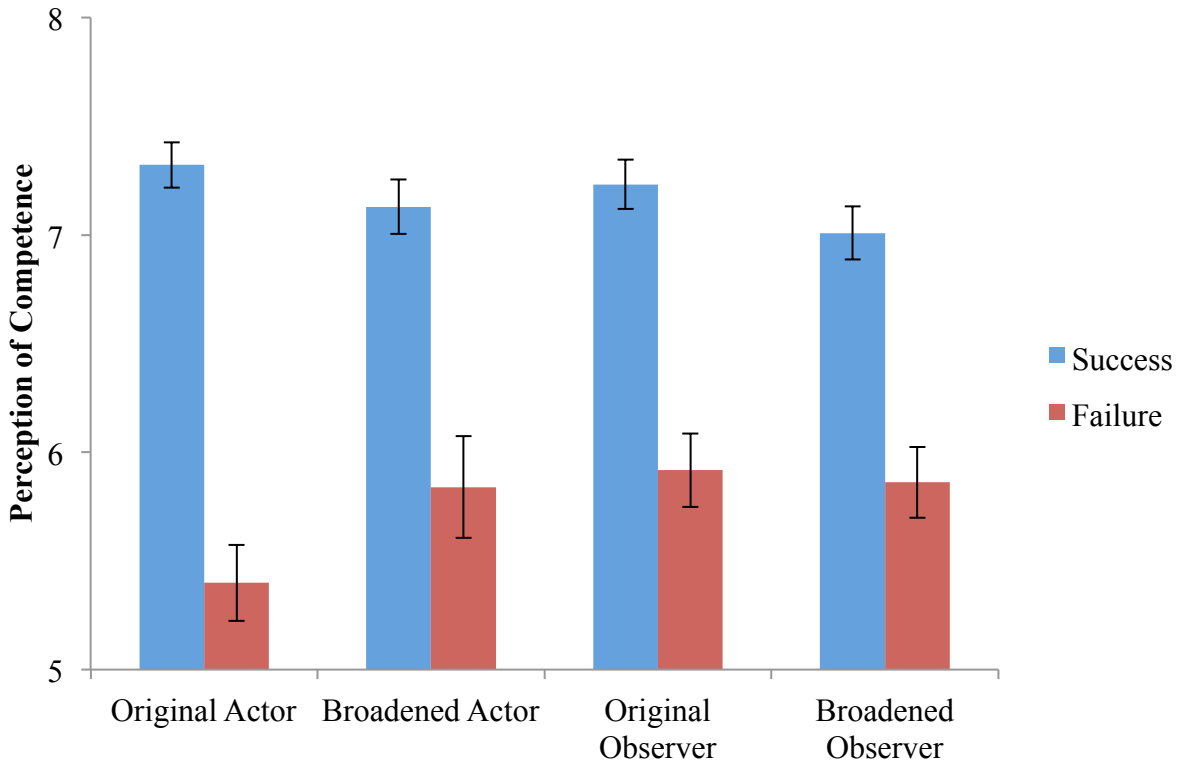


Figure 8. Perceptions of actors' competence by feedback condition, perspective, and working trait definition intervention (Study 6). The broadening intervention eliminated the overblown implications effect, as seen by the greater gap between the two bars for original actors compared to gap between the two bars for the other three perspective-intervention combinations.

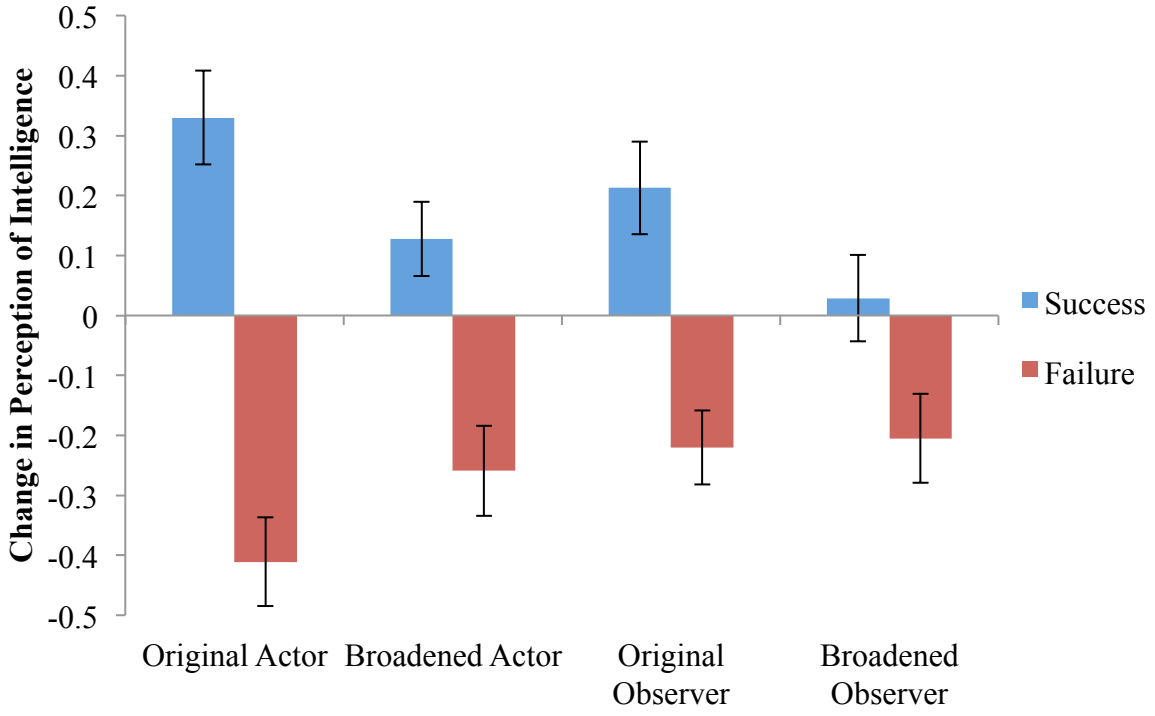


Figure 9. The change in perception (final – baseline) of actors' intelligence by feedback condition, perspective, and working trait definition intervention (Study 7). The broadening intervention eliminated the overblown implications effect, as seen by the greater gap between the two bars for original actors compared to gap between the two bars for the other three perspective-intervention combinations.