



Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot ☆

Emily Pronin *, Matthew B. Kugler

Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544, USA

Received 3 September 2005; revised 16 May 2006

Communicated by Fiedler

Abstract

People see themselves as less susceptible to bias than others. We show that a source of this *bias blind spot* involves the value that people place, and believe they should place, on introspective information (relative to behavioral information) when assessing bias in themselves versus others. Participants considered introspective information more than behavioral information for assessing bias in themselves, but not others. This divergence did not arise simply from differences in introspective access. The blind spot persisted when observers had access to the introspections of the actor whose bias they judged. And, participants claimed that they, but not their peers, *should* rely on introspections when making self-assessments of bias. Only after being educated about the importance of nonconscious processes in guiding judgment and action—and thereby about the fallibility of introspection—did participants cease denying their relative susceptibility to bias.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Bias blind spot; Introspection illusion; Self—other; Nonconscious influences; Self-perception

In a recent news story, a federal judge defended his ability to be objective in handling a case involving a long-time friend. In another story, government scientists defended their impartiality in evaluating drug companies from whom they received large consulting fees. In each of these stories, the relevant actors were convinced of their own objectivity, while outside observers were quick to accuse them of bias. Such accusations of bias in people who are confident in their objectivity are as common in everyday life as they are in the news. When our colleague judges the average work of his friend as better than the stellar accomplishments of a mere acquaintance, or when our neighbor argues that the new bright-red fire hydrant should go in front of any house but hers, we are struck by those individuals' blindness to their own biases.

The tendency to see bias in others, while being blind to it in ourselves, has been shown across a range of cognitive and motivational biases (for a review, see [Pronin, Gilovich, & Ross, 2004](#)). This article concerns a related question: *Why* do people show this bias blind spot?

Introspection and bias perception

This research explores one possible mechanism contributing to the tendency for people to acknowledge bias more readily in others than in themselves. This mechanism involves the value that people assign to introspective information, relative to behavioral information, when drawing conclusions about the presence or absence of bias. Consider the following example. In judging whether you have been biased in your hiring of a new associate, your introspections might involve the feeling that you tried to be objective in reviewing the applicants or that you never felt any bias clouding your assessments. Your behavior might be that you hired an old college buddy. As this example portends, introspections will be less likely than behavior to yield evidence of

☆ We thank Elana Jacobs, Jessica Karpay, and John Fleming for research assistance. We thank Susan Fiske, Kathleen Kennedy, and Joshua Rabinowitz for helpful comments on this research. Portions of this research were supported by a grant to Pronin from the NASD.

* Corresponding author.

E-mail address: eppronin@princeton.edu (E. Pronin).

bias. Biases typically operate nonconsciously, thereby leaving their influence hidden from introspection (Wilson & Brekke, 1994; Wilson, Centerbar, & Brekke, 2002).

Access to introspections

It is a truism that we have far more access to the contents of our own introspections than others have to our introspections. We generally know what we are thinking and feeling, and what we are intending to do and hoping to accomplish, better than others know these things about us. This rich introspective access puts us in a unique position, though not always a better one, for understanding our past attitudes and actions and for predicting our future ones (Jones & Nisbett, 1972; Nisbett & Ross, 1980, chapter 9). One might expect, then, that the proposed self—other difference in weighting of introspective information is simply a matter of differences in introspective access. This would suggest that if one knew a person's bias-related introspections (such as those of the boss who hired his college buddy), one's judgment of that person's bias would resemble his or her own.

Valuation of introspections

The hypothesis of the current research, however, differs from this account involving introspective access. We suggest, for example, that knowing the boss's desire to be fair and his faith in the talents of an old buddy (and knowing his *absence* of any willful bias) will not radically change one's perception of bias. Instead, we predict that people's lack of reliance on others' introspections is in part due to a diminished *valuation* of those introspections. When assessing bias in someone else, people are likely to view behavior (combined with theories of what biased behavior looks like) as another important source of information. This hypothesized mechanism might be referred to as an "introspection illusion," since it involves an illusion about the relative diagnostic value of one's own versus others' introspections (Pronin, Gilovich et al., 2004). In particular, it suggests that people over-value thoughts, feelings, and other mental contents, relative to behavior, when assessing their own actions, motives, and preferences, but not when assessing others'. The term introspection illusion thus involves a self—other asymmetry in the relative valuation of introspective versus behavioral information. Although the term uses the word "introspection" rather than "behavior," the illusion inherently involves a tendency for the self, but not others, to disregard behavior (and to place unwarranted value on introspections).

Previous research

Previous research has demonstrated the tendency for people to make inaccurate self-predictions and self-assessments when introspective information is available but misleading. This phenomenon has been shown for introspective information involving intense emotional reactions (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998),

careful reasoning (Wilson, Hodges, & LaFleur, 1995), positive intentions (Epley & Dunning, 2000; Kruger & Gilovich, 2004), optimistic plans (Buehler, Griffin, & Ross, 1994), personal strength versus weaknesses (Kruger, 1999), wishful thinking (Pronin, Wegner, McCarthy, & Rodriguez, 2006), and even the *absence* of seemingly relevant introspective information (Latane & Darley, 1968; Nisbett & Wilson, 1977a, 1977b). The thesis of this article rests in part on people's lack of awareness of the limitations of their own introspections, and on people's consequent over-valuation of their own (but not others') introspective contents.

The studies reviewed above converge on the idea that actors' consideration of internal information leads them to make different assessments than observers. These studies have generally assumed that those different assessments are attributable to differences in introspective access. This assumption has also been made for other classic self—other asymmetries, such as the tendency for people to describe themselves in terms of private and unobservable qualities but to describe others in terms of public and observable ones (McGuire & McGuire, 1986; Prentice, 1990), and the tendency for people to describe social interactions in terms of private reactions in the case of self but public responses in the case of others (Malle & Pearce, 2001).

We do not wish to argue that differences in awareness of internal information play no role in the effects described above. What we would like to argue, however, is that differences in valuation of internal information also play a role—particularly in the case of bias perception. Thus, when a person tells us she is unbiased in taking 75% of the credit for a project completed by a team of six peers, we are likely to consider her assessment biased even if we have access to her thoughts about how much she contributed and about her efforts to provide a fair assessment.

Researchers have provided evidence suggesting that the tendency to perceive bias more in others than the self is rooted in differences in the information that people use in making assessments of bias in the self versus others (Ehrlinger et al., 2005; Pronin, Gilovich et al., 2004; Van Boven et al., 1999). However, experiments have not examined the role of valuation in this effect. Yet, some research suggests that people may value internal information more, and observable actions less, when assessing the self relative to others. People have been shown to view themselves as defined by their thoughts, feelings, goals, fears, and other unobservable qualities, even while they are inclined to view others as defined by their observable behavior (Andersen & Ross, 1984; Pronin, Kruger, Savitsky, & Ross, 2001). People have also been found to view their unobservable intentions as more indicative than their observable actions for assessing their own (but not others') traits (Kruger & Gilovich, 2004).

Mechanism underlying the mechanism

Why might people value their own introspections more than others' introspections? Or, to put it another way, why might they disregard information about their behavior even

when it is available, and instead give undue weight to introspective “evidence”? One reason could be that a reliance on introspective evidence for assessing bias tends to elicit more flattering assessments than a reliance on behavioral evidence, and people are motivated to see themselves in a positive light. People also may place a good deal of weight on their introspections because they truly are unduly confident in them. People often fail to recognize that while they do have access to their mental contents, they do not have access to the mental processes that underlie those contents and therefore to whether those processes involve bias (Nisbett & Wilson, 1977a). We do not view these two sources of our hypothesized mechanism as in conflict with each other. Indeed, people’s reliance on introspective evidence (at the expense of behavioral evidence) for assessing bias is likely to involve an interplay of cognitive and motivational factors (e.g., Dunning, Meyerowitz, & Holzberg, 1989; Kunda, 1987). People may be motivated to rely on introspections because those yield less evidence of bias, but they are likely to feel that this reliance is defensible because of their undue confidence in the reliability of their introspections. If they were aware of the fallibility of their introspections, we suggest, their efforts to view themselves in a positive light (by turning to introspections as a source of information about bias) would be thwarted.

The present research

This article examines the hypothesis that an introspection illusion – i.e., an over-valuing of introspections relative to behavior for self, but not others – contributes to the bias blind spot. Study 1 seeks to demonstrate that people report considering introspective information more than behavioral information when assessing their own biases, but not others’ biases. Study 1a aims to show that a similar self–other asymmetry is apparent in people’s judgments of how *valuable* introspective versus behavioral information is for assessing bias in the self versus others. Studies 2 and 3 examine whether people ignore behavioral information when assessing their own bias, but not their peers’ bias, and whether asymmetric perceptions of bias persist even when others know their peer’s bias-related introspections. Study 4 examines people’s lay definitions of bias in an investigation of whether those definitions differ (in terms of their emphasis on thoughts vs. actions) depending on whether participants are primed to think about self vs. others. Finally, Study 5 educates participants about the importance of nonconscious processes and then examines whether this suggestion of the weak value of introspection diminishes the bias blind spot.

Study 1: Reporting bias assessment

This study sought preliminary evidence of a self–other asymmetry in how people go about assessing bias. We predicted that participants would report considering information about thoughts more than behavior in the case of self-assessments, but not other-assessments.

Method

Participants

A total of 247 Harvard University undergraduates participated in exchange for course credit or candy. They responded to questions either about the self-serving bias ($n = 150$) or about both the positive halo effect and the fundamental attribution error ($n = 97$).

Procedure and questionnaire

Participants completed a questionnaire that began by describing one of three different biases. The descriptions were taken from earlier research on the bias blind spot (Pronin, Lin, & Ross, 2002). For example, the description of the self-serving bias was:

Psychologists have claimed that some people show a “self-serving” tendency in the way they view their academic or job performance. That is, they tend to take credit for success but deny responsibility for failure; they see their successes as the result of personal qualities, like drive or ability, but their failures as the result of external factors, like unreasonable work requirements or inadequate instruction.

After reading one of these descriptions, participants were asked a series of questions either about themselves or about the average Harvard student (depending on experimental condition). First, they were asked: “To what extent do you believe that [you show/the average Harvard student shows] this effect or tendency?” (1 = *Not at all*, 5 = *Somewhat*, 9 = *Strongly*). They were then asked how much they used two different strategies in arriving at this assessment (using the same scale as the previous question). The first strategy involved “trying to [‘get inside my head’/‘get inside the heads’ of particular Harvard students] to find evidence of the sorts of thoughts and motives that could underlie this tendency.” The second strategy involved “considering how well this description fits the way that people in general tend to behave.”

Results and discussion

Participants showed a bias blind spot. They reported being far less susceptible to the relevant biases than their collegiate peers ($M_s = 5.05$ vs. 6.83), $F(1, 245) = 76.45$, $p < .0001$. This blind spot was evident for all three individual biases: the self-serving bias ($M_s = 4.99$ vs. 7.12), $F(1, 148) = 56.52$, $p < .0001$, the halo effect ($M_s = 6.07$ vs. 6.80), $F(1, 95) = 5.40$, $p = .02$, and the fundamental attribution error ($M_s = 4.19$ vs. 5.78), $F(1, 95) = 17.47$, $p < .0001$.

More relevant to our present concerns, participants reported using different information depending on whether it was their own or their peers’ bias that they had assessed. Those who assessed their own bias reported considering thoughts and motives more than human behavior in general ($M_s = 5.17$ vs. 3.91), $F(1, 95) = 18.39$, $p < .0001$. By contrast, those who assessed their peers’ bias reported considering thoughts and motives less than behavior in

general ($M_s = 3.48$ vs. 6.10), $F(1, 97) = 57.31$, $p < .0001$. This interaction was significant, $F(1, 244) = 70.31$, $p < .0001$, and it emerged for all three biases: self-serving, $F(1, 147) = 53.95$, $p < .0001$, halo, $F(1, 94) = 13.58$, $p = .0006$, and fundamental attribution error, $F(1, 94) = 8.07$, $p = .007$.

Finally, we expected that the information participants reported considering when making their assessments of bias would be predictive of the amount of bias they imputed. We expected that the tendency to give heavy weight to one's assumptions about relevant thoughts and intentions, rather than to assumptions about how people behave, would yield lower assessments of bias (since evidence of bias rarely emerges in conscious thought). This prediction was supported. The more participants reported considering thoughts and motives relative to human behavior (as measured by a difference score between the two), the less bias they imputed, $r(247) = -.43$, $p < .0001$. This correlation was apparent, albeit to somewhat differing degrees, when participants assessed their own susceptibility to bias, $r(132) = -.19$, $p = .07$, and when they assessed their peers' susceptibility, $r(115) = -.29$, $p = .003$. And, participants' reports of considering thoughts versus behavior mediated the observed bias blind spot, according to the Sobel test advocated by Baron and Kenny (1986), $z = 3.06$, $p = .002$.

Does individuals' consideration of internal information versus observable behavior (and general ideas about what biased behavior "looks like") simply reflect natural differences in the availability of information? We suggest that it may be rooted in people's beliefs about the probative value of information. In Study 1a, we sought to more directly investigate whether the self–other difference in reported consideration of internal versus behavioral information is accompanied by a self–other asymmetry in the perceived value of these two sources of information.

Study 1a: Assessing bias assessments

Rather than rating how much they used different sources of information in assessing bias, participants rated how *valuable* each of those sources would be for making self-assessments of bias. We predicted that they would report that considering introspections was the better strategy for assessing personal bias when it was themselves making that assessment, but that considering behavior was the better strategy for peers making such self-assessments.

Method

Participants

A total of 83 Harvard undergraduates participated in exchange for candy.

Procedure and questionnaire

Participants completed a questionnaire that began with the same description of the "self-serving tendency" used in Study 1. They were then asked to consider either themselves or an average Harvard student (depending on experimental

condition): "Imagine that [you have/an average Harvard student has] to provide an accurate assessment of how susceptible [you are/he or she is] to the above tendency. How valuable do you think it would be for [you/that person] to use each of the following strategies?" They were provided with descriptions of the two strategies described in Study 1 and asked to rate each in terms of how valuable it would be if they (or an average Harvard student) had to make the relevant self-assessment (1 = *not at all*, 9 = *strongly*).

Results and discussion

Participants showed an asymmetry in their perceptions of what information would be most valuable for their own versus others' self-assessments of bias. They reported believing that getting inside their head was a more valuable strategy than considering their beliefs about human behavior for arriving at their own self-assessment of bias ($M_s = 6.50$ vs. 5.05), $F(1, 40) = 16.19$, $p = .0002$. By contrast, when contemplating the average Harvard student assessing his or her own bias, they tended to believe that it would be less useful for that person to try to get inside his/her head than for that person to look to how people in general behave ($M_s = 5.34$ vs. 6.02), $F(1, 39) = 3.77$, $p = .06$. This interaction was significant, $F(1, 81) = 17.93$, $p < .0001$.

Participants in this study made very different claims about what information they versus their peers should use for assessing personal bias. This result suggests that Study 1 participants' reported consideration of internal information when assessing their own bias, but of behavioral information when assessing others' bias, did not simply reflect an asymmetry in availability but rather reflected an asymmetry in beliefs about which source of information was more probative.

These studies involved fairly abstract assessments of bias. Participants rated their own and others' commissions of bias in general rather than with respect to a specific instance of potential bias. In addition, participants were asked about their reliance on observations of human behavior "in general" rather than about their observations of the actions of the target being assessed. Study 2 sought to investigate the proposed underlying mechanism with respect to a specific commission of bias associated with a specific action taken by a specific actor.

Study 2: Examining others' introspections

Actor participants listed their ongoing thoughts about how they rated on various traits before rating themselves on those traits. Yoked observers saw either an actor participant's thoughts and self-ratings, or only that actor's self-ratings. Finally, actors and observers assessed the level of "better-than-average" bias in actors' ratings. We expected that actors' assessments of bias would be unaffected by the degree of bias apparent in their behavioral responses but that observers' assessments would be affected by that information—even when they had access to a sample of the actors' relevant introspections.

Method

Participants

Ninety-six Princeton University undergraduates participated in exchange for course credit or candy.

Procedure

Self-assessment condition. Upon agreeing to participate, participants were told that they would be asked to rate themselves on various personality characteristics. They were further told:

We are also interested in your thoughts while you think about these ratings, and we want you to write down all of your thoughts. Record all of your thoughts as they go through your head. The thoughts do not have to be sentences or even fully formed; they can be words or notes, whatever goes through your head.

To familiarize participants with the thought-listing task, they were shown a sample hand-written thought-listing, regarding the question: “How much do you like soda (relative to the average Princeton student)?” They then were asked to record their thoughts, and rate themselves, on four traits. Each thought-listing was preceded with a question—viz., “How [considerate/deceptive/able to get along with others/snobby] are you (relative to the average Princeton student)?” and then a half-page box labeled “Your thoughts.” Below each box, they were asked to rate themselves on the trait (1 = *Much less than average Princeton student*, 2 = *Somewhat less than average Princeton student*, 3 = *Same as average Princeton student*, 4 = *Somewhat more than [etc.]*, and 5 = *Much more than [etc.]*). Finally, they completed our measures.

Other-assessment conditions. Upon agreeing to participate, participants were told that they would be asked to read another student’s ratings of him or herself on several personality characteristics. Participants who had been randomly assigned to see a sample of their yoked actor’s introspections were also informed that they would see a report of that student’s thoughts before he or she provided the ratings. Those participants were also provided with the verbatim thought-listing instructions that we used. Participants were provided with a yoked actor’s self-ratings and, depending on condition, thought-listings. They then completed our measures.

Dependent measures

The dependent measures began by informing participants about the “better-than-average effect” in self-ratings. Participants were told that “75–85% of people rate themselves as ‘better-than-average’ relative to others within their group.” They were told that the way people do this is that: “When it comes to positive traits, they rate themselves as having these more than average, and when it comes to negative traits, they rate themselves as having these less than average.”

The questions that followed were identical across the three conditions, except for minor wording changes prompting actors to assess their own trait ratings and prompting observers to assess a yoked actor’s ratings. Participants were told: “We are interested in obtaining the most accurate personality assessments possible. We would like to know—to the best of your assessment—whether you think [your/the student’s] personality ratings were affected by this tendency.” They then were asked: “Do you think that [your/the student’s] personality ratings were affected by this tendency?” (*No, I don’t think they were* vs. *Yes, I think they were*), and: “To what extent would you say [your/the student’s] ratings were affected by this tendency?” (1 = *Not at all*, 7 = *Very much*). Finally, two questions probed for what information they considered when making the above assessments: “How much did you take into account the thoughts that [you/the student] had before providing each of [your/his/her] ratings?” and “How much did you take into account whether [your/the student’s] personality ratings claim [you/him/her] to be better than the average Princeton student?” (1 = *Not at all*, 7 = *Very much*).

Results and discussion

Better-than-average bias

We first sought to confirm that participants showed the better-than-average effect. To do that, participants’ self-ratings across the four traits were summed, after first re-scoring their 5-point scale ratings such that responses of “same” as average were coded as 0, ratings of “somewhat” better than average were coded as +1, ratings of “much” better were coded as +2, ratings of somewhat worse were coded as –1, and ratings of much worse were coded as –2. Consistent with past research, participants viewed themselves as better than their collegiate peers ($M = 3.18$), $t(30) = 8.59$, $p < .0001$. In fact, only one participant out of 32 failed to rate him or herself as better-than-average on the majority of the traits!

Bias blind spot

Participants showed a bias blind spot in recognizing their commissions of bias. While bias was imputed by a full 77% of participants who read a fellow student’s ratings, only 52% of participants who provided self-ratings imputed such bias, $\chi^2(1, N = 96) = 5.53$, $p = .02$. The blind spot was also evident on our more continuous measure, where actors imputed less bias to themselves than observers imputed to them ($M_s = 3.28$ vs. 4.45), $F(1, 62) = 10.21$, $p = .002$.

Attention to introspections

We next conducted analyses of our introspection illusion hypothesis. First, we predicted that actors, but not observers, would report making bias assessments based on introspective information (i.e., their thoughts before providing the ratings) rather than behavioral information (i.e., how much they rated themselves as better-than-average). Consistent with this hypothesis, actors reported considering

introspections more than behavior in assessing whether they were biased ($M_s = 4.25$ vs. 3.52), $F(1, 30) = 6.29$, $p = .02$. Observers showed no such privileging of introspections, and leaned towards considering introspections less than behavior ($M_s = 4.38$ vs. 4.80), $F(1, 62) = 2.58$, $p = .11$. This interaction was significant, $F(1, 94) = 10.78$, $p = .001$.

Analyses involving participants' actual assessments of bias provided further evidence for these self–other differences in information consideration. Whereas the amount of bias evident in actors' ratings was not correlated with actors' self-assessments of bias, $r(30) = .17$, $p = .34$, it was correlated with observers' assessments of bias, $r(62) = .48$, $p < .0001$. Thus, actors ignored their behavior in assessing their bias, whereas observers used that information (see Fig. 1).

Based on the hypothesized introspection illusion, we further predicted that observers would impute a large degree of bias to actors (relative to how much bias the actors saw in themselves) even when the observers were given access to actors' ongoing thoughts. Consistent with this hypothesis, observers' assessments of bias did not differ depending on whether they had such access, $F < 1$ (see Fig. 2). Moreover, observers perceived more bias in actors than actors did in themselves regardless of whether or not those observers were provided with actors' introspections, $F_s(1, 30) = 8.26$ and 8.06 , $p_s = .008$.

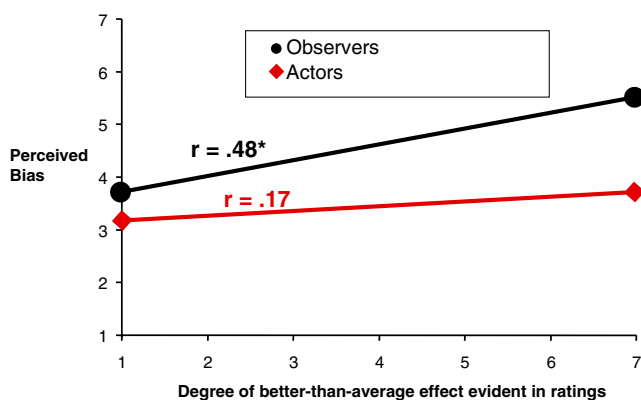


Fig. 1. Attention to behavior (in the form of personality ratings) on the part of actors versus observers assessing susceptibility to the better-than-average effect (Study 2).

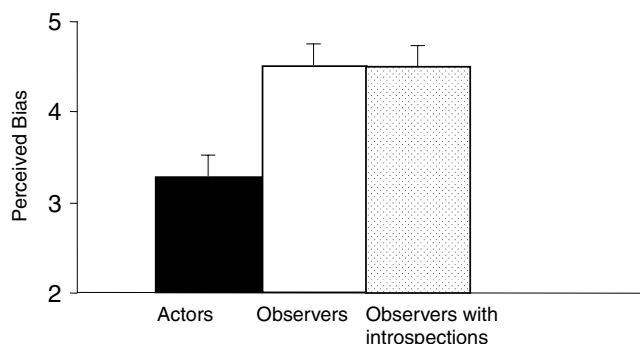


Fig. 2. Actors' versus observers' reports of whether actors' personality ratings were influenced by the "better-than-average" effect (Study 2).

Along with the results of Study 1, these results provide further evidence that while people tend to rely on their own inner thoughts and motives when assessing bias in themselves, they are more likely to rely on observable actions when assessing bias in others. Furthermore, these results suggest that people display a self–other asymmetry in perceptions of bias regardless of whether they have access to the introspections of the "other" whom they are judging.

Despite our efforts to provide observers with genuine and useful accounts of actors' introspections, it is worth considering whether this goal was accomplished. If actors failed to provide understandable accounts of their ongoing thoughts (but rather simply wrote a few words and then proceeded to their next task), or if their accounts did not seem to genuinely reflect their mental contents, then observers' failure to value such reports would tell us little. We thus next examined actors' introspective reports more closely. In so doing, we also sought to ensure that those reports did not yield evidence that actors' were aware of their bias (since our theoretical account posits that actors' denials of personal bias involve its nonconscious occurrence).

Actors' introspective reports

Quantity. To explore whether actors provided useful introspective reports, we first recorded the length of their reports. The median length was 96 words ($SD = 50$). One report fell precisely at the median:

- I'm pretty considerate, and I expect others to be as well. One of my pet peeves is when people do not say thank you when you hold the door for them. When I'm in a bad mood my considerate qualities kind of fly out the window.
- I don't think I'm very deceptive at all. I'm a horrible liar and feel guilty even thinking about lying.
- I'm not very patient, but generally I get along w/ others pretty well.
- I don't think I have anything to be snobby about, especially compared to some of the people here.

We next examined whether the results of this study would persist among observers who read lengthy introspective reports. We thus conducted our analyses again—but this time only with those actor–observer triads in which actors provided reports in the upper tertile for word length. The results supported the hypothesis that observers' disregard of introspective reports was not a function of those reports' inadequacy. Observers who read long introspective reports still saw the actors who provided them as more biased than those actors saw themselves ($M_s = 4.91$ vs. 3.18), $F(1, 10) = 8.60$, $p = .02$. And, observers with access to long introspective reports attributed similar amounts of bias as did observers with no introspective access ($M_s = 4.91, 4.80$), $F = 0$.

Quality. Next, we examined the question of whether observers disregarded actors' introspective reports because

they thought those reports were not faithful accounts of the actors' thoughts. Two research assistants rated each introspective report in terms of whether or not it appeared to be "*an understandable and descriptive account of what the person was thinking*," and whether or not it appeared to be "*a genuine reflection of what the person was thinking*."

Of the 32 introspective reports, 75% were viewed by both raters as understandable and descriptive accounts of the actor's thoughts, and 78% were viewed by both raters as honest and genuine accounts. When we limited our analyses to only those triads in which actors' accounts were viewed as understandable/descriptive and genuine (72% of triads), our predictions were again supported. Observers who received such reports saw the actors who provided them as more biased than those actors saw themselves ($M_s = 4.52$ vs. 3.39), $F(1, 22) = 6.18$, $p = .02$. And, observers who received those reports attributed similar amounts of bias as did observers who received no reports ($M_s = 4.52, 4.41$), $F = 0$.

Nonconscious bias. Our introspection illusion account is rooted in the assumption that actors' commissions of bias typically occur nonsconsciously and that, as a result, attention to introspections will not yield evidence of bias. Consistent with this hypothesis, a research assistant coder of actors' written accounts found that only one of the 32 actors indicated any conscious awareness of the possibility that her ratings may have been positively biased; that participant wrote: "Hmm... I wonder if I'm snobbier than I perceive."

These analyses lend credence to our suggestion that a lack of good introspective access was not the only cause of observers' weighing introspections less than actors. Our next study aimed to obtain yet better introspective access by having actors verbalize their ongoing thoughts.

Study 3: Richer access to others' introspections

Actor participants took a purported test of social intelligence and received a low score. They then were asked to think aloud about the test's validity, and then to rate its validity. Observer participants were shown the test, a yoked actor's purportedly low score on it, and that actor's rating of the test's validity. Half of observers also heard their yoked actor's thoughts. All participants then rated their own (or a yoked actor's) bias in evaluating the test.

Method

Participants

A total of 81 Princeton undergraduates participated in exchange for course credit.

Procedure and questionnaire

Self-assessment condition. Upon agreeing to participate in a study concerning social intelligence, participants were told that: "Social intelligence involves the ability to judge other

people accurately. Aspects of this ability include the ability to make quick judgments about people based on their facial expressions, the way in which they present themselves, apparent categories to which they belong, and other aspects of their appearances and/or behaviors." They were told that we were investigating the validity of different measures of this ability.

Participants were then presented with our purported test of social intelligence (adapted from Pronin, Lin et al., 2002) and told what it involved (i.e., matching people's appearances with their alleged written self-descriptions). They were told that their performance would be assessed by comparing it to that of other Princeton students who had taken the test "for the purpose of building standardized norms on it." Before leaving the participant to take the test, the experimenter commented that social intelligence is "determined mostly by unconscious processes" and that it is thus difficult to gauge how one is performing on the test.

After the participant completed the test, the experimenter appeared to score it, and she gave the participant a score sheet indicating that 7 of 18 matches had been made correctly, placing the participant in the 30th percentile among Princeton students who had taken the test. To obtain a recording of the participant's thoughts about the test validity, the experimenter said:

To help get your perspective on the validity of this social intelligence test, I'd like to give you a chance to think about your opinion as to whether this might or might not be a valid test of social intelligence, and I'd like you to record your thoughts about this as they go through your head. Your thoughts don't have to be complete sentences or even fully formed thoughts; words or phrases are fine, just whatever goes through your head.

After providing their tape recording (in privacy), participants were asked: "Do you think this task is a valid measure of social intelligence?" (1 = *Not at all valid*, 7 = *Very valid*). Then, in an ostensible effort to inform them about the study, the experimenter provided a description of the self-protective tendency in evaluations of test validity (see Pronin, Lin et al., 2002). Participants then assessed whether this bias had affected them: "To what extent do you think your own score influenced your evaluation of the test?" (1 = *Not at all influenced*, 7 = *Very much influenced*).

Other-assessment conditions. These conditions resembled the self-assessment condition, except that rather than being left to take the test, participants were told: "You're not actually taking the test." They were told (underlined portions for introspective-access participants only):

To help us get an observer's perspective on the validity of this social intelligence test, I'm going to show you one student's score on the test and the rating of validity that that student gave the test. Before providing this rating, the student was told, "I'd like to give you a chance to think about your opinion as to whether this might or

might not be a valid test of social intelligence, and I'd like to record your thoughts about this as they go through your head. Your thoughts don't have to be complete sentences or even fully formed thoughts; words or phrases are fine, just whatever goes through your head." After recording their thoughts about the test, the student made the validity rating of the test that you'll see. So look over the test itself, listen to the recording of the student's thoughts about the test, read this additional information about the test, look at the student's score and validity rating, and then there are a couple of questions for you to answer.

The "additional information about the test" was the description of the self-protective tendency given to test-takers. Observers were asked to make the same assessment of bias: i.e., "To what extent do you think the participant's score influenced his/her evaluation of the test?"

Results and discussion

Test evaluation bias

After receiving a poor score on the "social intelligence test," test-takers tended to doubt its validity. Their evaluations fell below the midpoint of the scale ($M = 3.52$ on a 7-point scale), indicating that they did not view the test as valid.

Bias blind spot

Participants imputed less bias to themselves than to their peers. Actors (test-takers) thought that their evaluations of test validity were less influenced by their test scores ($M = 3.59$) than observers thought those actors were influenced ($M = 4.69$), $F(1, 52) = 14.35$, $p = .0004$.

Attention to introspections

We predicted that participants would ignore behavioral information in assessing their own bias but not in assessing others' bias. In this study, the relevant behavioral information involved actors' evaluations of test validity. As predicted, actors' evaluations did not predict the degree to which they reported being influenced by the relevant test-evaluation bias, $r(25) = -.09$, $p = .65$, but their evaluations did predict the degree to which yoked observers felt they were influenced by that bias, $r(52) = -.34$, $p = .01$ (see Fig. 3).

We further predicted that observers would detect more bias in actors than the actors would in themselves, even when those observers had access to the actors' bias-related introspections. Consistent with this prediction, there was no difference between the bias assessments of observers who had access to the test-takers' think-aloud introspections and the bias assessments of those who did not have such access, $F(1, 25) = 1.25$, $p = .27$ (see Fig. 4). Test-takers saw themselves as less biased than did observers who heard their thoughts, $F(1, 25) = 6.09$, $p = .02$, and observers who did not hear their thoughts, $F(1, 25) = 20.31$, $p = .0001$.

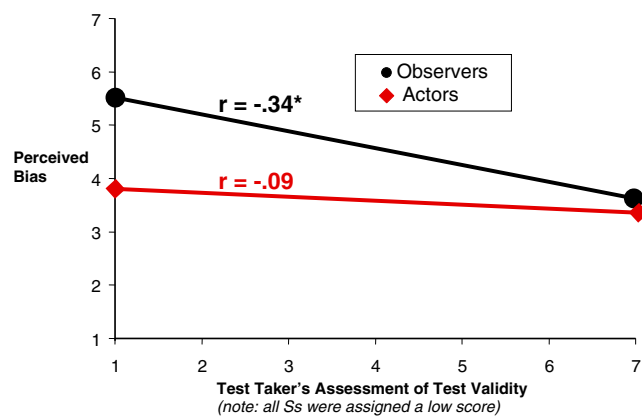


Fig. 3. Attention to behavior (in the form of test assessments) on the part of actors versus observers assessing susceptibility to a self-protective bias (Study 3).

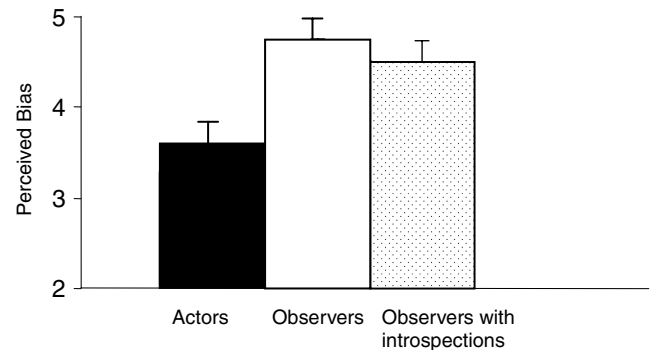


Fig. 4. Actors' (test-takers') reports, versus observers' reports, of whether actors' test validity assessments were influenced by a self-protective bias (Study 3).

As in Study 2, these results provide support for the predicted introspection illusion. Also as in Study 2, we next sought assurance about the quantity and quality of actors' thought reports.

Test-takers' introspective reports

Quantity. The median length of actors' introspective accounts was 141 words ($SD = 54$). The report that fell precisely at the median was:

I think this test is hard to do because you don't actually get to meet the person and I think interaction is probably key in the in determining who people are or what they do stuff like that um uh yeah I don't know on the other hand I guess um its its also just kind of hard to judge peoples occupations by simply facial descriptions er facial pictures and the descriptions they give about themselves I guess I find it surprising that people would be able to get um almost perfect scores on the test um impressive though I would have to say um um I I guess probably most of the judgments I would make of the people on the test were based on stereotypes or past contacts with people who who have the same occupations that are identified.

To examine whether the effects observed in this study would persist in the case of observers who were provided with lengthy introspective reports, we conducted our analyses again—but this time only including those actor-observer triads in which actors provided reports in the upper tertile for length. The blind spot effect again appeared. Observers who heard lengthy introspective reports nevertheless saw their yoked actors as more biased than those actors saw themselves ($M_s = 5.00$ vs. 3.78), $F(1, 16) = 4.99$, $p = .04$. Moreover, those observers attributed similar amounts of bias to actors as did their peers without such access (both $M_s = 5.00$), $F = 0$.

Quality. Next, we examined whether observers may have disregarded actors' introspective reports because they viewed those reports as less than genuine. We recruited a new set of Princeton students ($N = 27$) to participate in the *other-assessment condition with introspective access*. This condition resembled that already reported in every way except that participants were asked one additional question: i.e., whether or not they thought the report they heard was “a genuine reflection” of the test-taker's thoughts.

Of these participants, 81% felt that their yoked peer provided a genuine introspective account. When the responses of these 81% were analyzed (along with the responses of their yoked actor and no-introspective-access observer peers), our usual pattern of results emerged. Participants who were satisfied with their yoked actor's introspective reports nevertheless saw those actors as more biased than those actors saw themselves ($M_s = 4.64$ vs. 3.59), $F(1, 42) = 8.87$, $p = .005$. Moreover, those satisfied observers attributed similar amounts of bias to actors as did their peers who had no introspective access ($M_s = 4.64, 4.82$), $F < 1$.

Nonconscious bias. Our introspection illusion account is rooted in the assumption that commissions of bias typically occur nonsconsciously and that, as a result, attention to introspections will not yield evidence of bias. Consistent with this assumption, a graduate student coder found that only 3 (out of 27) actors mentioned in their recordings the possibility that their test score may have influenced their perception of the test. No participant used the word *bias*.

The forgoing studies provide evidence for our hypothesis of a self—other asymmetry in valuation of introspective versus behavioral information for assessing bias. Our participants seemed to act as though they had two different theories of what it means to be biased: one, relying on introspections, that they applied to themselves, and another, relying on behavior, that they applied to their peers. Study 4 explored this hypothesis more directly.

Study 4: Theories of bias in self versus others

This study aimed to further explore the hypothesis that people are more likely to view bias as defined by introspections (rather than behavior) when they are defining it with respect to themselves (rather than others). Participants read

about situations in which bias might arise (e.g., hiring decisions). Each one was described as involving either the participant or another person. Participants were asked to choose between an action-oriented versus an introspection-oriented description of what it might mean to “be biased” in that situation.

Method

Participants

Thirty-eight Princeton undergraduates participated in exchange for candy.

Procedure and questionnaire

Participants completed a questionnaire describing eight situations in which bias might occur: Hiring job applicants from different groups (gender or race bias); Evaluating the quality of a test after doing poorly on it (self-protective bias); Feelings about two previously liked products after choosing one of them (dissonance reduction); Reactions to negative self-relevant information (motivated processing); Attributions about crime victims (fundamental attribution error); Evaluations of collaborative contributions (egocentric bias); Betting on a roulette wheel (gambler's fallacy); Interest in suddenly available romantic partners (reactive devaluation).

For participants in the *self-assessment condition*, the scenarios were worded such that they themselves were the actor in each scenario. For participants in the *other-assessment condition*, another person (with a different name in each scenario) was always the actor. After reading each scenario, participants were asked, depending on condition, “What might it mean for you to be biased in this situation?” or “What might it mean for [Carol/John/etc.] to be biased in this situation?” They were given two response options, one describing an action and another describing a thought, feeling, or motive. For example, the dissonance reduction scenario (in the self-assessment condition) was as follows, excluding the bracketed information:

You recently had to make a purchase, and you couldn't decide whether to choose item X or item Y. You ultimately select item X.

What might it mean to be biased in this situation?

—A. After selecting item X, you try to make yourself feel better about your choice by thinking about all of the positive qualities of item X and all of the problems with item Y [Thought/feeling/motive]

—B. You tell people all the great things about the item you chose (item X), and the negative things about the one you didn't choose (item Y). [Action]

The other seven scenarios can be found in the Appendix

Results and discussion

As predicted, participants showed a self—other asymmetry in how they defined bias. They were more likely to



Fig. 5. Selection of behavioral versus introspective definitions of bias when considering the self versus others (Study 4).

define bias in terms of a thought, feeling, or motive, as opposed to an action, when primed to think about themselves rather than another person, $F(1, 36) = 4.33$, $p = .04$ (see Fig. 5). While 62% of participants in the self-assessment condition defined bias in terms of introspective information, 51% of participants in the other-assessment condition did so.

These results provide further support for the introspection illusion as a source of the bias blind spot. Participants in this study literally claimed that introspective definitions were simply better than behavioral definitions for defining bias in the case of themselves but not others.

In our next and final study, we took a somewhat different approach to examining the introspection illusion as a source of the bias blind spot. According to the introspection illusion hypothesis, individuals deny their susceptibility to bias because they place heavy (but misplaced) value on their introspections as a source of evidence about bias. In our next study, we sought to educate participants about the common tendency for psychological processes to elude introspective awareness. We predicted that such education would reduce the bias blind spot (by teaching participants about the problem of over-valuing introspections).

Study 5: Education about the limits of introspection

Based on the hypothesis that reliance on introspective evidence when assessing personal bias is a source of the bias blind spot, this study attempted to reduce the blind spot by providing an education about the limits of introspection. Half of participants read a putative article from *Science* reviewing social psychological phenomena that have been demonstrated to operate nonconsciously; the other half were in a control condition. Then in an apparently unrelated study, all participants rated their own susceptibility, relative to their peers, to a range of common biases.

Method

Participants

A total of 78 Princeton undergraduates received course credit for participating. One additional participant was

excluded prior to data analysis due to suspicion about the link between the introspection article and the bias blind spot measure.

Procedure

Upon arriving at the laboratory, participants were told that they would be completing two unrelated studies. The first, they were told, was a “message comprehension study.” For it, they were given a scientific article or articles to read (see below), with each one followed by a series of true/false questions included to bolster our cover story. The second study, they were told, was a “perceptual questionnaire.” In reality, the first task constituted our experimental manipulation, and the second our dependent measure.

Article stimuli

All participants were given a real article from *Nature* magazine to read and respond to, and those in the experimental condition were also given a fake article that appeared to be from *Science*. The *Nature* article described a technique for assessing changes in environmental pollution by measuring lead in different wine vintages (Lobinski et al., 1994). It was chosen for its brevity and accessibility to non-scientists. The article and the true/false questions that followed it had been used in previous research, also for the purposes of adding experimental control and providing a cover story (Pronin, Steele, & Ross, 2004).

The alleged *Science* article was titled “Unaware of Our Unawareness.”¹ It referenced a wide variety of social psychological studies concerning nonconscious influences on attitudes and behavior (including the titular reference to Wilson et al., 1995). The article began by suggesting that much of what goes on in the human mind is inaccessible via introspection:

“I’ll know it when I see it,” runs the popular refrain. It’s been used to explain how we can recognize everything from obscenity to true love. But how much can we trust what we see or, rather, what we think we see? For decades, cognitive psychologists have been discovering that there is more going on in our brains than we could ever be consciously aware of, even for a moment.

The article reviewed a number of genuine psychological findings involving the effects of nonconscious influences. These included the effects of nonconscious cues on bystander intervention (Darley & Latane, 1968; Darley, Teger, & Lewis, 1973), interpersonal rudeness (Bargh, Chen, & Burrows, 1996), achievement at a verbal task (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Troetschel, 2001), cooperation vs. competition in a prisoner’s dilemma game (Kay & Ross, 2003), status perceptions (Chen, Lee-Chai, & Bargh, 2001), walking speed (Bargh et al., 1996), aggressive and violent behavior (Anderson & Bushman, 2002; Donnerstein, Slaby, & Eron, 1994), aversive racism

¹ A copy of the article is available from the authors on request.

(Dovidio & Gaertner, 2004; Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005; Nail, Harton, & Decker, 2003), social judgment (Nisbett & Wilson, 1977b), and pantyhoose preferences (Nisbett & Wilson, 1977a). Thus, for example, Bargh and colleagues (1996) classic experiment was described as follows:

In one study, subjects were led to think about elderly people. The subjects were then told that the experiment was over and that they were free to go. What did the experimenter measure? How quickly the person walked down the hall when leaving the experiment. And yes, the people with the elderly thoughts took longer (16).

These various nonconscious effects were described as arising from a range of different sources that participants failed to recognize as causally responsible, including subliminal primes, the media, and behavioral mimicry. Evidence for these effects was described as deriving from a variety of scientific methods, including laboratory experiments, fMRI studies, survey research, and longitudinal data sets. The central message of the article was that there are many factors influencing our behavior and judgments that do not register in our introspections. The concluding paragraph included the following summary:

Try as they might, experimenters have not been able to make people realize when they have been unintentionally influenced. We believe that we would know if our actions were being so drastically altered. Yet these studies and hundreds of others have all shown that we know less about our motivations and about the sources of our actions, judgments, and decisions than we thought.

The article was formatted to appear as though it were photocopied from *Science*. It was 1643 words excluding references. As with the *Nature* piece, it was followed by a series of true/false questions. Most participants (93%) answered at least 8 of the 9 questions correctly.

Dependent measure

Our dependent measure described 10 different biases and called for participants to rate their own susceptibility, relative to that of their Princeton peers, to each one. Participants were told that the survey dealt with “several tendencies that can affect judgment and thought” and that they would “be asked to indicate how much you think that YOU PERSONALLY show each tendency relative to other students AT PRINCETON.” The measure included the bias descriptions used in Study 1 (for the self-serving bias, fundamental attribution error, and positive halo effect), as well as descriptions of the disconfirmation tendency (biased assimilation), the spreading of alternatives after free choice (dissonance reduction), self-interest, the hostile media effect, the anchoring effect, the better-than-average effect, and in-group favoritism (Cronbach’s $\alpha = .70$). Each bias description was followed by the question: “To what extent do you believe that you show this tendency compared to the average Princeton student?”

(1 = much less than the average student, 6 = same as the average student, 11 = much more than the average student).

Results and discussion

Consistent with past research, participants in the control condition showed a bias blind spot. They reported being less susceptible than their collegiate peers to the set of biases ($M = 5.29$, where 6 = “same” as peers), $t(38) = 4.67$, $p < .0001$. Our primary concern in this research involved participants in the experimental condition, which had educated them about the importance of nonconscious processes. Consistent with our hypothesizing, those participants did not show a bias blind spot ($M = 5.88$), $t < 1$. Moreover, participants who were educated about the role of nonconscious processes not only avoided committing the bias blind spot but also were more likely than their control condition peers to avoid that commission, $F(1, 76) = 6.69$, $p = .01$.

These results provide further evidence for the introspection illusion as a source of the bias blind spot. When participants were taught that valuing introspections is likely to lead one astray in making judgments about influences on the self, they ceased claiming that they were less susceptible to bias than their peers. These results also demonstrate an effective method for eliminating the bias blind spot. This may be a useful goal in its own right, as the tendency to claim objectivity in one’s own judgments, while readily imputing bias to others, can exacerbate conflict (Pronin, Kennedy, & Butsch, 2006; Pronin, Puccio, & Ross, 2002).

General discussion

The research reported here supports the contention that the bias blind spot is rooted at least in part in an introspection illusion. These results suggest that the tendency to impute bias more to others than to the self is rooted in people’s tendency to value introspective information—at the expense of considering actions—when assessing bias in themselves, but not others.

First, participants showed a self—other asymmetry in the information they considered when making assessments of bias. In Study 1, they reported attending to introspective information more than behavioral information for assessing bias in themselves but not others, and this self—other difference mediated the effect of self/other on imputations of bias. The results of Studies 2–3 extended these findings on information consideration to participants assessing a specific instance of bias in themselves or in a peer. Participants reported considering introspections more than behavior for self but not a peer (Study 2). And, in Studies 2 and 3, actors ignored their biased actions in drawing conclusions about whether they had been biased, whereas yoked observers took such information into account.

Second, the self—other difference in participants’ reliance on introspective information was not simply a matter of observers’ lacking access to actors’ introspections.

Participants in Study 1a reported that introspections were a more valuable source of information about their own bias than others' bias. In Studies 2–3 the observed self–other asymmetry in bias perception emerged not just among actors with introspective access and observers without it, but also when observers were provided with introspective access. Indeed, these effects persisted even when observers were provided with substantial introspective reports (more than 150 words) and when they believed that those reports were faithful accounts of actors' ongoing thoughts. In Study 4, participants literally chose to define bias in terms of introspective contents rather than actions when primed to think about bias in themselves rather than others. Finally, the results of Study 5 demonstrated that when participants were informed (in an allegedly separate study) that it would be unwise to place so much value on their introspections, they overcame the usual tendency to see themselves as less biased than their peers.

The observed self–other asymmetry in people's willingness to rely on introspective information has been referred to as an introspection *illusion* because the faith people have in the diagnostic value of their introspections is misplaced. It is true that introspective contents, such as thoughts, feelings and intentions, can provide a useful source of information about the self, for example when it comes to understanding the causes of our behavior or predicting how we will act in the future (Jones & Nisbett, 1972). But, it is also true that such information can mislead us in those very same cases, when our behavior has been influenced by non-conscious cues or when future circumstances prevent our good intentions from translating into good behavior.

Behavioral disregard

In the present research, participants not only weighted introspective information more in the case of self than others, but they concurrently weighted behavioral information more in the case of others than self. Our participants even seemed to claim that such “behavioral disregard” was appropriate for judging themselves, but not others. This claim is clearly unjustified. Apart from the obvious problem of holding self and other to different standards for judging bias (standards which will clear the self of bias more often than others), treatment of introspective information as sovereign leads to frequent misjudgments, as we have already noted. Although some research suggests that observers are overly prone to view actors as biased (Kruger & Gilovich, 1999; Van Boven, Kamada, & Gilovich, 1999), the present studies show that actors are overly prone to deny their own bias. Even when actors collectively rated themselves as better-than-average, and even when they questioned the validity of a test on which they had performed poorly (a test that they would have praised had they performed well, as shown by Pronin, Lin et al., 2002), they denied showing signs of bias. Moreover, these denials were associated with their self-reported disregard of behavior and reliance on

thoughts (Study 1), and with their actual disregard of specific biased actions (Studies 2 and 3).

These findings of behavioral disregard are noteworthy in the general context of lay social judgment. In most of life, we are judged by our actions rather than by our intentions, hopes, or feelings. We are generally characterized as good or evil, generous or greedy, and wild or dull, not by what we think about but by what we actually do. In this light, it is striking that people ignore their own actions when making assessments of bias. Although the term introspection illusion emphasizes self–other differences in the faith that people place in their introspections, a necessary component of it involves the neglect with which people treat their own behaviors.

Self-enhancement and bias perception

The tendency to rely heavily on one's own introspections when making bias assessments, while ignoring one's behavior, may have a motivational component to it. Biases are generally viewed as undesirable and thus people are likely to want to deny them. Indeed, past research has shown that people are more likely to deny their relative susceptibility to negatively valenced biases than to positively valenced ones (Pronin, Lin et al., 2002). In this regard, motivational concerns may provide a mechanism that underlies the introspection illusion mechanism we have shown. Given a choice between considering their behavior versus considering their thoughts, actors may have preferred to consider thoughts and thereby see themselves as bias-free (rather than considering behavior and thereby being compelled to acknowledge possible bias).

Yet, we hasten to note, people's motive to deny personal bias probably could not be served by their relying on introspective information if they were well-aware of the striking fallibility of such information. It would be difficult to feel good about one's self-proclaimed lack of bias if one knew that this proclamation relied on false information. Thus, we imagine that the self–other asymmetry in valuing thoughts versus actions likely involves an interplay of both cognitive factors (involving the felt validity of personal introspections) and motivational factors (involving the increased possibility for denying bias when introspections provide the basis for judgment). Though our participants were likely motivated to see themselves in a positive light, their efforts could easily have been thwarted were they fully aware of the weak value of their own introspections. The results of Study 5 support this hypothesis. When participants were informed that introspective information generally is a poor source of information about influences on human judgment and action, they no longer showed the tendency to deny their relative susceptibility to bias.

Of course, the judgmental biases that are the focus of this article are not the only influences that can elude introspective awareness. Decades of research have revealed that much of human behavior is guided by mental processes and contextual cues that are beyond conscious awareness (for

classic and modern reviews, see Nisbett & Wilson, 1977a; Wegner & Bargh, 1998). The present research suggests the hypothesis that individuals may deny the influence of a whole host of nonconscious influences on their own judgments and actions, while readily imputing those influences to other people. Recent work suggests that people may see others as far more susceptible than themselves to conformity and implicit social influence (Pronin, Berger, & Molokai, 2006). Perhaps individuals also view themselves as less susceptible to implicit prejudice and to “priming effects” in general. Future studies could examine whether when we deny our own susceptibility to racial or gender prejudice, or even to the subtle influence of the weather on our mood, we are alone in those denials. The present research suggests that such self–other asymmetries are especially likely to occur in cases where the relevant influence is likely to operate below conscious awareness (while leaving behavioral evidence).

Conflict and misunderstanding

The tendency to deny bias in oneself while imputing it to others is likely to foster interpersonal conflict and misunderstanding. When others have shown just the same amount of self-interest in their political views as we have, or when they have shown just the same amount of intergroup bias in their decisions about whom to reward and whom to punish, we are likely to see them as biased and ourselves as objective – especially when those others’ views and decisions differ from our own. Conflict is likely to ensue, and feelings of enmity are likely to worsen, as we resent their accusations of us as biased, when we are certain that we have not been, and when we are certain that they *have* been.

Here, one lesson of this research is clear. Whether we choose to define bias according to behavior or introspection, fairness dictates that we apply the same definition to others that we apply to ourselves. If we are unwilling to accept that our adversaries are free of bias because their thoughts were pure, we should not use this argument to defend our own freedom from bias. Likewise, if we think it is reasonable to accuse others of bias based on their actions, we should be prepared to consider our own actions as equally capable of betraying signs of bias.

Appendix A. Questionnaire Items (Study 4)

1. Imagine that you are the manager at a store, and you are sifting through job applications to decide whom to hire. What might it mean to be biased in this situation?
 - A. You end up hiring people mainly of your race and gender.
 - B. While you are reading the applications, you tend to feel a preference for people of your race and gender.
2. You just received your final exam grades, and you did very poorly on one exam. What might it mean to be biased in this situation?

- A. You feel motivated to make yourself feel better, so you try to think of reasons why the exam was unfair.
- B. You complain loudly and strongly about the fairness of the professor’s grading criteria for the exam.

3. You just read a magazine article about the declining benefits of an Ivy League education. What might it mean to be biased in this situation?
 - A. You throw the article in the garbage and say it is “absurd.”
 - B. You feel the need to find flaws in the article’s arguments.
4. You’re sitting on a jury in a case where someone’s home was robbed. What might it mean to be biased in this situation?
 - A. You point out to the jury that the victim’s front door had been left unlocked and so the robbery is due to the victim’s own foolishness.
 - B. You personally believe that people get what they deserve, and so you feel that the robbery was the victim’s fault because the door was left unlocked.
5. You just turned in a group project that you know the professor likes, and the professor asks every member of the four-person group to write down how much he or she contributed to the final project. What might it mean to be biased in this situation?
 - A. You write down that you did about 60% of the work.
 - B. You want to see yourself as having contributed more than everyone else to the project.
6. You’re gambling, and the roulette wheel has now landed on red four times in a row. What might it mean to be biased in this situation?
 - A. Although your prior bets have all been small, you now place a rather large bet—on black.
 - B. You think that since there has been a streak of red, black must be due to come up next.
7. You’re at a school dance and you immediately notice someone great-looking who you point out to all your friends. The person comes right over and asks you to dance. What might it mean to be biased in this situation?
 - A. You suddenly think that the person must not be as cute as you thought, or else they would be more “hard to get.”
 - B. You suddenly tell all your friends that you’re not interested in the person at all, and you set your eyes on someone new.

Note. Excludes one item reported in Methods section. Above items are for the “self” version.

References

- Andersen, S. M., & Ross, L. (1984). Self knowledge and social inference: I. The impact of cognitive/affective and behavioral data. *Journal of Personality and Social Psychology*, 46, 280–293.

- Anderson, C. A., & Bushman, B. J. (2002). Media violence and the American public revisited. *American Psychologist*, 57(6-7), 448–450.
- Bargh, J., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- Bargh, J., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, 81, 1014–1027.
- Baron, R. M., & Kenny, D. A. (1986). The moderator mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 29, 61–82.
- Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the “planning fallacy”: Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, 67, 366–381.
- Chen, S., Lee-Chai, A., & Bargh, J. A. (2001). Relationship orientation as a moderator of the effects of social power. *Journal of Personality and Social Psychology*, 80, 173–187.
- Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personality and Social Psychology*, 8, 377–383.
- Darley, J. M., Teger, A. I., & Lewis, L. D. (1973). Do groups always inhibit individuals’ responses to potential emergencies? *Journal of Personality and Social Psychology*, 26, 395–399.
- Donnerstein, E., Slaby, R. G., & Eron, L. D. (1994). The mass media and youth aggression. In L. D. Eron, J. H. Gentry, & P. Schlegel (Eds.), *Reason to hope: A psychosocial perspective on violence and youth* (pp. 219–250). Washington, DC: APA.
- Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (36, pp. 1–52). San Diego, CA: Elsevier Academic Press.
- Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving appraisals of ability. *Journal of Personality and Social Psychology*, 57, 1082–1090.
- Ehrlinger, J., Gilovich, T., & Ross, L. (2005). Peering into the bias blind spot: People’s assessments of bias in themselves and others. *Personality and Social Psychology Bulletin*, 31, 680–692.
- Epley, N., & Dunning, D. (2000). Feeling “holier than thou”: Are self-serving assessments produced by errors in self- or social prediction? *Journal of Personality and Social Psychology*, 79, 861–875.
- Gilbert, D. T., Pines, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75, 617–638.
- Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the cause of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.
- Kay, A. C., & Ross, L. (2003). The perceptual push: The interplay of implicit cues and explicit situational construals on behavioral intentions in the Prisoner’s Dilemma. *Journal of Experimental Social Psychology*, 39, 634–643.
- Kruger, J. (1999). Lake Wobegon be gone! The “below-average effect” and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, 77, 221–232.
- Kruger, J., & Gilovich, T. (2004). Actions, intentions, and self-assessment: the road to self-enhancement is paved with good intentions. *Personality and Social Psychology Bulletin*, 30, 328–339.
- Kruger, J., & Gilovich, T. (1999). “Naïve cynicism” in everyday theories of responsibility assessment: on biased assumptions of bias. *Journal of Personality and Social Psychology*, 76, 743–753.
- Kunda, Z. (1987). Motivated inference: self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology*, 53, 636–647.
- Latane, B., & Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology*, 10, 215–221.
- Lieberman, M. D., Hariri, A., Jarcho, J. M., Eisenberger, N. I., & Bookheimer, S. Y. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience*, 8, 720–722.
- Lobinski, R., Witte, C., Adams, F. C., Teissedre, P. L., Cabanis, J. C., & Boutron, C. F. (1994). Organolead in wine. *Nature*, 370, 24.
- Malle, B. F., & Pearce, G. E. (2001). Attention to behavioral events during interaction: Two actor-observer gaps and three attempts to close them. *Journal of Personality and Social Psychology*, 81, 278–294.
- McGuire, W. J., & McGuire, C. V. (1986). Differences in conceptualizing self versus conceptualizing other people as manifested in contrasting verb types used in natural speech. *Journal of Personality and Social Psychology*, 51, 1135–1143.
- Nail, P. R., Harton, H. C., & Decker, B. P. (2003). Political orientation and modern versus aversive racism: tests of Dovidio and Gaertner’s (1998) integrated model. *Journal of Personality and Social Psychology*, 84, 754–770.
- Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- Nisbett, R. E., & Wilson, T. D. (1977a). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 8, 231–259.
- Nisbett, R. E., & Wilson, T. D. (1977b). The halo effect: evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, 35, 250–256.
- Prentice, D. A. (1990). Familiarity and differences in self- and other-representations. *Journal of Personality and Social Psychology*, 59, 369–383.
- Pronin, E., Berger, J. A., & Molouki, S. (2006). Alone in a crowd of sheep: Asymmetric perceptions of conformity and their roots in an introspection illusion. Under review.
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: divergent perceptions of bias in self versus others. *Psychological Review*, 111, 781–799.
- Pronin, E., Kennedy, K., Butsch, S. (2006). Bombing versus negotiating: How preferences for combating terrorism are affected by perceived terrorist rationality. *Basic and Applied Social Psychology*, in press.
- Pronin, E., Kruger, J., Savitsky, K., & Ross, L. (2001). You don’t know me, but I know you: the illusion of asymmetric insight. *Journal of Personality and Social Psychology*, 81, 639–656.
- Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: perceptions of bias in self versus others. *Personality and Social Psychology Bulletin*, 28, 369–381.
- Pronin, E., Puccio, C., & Ross, L. (2002). Understanding misunderstanding: social psychological perspectives. In D. Griffin & T. Gilovich (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 636–665). New York, NY: Springer.
- Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: women and mathematics. *Journal of Experimental Social Psychology*, 40, 152–168.
- Pronin, E., Wegner, D. M., McCarthy, K., & Rodriguez, S. (2006). Everyday magical powers: the role of apparent mental causation in the overestimation of personal influence. *Journal of Personality and Social Psychology*, 91, xxx–xxx (in press).
- Van Boven, L., Kamada, A., & Gilovich, T. (1999). The perceiver as perceived: everyday intuitions about the correspondence bias. *Journal of Personality and Social Psychology*, 77, 1188–1199.
- Wegner, D. M., & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.), *Handbook of social psychology* (4 ed., Vol. 1, pp. 446–496). New York: McGraw-Hill.
- Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. *Psychological Bulletin*, 116, 117–142.
- Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In D. Griffin & T. Gilovich (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 185–200). New York: Cambridge.
- Wilson, T. D., Hodges, S. D., & LaFleur, S. J. (1995). Effects of introspecting about reasons: inferring attitudes from accessible thoughts. *Journal of Personality and Social Psychology*, 69, 16–28.