

Personality and Social Psychology Bulletin

<http://psp.sagepub.com/>

People Claim Objectivity After Knowingly Using Biased Strategies

Katherine Hansen, Margaret Gerbasi, Alexander Todorov, Elliott Kruse and Emily Pronin

Pers Soc Psychol Bull 2014 40: 691 originally published online 21 February 2014

DOI: 10.1177/0146167214523476

The online version of this article can be found at:

<http://psp.sagepub.com/content/40/6/691>

Published by:



<http://www.sagepublications.com>

On behalf of:



<http://www.spsociety.org>
Society for Personality and Social Psychology

Additional services and information for *Personality and Social Psychology Bulletin* can be found at:

Email Alerts: <http://psp.sagepub.com/cgi/alerts>

Subscriptions: <http://psp.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

>> [Version of Record](#) - May 19, 2014

[OnlineFirst Version of Record](#) - Feb 21, 2014

[What is This?](#)

People Claim Objectivity After Knowingly Using Biased Strategies

Katherine Hansen¹, Margaret Gerbasi¹, Alexander Todorov¹, Elliott Kruse¹, and Emily Pronin¹

Personality and Social Psychology Bulletin
2014, Vol. 40(6) 691–699
© 2014 by the Society for Personality and Social Psychology, Inc
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0146167214523476
pspb.sagepub.com



Abstract

People tend not to recognize bias in their judgments. Such “bias blindness” persists, we show, even when people acknowledge that the judgmental strategies preceding their judgments are biased. In Experiment 1, participants took a test, received failure feedback, and then were led to assess the test’s quality via an explicitly biased strategy (focusing on the test’s weaknesses), an explicitly objective strategy, or a strategy of their choice. In Experiments 2 and 3, participants rated paintings using an explicitly biased or explicitly objective strategy. Across the three experiments, participants who used a biased strategy rated it as relatively biased, provided biased judgments, and then claimed to be relatively objective. Participants in Experiment 3 also assessed how biased they expected to be by their strategy, prior to using it. These pre-ratings revealed that not only did participants’ sense of personal objectivity survive using a biased strategy, it grew stronger.

Keywords

bias blind spot, bias correction, objectivity illusion, mental contamination

Received December 18, 2013; revision accepted January 20, 2014

Following rejection after a job interview, one might seek consolation from friends. Suppose a sympathetic friend points out that although you had been excited about the job, you also had some reservations. She prompts you to focus on those concerns, while putting aside the positives. You realize that this is a biased exercise, but you go along. Afterward, you come to feel that things worked out for the best. Moreover, although you recognize that your exercise was biased, you think it led to the objectively right conclusion.

What happens when a person *knowingly* engages in a biased thought process, as in the above scenario? That scenario differs from one in which a person “automatically” engages in a biased process—for example, automatically focusing on the downsides of a job after bombing the interview. Past research has found that when people’s judgments result from biased strategies that are recruited automatically, people can be blind to the bias that those strategies produce (e.g., Ehrlinger, Gilovich, & Ross, 2005; Pronin, Gilovich, & Ross, 2004; Wilson, Centerbar, & Brekke, 2002). In this article, we propose that even when people are aware that their thought processes preceding a particular judgment are biased, they still tend to view the judgment they reach as objective (even if that judgment is, in fact, biased).

The particular sort of biases that this article concerns are those that actors themselves would label as biases (regardless of whether some normative criterion for objectivity is

violated). We suggest that people may recognize and label bias in a procedure or strategy they use to make a judgment, but they may nevertheless claim that they have been relatively objective in the face of using that procedure or strategy.

Previous research concerning the bias blind spot has primarily looked at it from an interpersonal perspective, focusing on the asymmetry between individuals’ sense of personal objectivity compared with their sense of others’ bias (e.g., Pronin, Lin, & Ross, 2002; West, Meserve, & Stanovich, 2012). This research has not explored the fundamentally *intra*-personal question of how, or indeed whether, individuals’ sense of objectivity can survive their own use of a patently biased strategy. Past participants have assessed the presence of bias in their own and others’ judgments and views, but they have not assessed the bias inherent in the judgmental strategies per se that they took to arrive at those views. In a typical experiment, participants might label other people’s political views as biased by ideology and self-interest, while maintaining objectivity in their own political views. But, those participants would not be asked to describe

¹Princeton University, NJ, USA

Corresponding Author:

Emily Pronin, Department of Psychology, Princeton University, Princeton, NJ 08540, USA.
Email: epronin@princeton.edu

(nor would they be told about) the strategies that they or others use to reach their views, nor, therefore, would they be asked to assess the bias inherent in those strategies *per se*. Thus, past research has not examined the important question of whether individuals may perceive their judgmental strategies as biased while perceiving their judgments following those strategies as relatively objective.

Rather than examining whether individuals are less inclined to detect biases in themselves than in others, the present experiments examine whether individuals are less inclined to detect biases in themselves than *in their own decision-making strategies*. If this type of persistent bias blind spot occurs, it could be important—not only because it would show that the bias blind spot can occur even when there is no “other” to denigrate but also because of its practical implications. To the extent that people deny being biased by processes and procedures that they readily (and correctly) see as biasing, the use of such processes and procedures may be unwarrantedly perpetuated. This research also holds implications for theories of bias correction. A variety of these theories point out that recognizing the potential for bias in one’s judgments is a critical step in correcting bias (e.g., Martin, 1986; Schwarz & Bless, 1992; Wegener & Petty, 1997; Wilson & Brekke, 1994). That step may be met with surprising infrequency, though, if recognition of bias in one’s judgmental strategies does not trigger similar recognition of potential bias in one’s own judgments.

We present three experiments investigating the psychological effects of making a judgment while knowing that the strategy behind it is biased. In Experiment 1, participants received false feedback on an alleged test of their social intelligence. They then were instructed to evaluate the test using either an explicitly biased strategy (e.g., evaluating the test solely based on its weaknesses), an explicitly objective strategy, or an unrestricted strategy. We predicted that participants in the explicitly biased condition would rate the evaluative procedure as relatively biased, engage in it, and then claim to have been relatively objective. In Experiment 2, participants rated various paintings using either a strategy that they viewed as explicitly biased (looking at the identity of the painter before assessing the painting’s inherent quality) or a strategy that they viewed as explicitly objective (shielding themselves from the painters’ identities). We again predicted that participants in the explicitly biased condition would rate the evaluative procedure as relatively biased, engage in it, and then claim to have been relatively objective (a claim that we predicted would be undermined by their ratings of the paintings). Participants in Experiment 3 also rated paintings, but in this study, they assessed bias at two time points: They not only assessed bias in themselves (and in their strategy) after rating the paintings, but they also assessed how biased they would be (and how biased the strategy is) prior to using it. We predicted that participants’ denials of personal bias would show strong persistence in the face of using the biased strategy.

Experiment 1

Method

Participants. A total of 101 Princeton undergraduates (63 female, 38 male) participated for course credit. Seven participants were excluded prior to data analysis because they strongly suspected that their test score was fake.

Procedure, conditions, and measures. Participants arrived at the laboratory and took a purported newly developed test of “Social Intelligence.” The test involved 18 photographs of people’s faces and 18 statements allegedly written by those individuals about their hobbies, careers, and preferences; participants were asked to match each face with the correct self-description. Afterward, they received false feedback on their performance. They then were asked to evaluate the quality of the test. Participants in three of the four experimental conditions received failure feedback (a score of 6 correct matches, ranking in the 27th percentile). They then were randomly assigned to evaluate the test using one of three strategies: In the *explicitly objective condition*, they were asked to list three strengths and three weaknesses of the test; in the *unrestricted condition*, they were asked to list strengths and weaknesses at their discretion; in the *explicitly biased (weakness-focus) condition*, they were asked to list only weaknesses.

Previous work using this alleged Social Intelligence test has demonstrated that individuals’ evaluations of this test are self-servingly biased such that individuals who believe that they have performed poorly evaluate the test more negatively than individuals who believe they have performed well (Pronin & Kugler, 2007; Pronin et al., 2002). To be sure that this bias emerged in the current experiment, we included a fourth experimental condition in which participants were led to believe that they performed well. Those participants were given success feedback (a score of 14 correct matches, 80th percentile) and asked to evaluate the test in an explicitly positively biased way—that is, by listing only strengths of the test (*explicitly biased strength-focus condition*).

Before engaging in their assigned strategy for evaluating the test, all participants assessed the objectivity of their assigned strategy (1 = *completely biased*, 9 = *completely objective*). After engaging in that strategy, they rated the quality of the test (1 = *very bad test*, 9 = *very good test*). Last, they assessed the objectivity of the rating of the test that they had provided (1 = *completely biased*, 9 = *completely objective*).

Results and Discussion

Manipulation check: Adherence to strategy. The number of strengths and weaknesses that participants listed when evaluating the test was coded as a manipulation check of their adherence to their assigned evaluative strategy. Participants complied with the strategy to which they were assigned, with the exception of those in the explicitly objective condition.

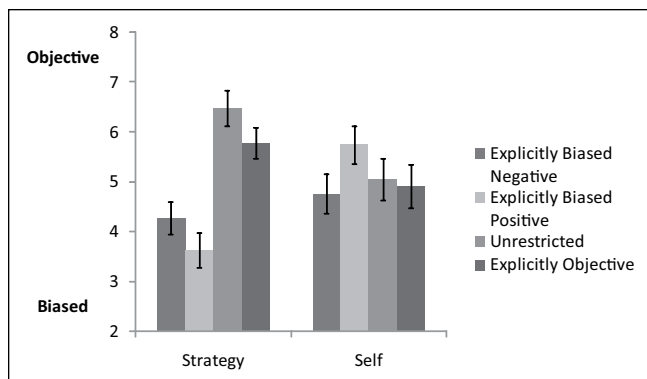


Figure 1. Participants' ratings of bias in their judgmental strategy, and in their subsequent self-judgment (Experiment 1). Note. Error bars indicate 1 SE above and below the mean.

In the explicitly biased weakness-focus condition, participants listed more weaknesses of the test ($M = 2.12$, $SD = 1.09$) than strengths ($M = 0.04$, $SD = 0.20$), $t(24) = 9.66$, $p < .0001$. In the explicitly biased strength-focus condition, participants listed more strengths ($M = 3.38$, $SD = 1.38$) than weaknesses ($M = 0.13$, $SD = 0.45$), $t(23) = 9.66$, $p < .0001$. Participants in the unrestricted condition responded in a manner consistent with the presence of an ego-protective bias: After receiving failure feedback, they listed more weaknesses than strengths ($M = 2.77$, $SD = 1.54$, vs. $M = 1.45$, $SD = 0.96$), $t(21) = 3.70$, $p = .001$. In the explicitly objective condition, participants were asked to write an equal number of strengths and weaknesses (three of each); however, they wrote slightly more weaknesses than strengths ($M = 3.22$, $SD = 0.42$, vs. $M = 3.04$, $SD = 0.37$), $t(22) = 2.15$, $p = .04$.

Assessments of bias. Our primary prediction concerned participants in the two explicitly biased conditions. Those participants either received failure feedback and then were asked to evaluate the test by focusing on its weaknesses, or they received success feedback and then were asked to evaluate the test by focusing on its strengths. We predicted that both these sets of participants would (a) rate their assigned strategy as biased compared with participants in the other conditions (i.e., the unrestricted condition and the explicitly objective condition), and (b) fail to rate their evaluation of the test as biased compared with participants in those other two conditions. Thus, we predicted a 4×2 interaction effect (Evaluation Strategy: *Explicitly Biased–Weakness Focus*, *Explicitly Biased–Strength Focus*, *Explicitly Objective*, *Unrestricted* \times Bias Assessment: *Self-Assessment*, *Assessment of Strategy*).

The predicted interaction effect was tested using a mixed-model ANOVA, with evaluation strategy as a between-subjects variable, and type of bias assessment as a within-subjects variable. The interaction was significant, $F(3, 89) = 13.02$, $p < .0001$ (see Figure 1). To further explore this interaction effect, we next analyzed participants' assessments of bias in

their strategy separately from their assessments of bias in their own evaluations.

We predicted that participants in the explicitly biased conditions would see their assigned strategies as biased compared with participants in the other conditions. An omnibus test of all four conditions was significant, $F(3, 90) = 15.76$, $p < .0001$. As predicted, participants in the explicitly biased weakness-focus condition rated their assigned strategy as less objective ($M = 4.28$, $SD = 1.62$) than did participants in the unrestricted condition ($M = 6.50$, $SD = 1.60$), $t(45) = 4.72$, $p < .0001$, and in the explicitly objective condition ($M = 5.78$, $SD = 1.48$), $t(46) = 3.35$, $p = .002$. Likewise, those in the explicitly biased strength-focus condition ($M = 3.63$, $SD = 1.72$) saw their strategy as less objective than did participants in the unrestricted condition, $t(44) = 5.87$, $p < .0001$, and in the explicitly objective condition, $t(45) = 4.62$, $p < .0001$.

Although participants in the explicitly biased conditions perceived greater bias in their assigned strategies than did their peers, we predicted that they would rate their test evaluations as no more biased. As predicted, an omnibus test of the effect of participants' experimental condition on their assessments of the objectivity of their test evaluations revealed no differences, $F(3, 89) = 1.21$, $p = .31$, with participants in all conditions rating their choices as similarly objective ($M_{\text{explicitly biased, weakness-focus}} = 4.76$, $SD = 1.99$, $M_{\text{explicitly biased, strength-focus}} = 5.75$, $SD = 1.85$, $M_{\text{explicitly objective}} = 4.91$, $SD = 2.02$, $M_{\text{unrestricted}} = 5.14$, $SD = 1.94$), $ts < 1.8$, $ps > .08$.

Bias in explicitly biased conditions. The present study concerns individuals' tendency to deny bias in their judgments even when they have acknowledged that the processes leading up to those judgments are biased. Although such a tendency to deny bias seems consistent with a bias blind spot, one could argue that such denials were accurate—if participants were able to be objective although their judgmental processes were biased. To address this possibility, we next examined whether participants in the explicitly biased conditions had been biased.

As predicted, participants in our explicitly biased conditions showed a "self-serving" bias whereby they saw the test as less valid after allegedly performing poorly on it rather than well. Those who failed the test and then were instructed to list test flaws rated the test as less valid than did those who succeeded on the test and then were instructed to list test strengths ($M = 4.72$, $SD = 1.24$ vs. $M = 5.58$, $SD = 1.53$), $t(47) = 2.17$, $p = .04$. It also may be worth noting that participants in the three failure conditions (those who were instructed to be biased, to be objective, or to use their own discretion) all evaluated the test similarly to each other, all $ts < 1.05$, ns , and more negatively than those in the one success condition, all $ts > 2.17$, $ps < .04$. It is not surprising that participants in the explicitly objective condition were as negative about the test as their peers. As noted earlier, they did not

fully comply with the objectivity instruction and instead listed more weaknesses than strengths.

In this experiment, participants in the explicitly biased conditions rated their judgmental strategies as biased relative to those in the explicitly objective conditions, and they then made biased judgments using those strategies; then, they claimed that their judgments were no less objective than those in the explicitly objective conditions. Importantly, the objectivity claims of participants in the bias conditions were uncorrelated with the actual amount of objectivity in their judgments. That is, greater displays of bias (i.e., more criticism of the test after doing poorly or less criticism after doing well) were not associated with decreased claims of objectivity, $r(51) = .05, p = .73$.

In this experiment, we were able to ascertain the presence of bias among participants in the explicitly biased conditions via a comparative analysis—that is, participants' ratings of the test were more negative when they had been told they did poorly rather than well. Experiment 2 aims to augment these results by using an external criterion to assess bias rather than doing so comparatively. It also introduces a new domain of bias commission to test our hypotheses outside the realm of self-serving biases and test performance.

Experiment 2

Method

Participants. Seventy-four Princeton undergraduates (47 women, 27 men), all of whom reported no knowledge of art history beyond “visiting museums,” participated for course credit.

Procedure, conditions, and measures. Participants arrived at the laboratory and rated a series of 80 paintings in terms of their “artistic merit.” They were told that “by artistic merit, we mean the *quality* of the painting.” Participants were then told that they would be asked to make those ratings after either choosing to see, or choosing not to see, the name of the artist who made the painting. In the *explicitly biased condition*, participants then were asked to always choose to look at the name of the artist prior to judging the paintings.¹ In the *explicitly objective condition*, participants instead were asked to always choose *not* to see the name of the artist.

Before seeing the paintings, participants rated the objectivity of their assigned strategy (1 = *completely biased*, 9 = *completely objective*). Then, for each painting, they were asked to indicate via button press whether they would like to see the name of the artist prior to seeing the painting. After making their choice, and therefore seeing the artist's name or not, they saw the painting for 5 s. Then, they evaluated its artistic merit (1 = *no artistic merit*, 9 = *very high artistic merit*). This process continued for all 80 paintings. Finally, participants assessed how biased versus objective they had

been in evaluating the paintings (1 = *completely biased*, 9 = *completely objective*).

Materials. The 80 paintings were retrieved from the websites of internationally recognized art museums. Each painting was selected from a larger set (of 120 paintings) that were rated for their perceived quality by a separate sample of undergraduates ($n = 17$) who were untrained in art history and uninformed of the painters' identities. These ratings were used to produce two sets of 40 paintings of comparable perceived quality, so that one set could be imputed to famous artists, and the other to non-famous artists. For the paintings allegedly made by non-famous artists, we consulted a phone book to obtain names to assign to those paintings. For the paintings allegedly made by famous artists, we used the true artist's name except in cases where that name was not recognized by a separate sample of undergraduates ($n = 143$) untrained in art history. In those cases, we substituted a more recognizable artist from the same era and style (e.g., Picasso substituted for Braque).

Results and Discussion

Manipulation check: Adherence to strategy. As a manipulation check on whether participants used the evaluative strategy to which they were assigned, analyses were conducted on participants' decisions to look versus not look at the artists' names prior to rating each painting. Participants complied with their assigned conditions: In the explicitly biased condition, they chose to look at the name of the artist 96% of the time, and in the explicitly objective condition, they chose NOT to look at the name of the artist 99% of the time.

Assessments of bias. As in Experiment 1, participants rated the objectivity/bias of their assigned evaluative strategy prior to engaging in it. We predicted that participants in the explicitly biased condition would rate the strategy they used as significantly less objective than participants in the explicitly objective condition, but that they would nonetheless rate their resulting evaluations as similarly objective. This predicted 2×2 interaction effect (Evaluation Strategy: *Explicitly Biased*, *Explicitly Objective* \times Bias Assessment: *Self-Assessment*, *Assessment of Strategy*) emerged, $F(1, 72) = 19.13, p < .0001$ (see Figure 2).

We further explored this interaction effect by looking separately at participants' assessments of bias in their strategy versus themselves. A simple effects test of the effect of experimental condition on participants' assessment of objectivity/bias in their evaluation strategy was significant, $F(1, 72) = 34.21, p < .0001$. That is, participants in the explicitly biased condition viewed their strategy as less objective than participants in the explicitly objective condition viewed *their* strategy ($M = 4.00, SD = 2.12$ vs. $M = 6.78, SD = 1.96$). Moreover, and also as predicted, participants in the explicitly biased condition did not view their own evaluations as any less objective than did participants in the explicitly objective

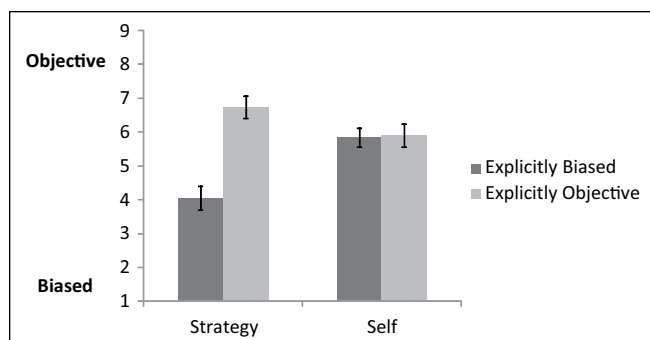


Figure 2. Participants' ratings of bias in their judgmental strategy and in their subsequent self-judgment (Experiment 2).

Note. Error bars indicate 1 SE above and below the mean.

condition ($M = 5.89$, $SD = 1.67$ vs. $M = 5.97$, $SD = 2.04$), $F(1, 72) = 0.03$, $p = .86$.

Bias in explicitly biased condition. Participants in the explicitly biased condition rated their assigned evaluative strategy as biasing, but they rated their own judgments resulting from that strategy as of similar objectivity to the ratings of participants who used an explicitly objective strategy. Did those who used the biased strategy *actually* show bias in their evaluations, or were they in fact no more biased than their peers? To answer that question, we examined whether participants in the explicitly biased condition saw greater “artistic merit” in the paintings attributed to famous artists than the paintings attributed to non-famous artists. Participants in the explicitly biased condition rated the artistic merit of paintings attributed to famous artists ($M = 6.19$, $SD = 0.87$) significantly higher than that of paintings attributed to non-famous artists ($M = 5.91$, $SD = 0.83$), $t(37) = 4.12$, $p = .0002$. Participants in the explicitly objective condition (who did not see the alleged names of the artists) rated the artistic merit of the two groups of paintings the same ($M = 5.82$, $SD = 0.96$ vs. $M = 5.80$, $SD = 1.05$), $t(35) = 0.36$, $p = .72$. The relevant interaction was significant, $F(1, 72) = 10.10$, $p = .002$. Moreover, the actual level of bias that participants displayed in the bias condition was not correlated with the amount of bias versus objectivity that they claimed. That is, greater displays of bias in evaluations of the paintings did not predict decreased claims of objectivity, $r(36) = .05$, $p = .75$.

In this experiment, participants who were instructed to use a biased strategy (for evaluating the artistic merit of paintings) rated that strategy as relatively biased, then engaged in it and showed the relevant bias, and, finally, denied having been biased. This experiment extends the results of the previous study by using an external method for assessing bias (i.e., a standard not reliant on between-condition comparisons) and by exploring our hypotheses in the context of a non-motivational bias.

One limitation of the preceding experiments is that the type of rating that participants made (i.e., whether they rated

themselves vs. their strategy) is confounded with the timing of their ratings, such that participants rated their strategy prior to engaging in the task but rated themselves afterward. This suggests the possibility that participants may have thought their strategy was biasing simply because they had not yet used it. If participants were to conclude, after using the strategy, that it was objective, then their rating of themselves as objective might simply reflect that new understanding of the strategy. To address this possibility, we conducted a third experiment in which participants rated both themselves and their strategy both before and after the task. This experiment also sought to reduce any possible tendency for participants to deny their bias as a means of self-enhancement by informing participants that the researchers would be able to empirically determine the participants' actual level of bias.

Experiment 3

Method

Participants. Eighty-five adults (52 women, 33 men) were recruited via Amazon's Mechanical Turk to complete the study online. The mean age was 35.66 ($SD = 12.30$).

Materials and procedure. The methods were the same as in Experiment 2, with three exceptions. First, participants rated both themselves and their decision strategy before and after the evaluation task. This change was made to clarify the role of task exposure on perceptions of bias. Second, to mitigate the possibility that participants did not understand the nature of the bias prior to the task (which would undermine the validity of the pre-task ratings), we provided further elaboration beforehand. We described the potential bias in detail by telling participants that looking at the name of the artist before seeing the painting might be biasing “because it could lead to different evaluations of the quality of the painting (in that paintings by famous painters could be rated more highly, regardless of their actual quality).” We also provided examples of several famous and non-famous artists' names that we randomly selected from the study so that participants could see for themselves the degree of fame and non-fame to which they might be exposed (e.g., Picasso, Matisse, Michelangelo; Zagia, Micon, Viele). Finally, the third change introduced in Experiment 3 involved reminding participants to be honest in providing their ratings of personal bias—and notifying them that we would “be able to assess the actual objectivity of [their] ratings.” This change was added to reduce any tendency to self-enhance by underreporting bias (based on the rationale that appearing dishonest would not be self-enhancing).

Results and Discussion

Manipulation check: Adherence to strategy. Participants generally complied with the assigned manipulation. In the

Table 1. Means and Standard Errors for Ratings of Bias, Separated by Timing of Rating (Pre- vs. Post-Task), Assessment Type (Self-Assessment vs. Assessment of Strategy), and Experimental Condition (Explicitly Biased vs. Explicitly Objective).

	Strategy ratings		Self-ratings	
	Objective condition	Biased condition	Objective condition	Biased condition
Prior to task				
M	3.93	8.20	4.98	8.40
SE	0.30	0.32	0.29	0.31
After task				
M	3.53	8.20	6.90	7.98
SE	0.27	0.29	0.27	0.29

explicitly biased condition, participants chose to look at the artists' names 83% of the time, and those in the explicitly objective condition chose NOT to look 99% of the time.

Assessments of bias. Participants rated the objectivity/bias of their assigned rating strategy and of themselves, both before and after using the strategy. A three-way interaction effect emerged (Timing of Assessment: *Pre-task*, *Post-task* × Evaluation Strategy: *Explicitly Biased*, *Explicitly Objective* × Bias Assessment: *Self-Assessment*, *Assessment of Strategy*), $F(1, 83) = 16.26, p < .0001$ (see Table 1). To decompose this interaction, we first separated the analyses (as in the previous studies) by the type of bias assessment in question (self-assessment vs. assessment of strategy). For participants' assessments of the bias in the strategy they used, there was a main effect of the type of strategy whereby, across time, participants in the explicitly biased condition rated their strategy as less objective than participants in the explicitly objective condition rated *their* strategy ($M = 3.73, SD = 2.31$ vs. $M = 8.20, SD = 1.35$), $F(1, 83) = 217.54, p < .0001$. There was no interaction between the timing of participants' assessments and the type of strategy they used, $F(1, 83) = 0.48, p = .49$.

For participants' assessments of their own bias, the effect of timing on bias assessments interacted with whether participants had been led to use an explicitly biased or an explicitly objective strategy, $F(1, 83) = 25.83, p < .0001$. For participants who were led to use an explicitly biased strategy, they reported being more objective after using the strategy than they predicted they would be prior to using it ($M_{pre} = 4.98, SD = 2.42$; $M_{post} = 6.89, SD = 1.87$), $t(44) = -5.01, p < .0001$. Thus, having using the biased strategy, they became yet more convinced of their objectivity. For participants who were led to use an explicitly objective strategy, there was a marginal tendency in the other direction, whereby assessments of objectivity were marginally reduced ($M_{pre} = 8.40, SD = 1.15$; $M_{post} = 7.98, SD = 1.76$), $t(39) = 1.83, p = .074$.

These results replicate the basic interaction pattern found in our previous experiments, in that participants saw their strategy as more biased in the biased condition than in the

objective condition, but they were far less likely to see *themselves* as more biased in the biased condition than in the objective condition—especially in the case of their post-task ratings, in which self-assessments of objectivity rose from pre-task.

Another way to look at these results is that, pre-task, participants in the explicitly biased condition felt that they would be somewhat more objective than the strategy they used—and, post-task, they felt yet even more objective relative to the strategy they used. This finding was not the result of a general increase in perceived self-objectivity, as it did not occur in the explicitly objective condition.

Bias in explicitly biased condition. Participants in the explicitly biased condition claimed objectivity after using their biased assessment strategy—indeed, they claimed more objectivity than they had prior to using the strategy. We next examined whether those participants had in fact been biased in the manner that they were asked about; that is, whether they rated the paintings that were randomly attributed to famous artists as of greater merit than those randomly attributed to non-famous artists. Participants in the explicitly biased condition rated paintings attributed to famous artists as of greater merit ($M = 6.24, SD = 1.10$) than those attributed to non-famous artists ($M = 6.00, SD = 1.06$), $t(44) = 3.99, p < .0001$. Participants in the explicitly objective condition (who did not see the alleged names of the artists) rated the artistic merit of the two groups of paintings the same ($M = 6.07, SD = .89$ vs. $M = 6.07, SD = .92$), $t(39) = -0.03, p = .98$. The relevant interaction was significant, $F(1, 83) = 9.40, p = .003$. Moreover, the actual level of bias that participants displayed in the bias condition was not correlated with the amount of bias versus objectivity that they then claimed. That is, greater displays of bias in evaluations of the paintings did not predict decreased claims of objectivity, $r(43) = .16, p = .28$.

These results, along with those of our previous experiments, suggest that individuals who use a biased strategy recognize it as biasing and yet perceive themselves to be relatively objective after using it. In this experiment, individuals who used a biased strategy claimed before using it that they would be reasonably objective in the face of it, and claimed afterward to have been even more objective than that. This temporal effect is of theoretical interest, and receives further attention in the ensuing General Discussion.

General Discussion

In the context of judgments and decisions in the real world of work and relationships, the costs of bias can be high. As a consequence, the ability to recognize one's biases—and, thereby, to work toward preventing, correcting, and overcoming them—is valuable. When individuals' judgments are influenced by strategies that they do not see as biased, it seems inevitable that they will fail to see their resulting judgments as

biased. This research suggests that the problem is deeper and more persistent than that. In a series of experiments, we found that even when individuals recognized bias in their judgmental strategies, they nonetheless claimed that their decisions emerging from those strategies were fairly objective. Moreover, we found that these claims of objectivity were on the whole false: Participants in the biased conditions did in fact display significant bias, and those who displayed more rather than less bias did not claim any less objectivity.

This research suggests the persistence of people's blindness to their own biases. Recognizing one's bias is a critical first step in trying to correct for it; these experiments make clear how difficult that first step can be to reach. The effects in this article were shown in two domains of bias commission. Experiment 1 showed the effects of knowingly engaging in a biased decision-making strategy in the context of a classic self-serving bias—that is, the tendency to criticize a test after performing poorly on it. Experiments 2 and 3 showed the effects in the context of a bias involving the tendency for people to view paintings as having more inherent artistic merit when those paintings are attributed to famous artists.

This research constitutes another chapter in a continuing story about the nature of the “bias blind spot” (e.g., Ehrlinger et al., 2005; Frantz, 2006; Pronin et al., 2002; West et al., 2012). Past research has shown that although individuals are quick to point out bias in others, they are far less likely to admit to bias in themselves. The current research supports and strengthens the notion that people have difficulty recognizing their own biases. It shows that even when people acknowledge that what they are about to do is biased, they *still* are inclined to see their resulting decisions as objective.

Those participants in our experiments who were induced to use biased decision-making strategies readily saw bias in those strategies but nonetheless claimed comparative personal objectivity. Could it be that their denials of personal bias partly reflected a desire to appear positively to the experimenter? Being biased is generally viewed as negative, and our participants may have wanted to distance themselves from this negative characteristic. A couple of features of our studies suggest that social desirability concerns are not the principle cause of our effects. First, it is unclear that bias denials would be viewed as socially desirable in the explicitly biased conditions. Rather, it might appear hypocritical for individuals to call their judgmental strategy biased and then, minutes later, claim objectivity in their resulting judgment. Second, participants in Experiments 2 and 3 knew that the experimenter would be able to discern whether their assessments had been biased by the identities of the painters. In Experiment 3, participants were explicitly warned about this eventuality and reminded that they should be honest about their degree of bias because the experimenters would, essentially, know if they were being deceptive. Therefore, especially in Experiment 3, concerns about social desirability should have magnified participants' interest in responding as

accurately as possible—to avoid the embarrassment of claiming objectivity when the experimenter could readily see their bias.

Although we cannot rule out the possibility that social desirability played a role in our effects, another explanation seems more compelling to us. Past research has traced people's bias blindness to their overweighting of introspective information when judging their own bias. Although bias tends to operate non-consciously (e.g., Ditto & Lopez, 1992; Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998; Williams & Gilovich, 2008; Wilson & Brekke, 1994), people look inward to judge their susceptibility, for example, consulting their *feelings* of whether they have been biased or their *intentions* to be objective. When judging others' bias, by contrast, people tend to look to those others' actions and to general theories about what bias “looks like” (Pronin, 2009; Pronin et al., 2004). When participants in our experiments assessed their own bias, they likely found a lack of introspective signs of bias and consequently inferred objectivity. Consistent with this theorizing, participants in Experiment 3 claimed to be more objective *after* having gone through the process of making judgments using a biased strategy than *before* doing so. Those participants used a strategy that they thought was biased, and thus they probably expected to feel some bias when using it. The absence of that feeling may have made them more confident in their objectivity. This explanation is consistent with past work showing that people admit to more bias in the abstract than they do in the context of a specific act of bias (Ehrlinger et al., 2005).

It is of further theoretical interest that participants in the biased condition in Experiment 3 did not come to view their *strategy* as more objective after using it than before using it. Perhaps those participants viewed their own objectivity as the result of overcoming (or avoiding) the influence of a biased process rather than as a result of the process itself being less biasing than they had expected. Past work has shown that introspective information related to bias is applied only to self-assessments of bias and does not decrease people's perceptions of others' bias. The present results suggest that this information also does not decrease people's perceptions of the bias associated with general strategies for making judgments.

The consequences of seeing oneself as objective while seeing one's judgmental strategy as biased may be serious. Consider a group of jurors who have just been exposed to testimony that they are now told to disregard as inadmissible. Each juror may admit a priori that being exposed to inadmissible evidence biases judgment. However, having been exposed to that information, each is likely to believe that his or her own judgment is objective (although that judgment likely took into account the “inadmissible” evidence; for example, Sommers & Kasson, 2001). Or, consider a team of human resources officials who have a hiring process in which they see photographs of applicants before evaluating the merits of their applications. Each official might view that

process as introducing a host of potential biases, but each might believe that they personally can overcome that bias (and, thereby, benefit from being able to “attach a face” to an application). This blindness may be especially likely to foster conflict between people who disagree after going through the same judgmental process. Each may not only be confident in his or her own objectivity but also quick to impute bias to the person who disagrees—and who was subject to a biased decision-making process.

Ironically, people’s recognition of bias in their judgmental strategies could strengthen their confidence in their personal objectivity. Consider the sports coach who explicitly decides to give his biggest players the best playing time during the pre-season and to evaluate, when pre-season ends, who should be in the starting line-up. The coach would readily acknowledge that his process is biased toward the big players. Compare him with a coach who does not make the same *explicit* choice but who, *implicitly*, ends up giving her biggest players the best playing time in the pre-season. The present research suggests that both coaches are likely to be biased toward the big players when they choose their starting line-ups, and that both coaches are likely to deny showing that bias. Ironically, the coach who recognized the potential for bias in his process may be especially confident in his objectivity—perhaps because he is keenly aware of the steps he took to avoid being influenced by that bias or even because he simply has the knowledge that he felt unbiased even in the face of it. This rings true in light of research by Uhlmann and Cohen (2007) which suggests that a feeling of personal objectivity leads individuals to view their beliefs as true and valid.

The current research explored actors’ self-assessments of objectivity following a biased judgmental strategy, but what of observers’ assessments? If a judge, hiring official, or athletic coach openly acknowledges the potential for bias in his or her judgment process, what do onlookers conclude about his or her resulting judgment? It seems possible that actors’ admissions of bias in their judgmental processes and their subsequent conclusions of objectivity could be mirrored by onlookers. Those observers might think something like: “If he was smart enough to know this bias existed, and honest enough to acknowledge it, surely he wouldn’t fall prey to it!”

This research contributes to an existing literature illustrating that efforts at debiasing are fraught with difficulty (e.g., Epley & Gilovich, 2006; Pronin et al., 2002; Stapel, Martin, & Schwarz, 1998; Wegener & Petty, 1995). People have difficulty correcting for biases that have already influenced their judgments, because they are unsure of how much to correct (and often deem any correction unnecessary). The present experiments suggest that people also have difficulty preventing biases from affecting their ongoing judgments, even when they can recognize a priori the potential for that bias. One message emerging from these findings is clear: Debiasing efforts may work best when they do not

aim to debias at all, but rather when they aim to prevent bias in the first place (e.g., Hansen & Pronin, 2012; Uhlmann & Cohen, 2005; Wilson & Brekke, 1994). When it comes to decisions in such varied realms as hiring, coaching, judging, and policy implementing, those wishing to be objective (or wishing to elicit objectivity from others) should insist on procedures that block biases from entering into the process. This strategy has been used, for example, in the context of symphony orchestras in which female under-representation has been dramatically reduced in part by having applicants audition behind a screen (Goldin & Rouse, 2000). Such efforts clearly can be an effective way to avoid bias. So why are we so resistant to using them? The present research provides one such answer: Such efforts are likely to seem needless when we believe that we can be objective even in the face of obviously biasing procedures.

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the National Science Foundation [BCS-0742394] and the Fuller Theological Seminary/Thrive Center in concert with the John Templeton Foundation [IH-114].

Note

1. In some instances, knowing the name of the artist (and, therefore, his or her degree of fame and/or talent) might not be a “biasing” influence on judgments of artistic merit, but rather a rational criterion to use in judging that merit. Because the present research concerns people’s *perceptions* about whether they are biased, we are less concerned with whether this influence is “objectively” biasing than with the fact that participants themselves would view it that way.

References

- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63, 568-584.
- Ehrlinger, J. E., 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., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, 17, 311-318.
- Frantz, C. M. (2006). I AM being fair: The bias blind spot as a stumbling block to seeing both sides. *Basic and Applied Social Psychology*, 28, 157-167.
- Gilbert, D. T., Pinel, 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.
- Goldin, C., & Rouse, C. (2000). Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90, 715-741.
- Hansen, K. E., & Pronin, E. (2012). Illusions of self-knowledge. In S. Vazire & T. D. Wilson (Eds.), *Handbook of self knowledge* (pp. 345-362). New York, NY: Guilford.
- Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology*, 51, 493-504.
- Pronin, E. (2009). The introspection illusion. *Advances in Experimental Social Psychology*, 41, 1-67.
- 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., & Kugler, M. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology*, 43, 565-678.
- 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.
- Schwarz, N., & Bless, H. (1992). Constructing reality and its alternatives: An inclusion/exclusion model of assimilation and contrast effects in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 217-245). Hillsdale, NJ: Lawrence Erlbaum.
- Sommers, S. R., & Kassir, S. M. (2001). On the many impacts of inadmissible testimony: Selective compliance, need for cognition, and the overcorrection bias. *Personality and Social Psychology Bulletin*, 27, 1368-1377.
- Stapel, D. A., Martin, L. L., & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin*, 24, 797-806.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16, 474-480.
- Uhlmann, E. L., & Cohen, G. L. (2007). "I think it, therefore it's true": Effects of self-perceived objectivity on hiring discrimination. *Organizational Behavior and Human Decision Processes*, 104, 207-223.
- Wegener, D. T., & Petty, R. E. (1995). Flexible correction processes in social judgment: The role of naïve theories in corrections for perceived bias. *Journal of Personality and Social Psychology*, 68, 36-51.
- Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. *Advances in Experimental Social Psychology*, 29, 141-208.
- West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, 103, 506-519.
- Williams, E., & Gilovich, T. (2008). Do people really believe they are above average? *Journal of Experimental Social Psychology*, 44, 1121-1128.
- 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 T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 185-200). New York: Cambridge University Press.