HARVARD | BUSINESS | SCHOOL

Is Overconfidence a Motivated Bias? Experimental Evidence

Jennifer M. Logg
Uriel Haran
Don A. Moore

Working Paper 18-099

Is Overconfidence a Motivated Bias? Experimental Evidence

Jennifer M. Logg
Harvard Business School

Uriel Haran
Ben-Gurion University of the Negev

Don A. Moore
University of California, Berkeley

Working Paper 18-099

Is Overconfidence a Motivated Bias?

Experimental Evidence

Jennifer M. Logg
Harvard University

Uriel Haran
Ben-Gurion University of the Negev

Don A. Moore
University of California, Berkeley

Revision Submitted 12.11.17

Author Note

Jennifer M. Logg, Harvard Business School, Harvard University; Uriel Haran, Guilford Glazer Faculty of Business and Management, Ben-Gurion University of the Negev; and Don A. Moore, Haas School of Business, University of California, Berkeley.

The Authors wish to thank Leif Nelson and Elizabeth Tenney for their helpful comments and insights and to Heather Yang and Silva Kurtisa for help running experiments. Thanks also to Sahaana Suri, Jonathan Wang, and Jennifer Georgevich for their assistance collecting data.

Correspondence concerning this article should be addressed to Jennifer M. Logg, Harvard Business School, Harvard University, Baker Library, Bloomberg Center 433, Harvard Business School, Boston, MA 02163. E-mail: jlogg@hbs.edu

Abstract

Are overconfident beliefs driven by the motivation to view oneself positively? We test the relationship between motivation and overconfidence using two distinct, but often conflated measures: better-than-average (BTA) beliefs and overplacement. Our results suggest that motivation can indeed affect overconfidence, but only under limited conditions. We find that motivation does indeed inflate BTA beliefs. However, introducing some specificity and clarity to the standards of assessment (Experiment 1) or to the trait's definition (Experiments 2 and 3) reduces or eliminates this bias in judgment. We find stronger support for a cognitive explanation for overconfidence, which emphasizes the effect of task difficulty. The difficulty of possessing a desirable trait (Experiment 4) or succeeding on math and logic problems (Experiment 5) affected overconfidence in ways that are consistent with the cognitive account proposed by prior research, above and beyond motivation. Finally, we find the lack of an objective standard for vague traits allows people to create idiosyncratic definitions and view themselves as better than others in their own unique ways (Experiment 6). Overall, the results suggest motivation's effect on overconfidence is driven more by idiosyncratic construals of assessment than by self-enhancing delusion. They also suggest that by focusing on vague measures (BTA rather than overplacement measures) and vague traits, prior research may have exaggerated the role of motivation in overconfidence.

Keywords: self-perception, overconfidence, motivation, Better-Than-Average effect, specificity

Abstract Word Count: 217/250

Is Overconfidence a Motivated Bias? Experimental Evidence

People claim to be better than others on a variety of traits and attributes, including honesty (Brown, 2011), leadership skills (Dunning, Heath, & Suls, 2004), popularity (Zuckerman & Jost, 2013), and safe driving (Svenson, 1981). Business people claim that their firms are better than the average firm (Cooper, Woo, & Dunkelberg, 1988; Larwood & Whittaker, 1977), engineers report that their work is superior to their peers' work (Zenger, 1992), and venture capitalists are overconfident in their ability to predict which entrepreneurs will succeed (Zacharakis & Shepherd, 2001). All of these predicted states of the world are desirable, but not always realistic. While prior theory and research has suggested that the desire to be better than others on a certain dimension drives individuals' inflated beliefs, we test this proposition experimentally and compare it with a cognitive account of overconfidence.

Overconfidence: BTA vs. Overplacement

Before examining what drives overconfidence, a clarification is needed; not all measures of overconfidence are created equal. The term *overconfidence* generally describes several constructs that measure inflated views of the self. This paper focuses on two: better-than-average (BTA) beliefs and overplacement. These terms are often used interchangeably, but differ in important ways (Burson, Larrick, & Klayman, 2006).

BTA beliefs are evident when the majority of people in a group claims that they are better than the median (e.g., when an entire class thinks they performed higher than the class's *median score*), which is mathematically impossible. Overplacement is manifested by an exaggerated estimate of one's standing relative to other individuals (e.g., when a student thinks she achieved a higher percentile ranking on a test than she *actually* did). BTA beliefs compare individual

beliefs to a *group-level* statistic, while overplacement compares individual beliefs to *individual-level* performance. The BTA measure is more vague than overplacement, and, though very widely used in research on overconfidence, is poorly suited to detecting bias in individual judgment (Benoît, Dubra, & Moore, 2015; Harris & Hahn, 2011). Overplacement allows researchers to differentiate between realistic and unfounded claims of superiority at the individual level (Krueger & Wright, 2011). Yet, researchers can only measure overplacement when an objective benchmark is available for assessing the accuracy of individual beliefs. We contribute to these two research streams by marrying them in experimental designs that allow us to identify interactions between motivation and specificity.

A Motivational Account of BTA beliefs

Many have argued that BTA beliefs are driven by the desire to view oneself positively (Dunning, 2005; Fabricius & Büttgen, 2013; Greenwald, 1980; Kunda, 1990; Radhakrishnan, Arrow, & Sniezek, 1996; Sedikedes & Gregg, 2008; Taylor & Brown, 1994). In fact, the assumption that flattering self-perceptions are motivated is so pervasive that some have claimed a "well-established connection between traditional optimism biases and motivated reasoning" (O'Brien, 2013) and that "the better-than-average bias is caused by our strong unconscious desire to maintain a positive self-view" (Chamorro-Premuzic, 2013). While it is entirely plausible that the motivation to view oneself in a positive light could drive excessively positive self-perceptions, causal evidence to support this claim is surprisingly sparse. Studies have found correlations between motivation and "better-than-average" beliefs for general traits, such as honesty and intelligence (Kunda, 1990). But as with all correlational evidence, these findings are amenable to several causal explanations. Other studies (e.g., Alicke & Govorun, 2005) have

found that BTA beliefs increase under ambiguity, but have not causally linked these increases to motivation.

We test the role of specificity as a moderator of the relationship between motivation and BTA beliefs. Our studies vary the ambiguity of the domain or construct, as well as the specificity of the measure that people use for rating themselves and others (BTA versus overplacement measures). The results allow us to reconcile discrepant findings and address limitations in the prior literature. In addition, we compare the effects of motivational influences on self-enhancement with well-established cognitive effects from the overconfidence literature.

Prior Evidence for a Motivational Account

Existing evidence for the motivational origins of BTA beliefs generally shares three limitations: reliance on correlational evidence, confounding trait commonness with importance, and vague performance standards. We first outline these shortcomings and then detail how our experiments address them.

Correlational evidence. Earlier studies on desirability and self-perception found that people are more likely to hold positive self-perceptions in domains they value (Alicke, 1985; Kunda, 1990; Mazar, Amir, & Ariely, 2008), including those valued distinctively by their own cultures (Sedikides, Gaertner, & Toguchi, 2003) and those useful for obtaining future goals (Dunning, 1995). These correlational findings, however, suffer from a number of shortcomings. One is that they leave open several different causal paths. People might express BTA beliefs for traits and skills in domains they initially value (e.g., if someone values honesty, they may overestimate how honest they are relative to others). Or they might assign greater importance to domains where they already consider themselves skilled (e.g., if someone observes their own honest behaviors, they may begin to place more value on honesty). Or they might, quite

sensibly, work to develop traits and skills in domains they believe are important (e.g., if someone values honesty, they may endeavor to behave honestly).

Dunning, Meyerowitz, and Holzberg (1989) found that people displayed a larger BTA effect for positive than negative traits. Specifically, participants ranked themselves in the 59th percentile, on average, for the traits "talented" and "athletic" but in the 39th percentile (below the median) for the traits "trouble handling money" and "socially anxious." As with so much of the evidence, this study did not exogenously manipulate motivation, but instead examined different domains that varied not only in valence, but also in other aspects. Consequently, there are many possible explanations for their results. It is possible, for instance, that Dunning et al. happened to select negative traits that were less vague or which people identified as rare (Kruger & Savitsky, 2009), which led them to consider that they embodied those traits less.

Confounded evidence. One noteworthy study manipulated motivation while holding a list of traits constant (Brown, 2011). BTA effects were larger for traits described as "important and rare" than as "unimportant and common." Unfortunately, this manipulation confounded importance with commonness, which appears to drive BTA beliefs more than does importance (Chambers, Windschitl, & Suls, 2004). People tend to think they are more likely than others to possess common traits and less likely than others to possess rare traits (Kruger & Burrus, 2004). It is therefore possible that Brown's results are more attributable to perceived commonness than to desirability or motivation.

Vague performance standards. A third limitation of the extant self-enhancement literature is a reliance on assessments within vague domains, using undefined, or poorly-defined criteria and measures of assessment (Alicke & Govorun, 2005). In most cases, BTA beliefs are elicited by asking participants how well certain traits (e.g., honest, kind, responsible, intelligent)

describe them. These traits are typically not defined and are open to different interpretations.

One problem with using vague personality traits and measures is that they are likely to overestimate bias if self-serving attributions are stronger for vague contexts and traits compared with more precise contexts and traits (Dunning et al., 1989; Sloman, Fernbach, & Hagmayer, 2010). Another problem is that ambiguous domains come with idiosyncratic assessment criteria (Weinstein, 1980; Weinstein & Lachendro, 1982). If people use different definitions to assess performance, then everyone can (correctly) claim that they are better than others (van den Steen, 2004). For example, some people may consider themselves honest if they fulfill their obligations, while others may consider themselves honest if they do not steal. In fact, if everyone has their own standards for what it means to be honest, then everyone can claim they are the most honest person in the world, and, by their own quirky standards, everyone would be correct.

A Cognitive Account of Overplacement

Although BTA beliefs and overplacement share psychological origins, the literatures examining these two phenomena have developed in parallel and have focused on different underlying mechanisms of the effect. Kahneman and Tversky (1996) viewed overconfidence as a cognitive bias, caused by errors in processing information. Work on overplacement, too, has offered a cognitive account for the bias (Chambers & Windschitl, 2004; Moore, 2007; Moore, Tenney, & Haran, 2016; Simon, Houghton, & Aquino, 2000), as did studies considering both BTA and overplacement (Burson, Larrick, & Klayman, 2006). These cognitive theories do a good job accounting for important features of the empirical evidence, such as the finding that people underestimate their performance on easy tasks and show underplacement when considering difficult tasks (Moore & Small, 2007; Moore & Healy, 2008). Do these cognitive

theories leave any room for motivational influences on self-assessment? We attempt to answer this question by comparing the effects of motivational and cognitive manipulations on overplacement. Our tests show why doing so matters for understanding the cause of overconfidence.

The Key to Motivational Effects: Vagueness

Prior research has found that vaguely-defined domains and traits exhibit greater self-enhancement (Dunning et al., 1989). Others assert that self-enhancement is driven by desirability and motivated reasoning (e.g., Brown, 2011). These two claims highlight an outstanding question: is ambiguity necessary to find an effect of motivation on BTA beliefs? Kunda (1990) argued that motivation cannot twist any fact to its end--some facts are more easily re-interpreted than are others (see also Armor & Sackett, 2006; Gilovich, Kerr, & Medvec, 1993). We suspect that vagueness is a crucial facilitating mechanism for allowing the expression of motivated impulses on biased beliefs. We test whether motivational effects found on BTA beliefs depend on vagueness.

Theorizing that vagueness is necessary for motivation to influence BTA beliefs, leads us to expect an interaction between motivation and specificity in our experiments. Prior research has not exogenously manipulated both motivation and specificity and thus has been unable to test this interaction. Some work in the motivated cognition literature has manipulated the importance of a single trait and measured *either* self-perceptions (Kunda & Sanitioso, 1989) *or* perceptions of others (Klein & Kunda, 1992). Yet, no research, to our knowledge, has manipulated the motivation to possess a single trait and then compared self-assessments with *reality*. In sum, the existing empirical record calls for further testing of the causal claim that motivation affects

overconfidence. It also calls for a test of how the specificity of construct criteria and measures (BTA beliefs vs. overplacement) interacts with motivation to affect overconfidence.

The Present Research: An Overview

Our work seeks to better understand what causes overconfidence by connecting the BTA and overplacement literatures that have developed in parallel. The self-enhancement literature (which relies primarily on BTA measures) has developed nuanced theories of motivation while the overconfidence literature (which relies on measures of overplacement) has focused more on cognitive causes and measurement issues. The two literatures are deeply related but rarely linked. We seek to connect them by providing an empirical test of a motivational explanation and directly comparing the strength of motivational and cognitive forces. In addition, we conduct experimental manipulations of both motivation and vagueness in order to test a causal connection between motivation and overconfidence in its different forms. If overconfidence is motivated, then people should display greater overconfidence for abilities or attributes they consider important. But manipulating motivation, rather than measuring its correlates, is key to identifying a causal relationship between motivation and overconfidence.

Specificity of performance standards. We systematically vary the specificity of people's self-assessments and examine its effects on the relationship between motivation and overconfidence. The experiments in this paper progress from vague to specific in their:

1) **Criteria** of assessment (vague traits to clearly defined traits). For example, people can assess how honest they are on a single measure or assess their honesty as defined by specific behaviors: "When I make a promise, I keep it" and "I do not say things I know to be untrue."

2) **Measures** of assessment (verbally labeled BTA measures to numeric elicitations of overplacement). For example, people can assess their honesty from "Not at all" to "Very" or assess their score on an 11-item honesty questionnaire from 0 to 11. The former measure is more common in the self-enhancement literature while the latter sort appears more often in the overconfidence literature.
3) **Domain** of assessment (personality traits to test performance). General personality traits, for example, are not objectively measureable. But when people assess how they did on a math test, their performance is based on the objective number of correctly answered questions.

In Experiments 1, 2, and 3, we manipulate motivation by varying how desirable it is to possess a trait. Experiment 1 employs assessments of a single trait (introversion), defined vaguely. Experiment 2 introduces a specific, numeric measure of assessment (overplacement) for introversion. Experiment 3 employs a new method for operationalizing specificity, by eliciting both BTA beliefs and overplacement for an unfamiliar trait, which we invented for purpose of the experiment.

Experiment 4 compares how motivational and cognitive factors affect BTA beliefs and overplacement. Experiment 5 compares how motivation influences BTA and overplacement measures within an objective domain, performance on a math and logic questions. Unlike subjective trait assessments, using an objective domain provides a benchmark for comparing participants' beliefs against reality. It also increases verifiability of assessments, which should suppress bias if motivation only affects vague self-assessments.

Our manipulations of importance represent an attempt to understand how motivation affects overconfidence more broadly. But any such attempt is incomplete without an

examination of what people mean when they claim they are better than others. Therefore, in Experiment 6, we vary trait criteria specificity within-subjects. Doing so allows us to observe the emergence of overconfidence on vague measures, its reduction through clearer criteria, and whether it subsequently re-emergence with measures that allow for idiosyncratic definitions. The within-subject manipulation also allows us to examine whether people's trait construals are idiosyncratic and whether these construals play a role in driving inflated relative self-perceptions. Table 1 summarizes the experimental designs.

For all of our experiments, we report how we determined sample sizes, pre-registered data exclusions, and all conditions. We determined sample sizes a priori, striving for at least 80% power. When possible, these power analyses relied on effect sizes revealed in prior experiments in the paper. Where that was not possible, we estimated smaller effects for a conservative test (using larger sample sizes). Final sample sizes include the number of participants after removing survey responses based on pre-registered exclusion criteria (Experiment 1: N = 200, Experiment 2: N = 666, Experiment 3: N = 391, Experiment 4: 359, Experiment 5: N = 111, and Experiment 6: 136). We report all exclusions as well as the results without any exclusions. Materials, data, syntax, and our pre-registrations for these experiments are posted on the Open Science Framework (https://osf.io/qayhz/?view_only=17e6b3b3969a4f81a8991a36f93a8d7c).

Experiment 1: Self-assessments of a vague vs. specific trait

Experiment 1 tests the effect of motivation and specificity on BTA beliefs. Prior research on overconfidence has found that people see positive traits as more characteristic of themselves than of others. However, specifying the definitions of such traits should reduce

variation in construals of performance, which should attenuate overconfidence (Preuss & Alicke, 2009) let alone any potential effect of motivation on overconfidence. Therefore, we varied the specificity of the trait's description, predicting that this would moderate the effect of motivation.

Whereas prior work compared different traits that might vary in importance, we focused on trait: introversion. Introversion has the advantage that people can view it as either desirable or undesirable (Cain, 2013). We manipulated the motivation to view oneself as introverted by varying the trait's perceived importance, and measured the extent to which people viewed themselves as more introverted than others.

Method

Participants

Two hundred twelve people (109 women, 103 men; *M* age = 30.22, *SD* = 12.20) completed two ostensibly unrelated surveys. Fifty-seven participated in the study in the lab at a West Coast university in exchange for course credit whereas the 155 others participated via Amazon's Mechanical Turk for a fee of $0.45.[1] Our sample size, determined ex-ante, sought two hundred and ten participants to detect a medium-sized effect ($f = 0.25$; $d = 0.5$) with 95% power. After excluding 12 participants (details below), a final sample of 200 remained.

Design

The experiment had a 2 (motivation: extroversion-important vs. introversion-important) × 2 (specificity: vague criteria vs. specific criteria) between-subjects design. We measured BTA beliefs by the mean difference between participants' ratings of introversion for themselves and others.

[1] Neither the BTA nor manipulation check measures differed significantly between these samples, ts < 1.44, ps > .14. The number of participants who failed the manipulation check did not differ between the mTurk and lab samples, either, $p = .31$, correcting for unequal variances.

Procedure and Materials

Motivation manipulation. Participants completed two ostensibly unrelated surveys on personality traits and leadership. The first survey manipulated the motivation to possess introversion by manipulating its importance. Participants read about one trait that helps people achieve success. In the introversion-important condition, participants read about introversion:

> …Skills associated with introversion may help people succeed in different areas of life. Introverted people are empathetic as well as good listeners (Wall Street Journal (WSJ), 2011), which allows them to gain trust from different kinds of people…

In the extroversion-important condition participants read about extroversion:

> …Skills associated with extroversion may help people succeed in different areas of life. Extroverted people are energetic and talkative which allows them to get along well with different kinds of people (New York Times (NYT), 2010)…

Details of the pre-tests for both manipulations are provided in Appendix A. Following the passage, participants listed two examples from their lives of how introversion [extroversion] made them or someone they know a good leader.

Criteria specificity manipulation. On the following page, participants took an ostensibly unrelated survey in which they rated how well various traits described them and others. The vague condition included only the trait names (introversion, outgoing [extroversion], conscientious, imaginative, agreeable, and honest), whereas in the specific condition, participants read each trait as specified with five relevant behaviors (e.g., introverted: I work alone when I can rather than with a group). Traits and behaviors appear in Table 2. Obviously, only introversion and being outgoing are relevant to our purposes here; we included the other traits to

reduce experimental demand and increase the plausibility of the claim that the two surveys were unrelated.

BTA measure (vague). Participants rated how well each trait described them and most other people on a verbally labeled scale ranging from 1 = *Not at all* to 5 = *Very*. Both the order of the traits and self- and other-ratings were randomized. We measured BTA beliefs indirectly by calculating the difference between self- and other-ratings. This sort of indirect measure is a more conservative measure of BTA beliefs than direct measures, which consist of a single comparative assessment and typically produce stronger BTA beliefs (Chambers & Windschitl, 2004; Moore, 2007; Otten & van der Pligt, 1996).

Motivation manipulation check. On the last page of the experiment, participants rated how important they thought it was for a person to possess the attributes of introversion and extroversion. Each rating was on scale ranging from 1 = *Not important at all* to 5 = *Extremely important*.

Results

Motivation Manipulation Check

Excluding twelve participants who failed the manipulation check,[2] those in the introversion-important condition rated introversion as more important (M = 3.05, SD = 0.71) than those in the extroversion-important condition (M = 2.58, SD = 0.92), $t(195.43) = 4.10$, $p < .001$, $d = 0.57$, correcting for unequal variances. Including those twelve participants in the analysis reduces the effect of the motivation manipulation on the manipulation check (high: M = 2.91, SD = 0.87; low: M = 2.69, SD = 1.03), $t(209.23) = 1.70$, $p = .09$, $d = 0.23$, correcting for

[2] Twelve participants failed the manipulation check by either rating introversion as a 1 (not important at all) or extroversion as 5 (extremely important) if they were in the introversion-important condition, or by rating extroversion as a 1 and introversion as a 5 in the extroversion-important condition.

unequal variances. The analyses below exclude the twelve but this exclusion does not materially alter the results.

BTA Effect

In aggregate, participants believed that they were more introverted (M = 3.23, SD = 1.62) than others (M = 2.77, SD = 0.90), $t(199) = 3.50$, $p = .001$, $d = 0.35$. In order to compare the magnitude of beliefs between motivation conditions, we subtracted ratings of others from self-ratings for each condition. A 2 (motivation: introversion-important vs. extroversion-important) × 2 (criteria specificity: vague vs. specific) between-subjects ANOVA yielded significant main effects of both motivation, $F(1, 196) = 18.37$, $p < .001$, *partial* $\eta^2 = .09$, and criteria specificity, $F(1, 196) = 3.93$, $p = .05$, *partial* $\eta^2 = .02$, and a significant interaction, $F(1, 196) = 5.57$, $p = .02$, *partial* $\eta^2 = .03$. As Figure 1 shows, the importance of introversion increases BTA beliefs when the trait was presented vaguely, $F(1, 196) = 20.35$, $p < .001$, *partial* $\eta^2 = .09$, but this effect is attenuated when introversion was specified and explicitly defined, $F(1, 196) = 2.03$, $p = .16$, *partial* $\eta^2 = .01$. Analyzing ratings of self and others separately in a repeated measures design yielded the same results.

Discussion

Our first challenge to the assumption that motivation affects BTA beliefs resulted in finding that such an effect exists, but in a limited capacity and under a stringent condition. We only found that motivation inflated BTA beliefs for a vaguely-defined trait, and not when its criteria were specified. These results extend Dunning et al.'s (1989) result that specificity decreases BTA beliefs; they show that vagueness *interacts* with motivation to inflate overconfident beliefs. In Experiment 2, we conduct a proper test of overconfidence by comparing self-reports of a vague trait with specific, objective scores on a test of the trait.

Experiment 2: Specifying Measures of Assessments for Introversion

Experiment 2 tests a new specificity-related moderator on the relationship between motivation and overconfidence. Whereas Experiment 1 found an effect of *criteria* specificity on BTA beliefs, Experiment 2 varies the specificity of *assessment measures* by administering both BTA and overplacement measures. We predicted that overly-generous self-ratings, produced by self-serving interpretations of a trait, diminish when BTA ratings are calibrated against a specific mode of assessment. We test this prediction by comparing people's assessments of themselves and others to actual scores on an introversion questionnaire. Eliciting assessments on vague (BTA beliefs) as well as specific (overplacement) measures of overconfidence enables us to test the effect of motivation on overconfidence at the individual level.

Method

Participants

Consistent with our pre-registered research plan, we collected data from 666 Mechanical Turk workers (342 women, 324 men; *M* age = 34, *SD* = 12.17), each paid $0.50. We estimated a sample size of 666 to detect an interaction with repeated measures that correlated, $r = -.193$, when z-scored, $d = .1679$, with 80% power. We based this calculation on the results of a prior study.

Design

The experiment had a mixed 2-cell (motivation: extroversion-important vs. introversion-important) between-subjects design. We asked people to assess themselves on vague and specific measures. As in Experiment 1, all participants read a manipulation passage and then completed measures of overconfidence. The main difference was that here, they answered both vague (BTA) and specific (overplacement) measures of overconfidence.

Procedure and Materials

The procedure and materials were the same as in Experiment 1 except for two major differences. First, rather than specifying what introversion meant by enumerating its specific behaviors, we administered McCroskey's (1997) 18-item introversion questionnaire.[3] Only after the questionnaire were participants informed that it measured introversion. Second, in addition to rating themselves and others on a vague scale, participants made specific estimates of their own and others' scores on the introversion questionnaire, directly following its completion.

Motivation manipulation. The manipulation was the same as Experiment 1. Participants completed what were described as two separate surveys on personality traits and leadership. Participants read that either introversion (introversion-important condition) or extroversion (extroversion-important condition) were conducive to personal success.

BTA measure (vague). As in Experiment 1, participants assessed how well each trait described themselves and most other people. They did so on a verbally labeled scale ranging from 1 = *Not at all* to 5 = *Very*.

Overplacement measure (specific). Participants estimated their own and others' scores on the McCroskey questionnaire. We measured overplacement by first calculating the difference between each participant's estimated and actual scores on the McCroskey questionnaire. The scores ranged from 12 to 36. We then subtracted the difference between participants' own and others' *actual* scores from the difference between the scores they *estimated* for themselves and others.

One key feature of this measure is that it cannot distinguish bias from error (Krueger & Wright, 2011). Unless participants are perfectly accurate estimating their relative placement, they

[3] We followed McCroskey's (1997) scoring scheme: we subtracted the sum for the extroversion items from the sum for the introversion items plus 40.

will show up as either over- or underplacing. Although a motivated bias is likely to produce overplacement, it can also result from cognitive error, especially when the task is an easy one (Heck & Krueger, 2015). Here, we observe an effect of motivation leading to bias through our manipulation of motivation. In Experiment 4, we cross a motivation manipulation with a manipulation of difficulty in order to compare the sizes of these effects and identify the relative influence of bias vs. error in driving estimates of placement relative to others.

Results

Motivation Manipulation Check

An independent samples t-test reveals that participants in the introversion-important condition rated introversion as more important (M = 3.11, SD = .78) than those in the extroversion-important condition (M = 2.54, SD = .88), $t(663.61) = 8.82$, $p < .001$, $d = 0.68$, correcting for unequal variances.

Effect of Motivation on Responses to the Introversion Questionnaire

We checked whether our manipulation influenced participants' responses on the McCroskey introversion questionnaire. Indeed, participants in the introversion-important condition had higher introversion scores (M = 25.88, SD = 6.43) than those in the extroversion-important condition, (M = 24.00, SD = 6.66), $t(663.61) = 8.82$, $p < .001$, $d = .15$, correcting for unequal variances.

Effect of Motivation on BTA and Overplacement

We standardized the vague and specific measures and submitted them to a 2 (motivation: introversion-important vs. extroversion-important) × 2 (specificity of measure: vague vs. specific) mixed ANOVA with specificity as a repeated measure. The analysis yielded a significant main effect of motivation, $F(1,664) = 49.25$, $p < .001$, *partial* $\eta^2 = .07$, and an

interaction between motivation and specificity, $F(1, 664) = 8.10$, $p = .006$, *partial* $\eta^2 = .01$.[4] Figure 2 shows that the specificity of the measure decreased the effect of motivation on overconfidence. Because participants' actual scores were affected by motivation in the same direction as their self-assessments, these exaggerated beliefs have emerged above and beyond differences in actual scores.

Discussion

Experiment 2 replicates Experiment 1's primary result, showing an effect of motivation on BTA beliefs for vague personality traits. Personality traits, as usually studied, are vague enough that they allow people to construe the meaning of a trait in a way that provides little verifiability of assessment. It is possible that high perceived importance of introversion led people to claim greater introversion because they were trying to manage impressions of the experimenter. Another possibility is that the motivation manipulation influenced the way participants thought about what it meant to be introverted such that they identified more ways in which their behavior could qualify as introverted. However, even within the subjective domain of personality traits, motivation affected vague (BTA) measures much more than it did specific (overplacement) measures. This result suggests that specificity of measures may suppress the effect of motivation on overconfidence, even within a subjective domain, and that the influence of motivation on overconfidence is tenuous.

An alternative explanation to the moderating effect of specificity is that participants were already familiar with their own level of introversion. We might find a stronger effect of motivation on overplacement for less familiar assessment domains (most people have probably

[4] On overplacement, those in the introversion-important condition did exaggerate the degree to which their introversion scores were higher than those of others ($M = 2.42$, $SD = 8.30$), and they did so more than those in the extroversion-important condition ($M = .72$, $SD = 8.74$), $t(657.57) = 2.57$, $p = .01$.

considered how introverted they are). We tested this proposition in Experiment 3 and presented participants with an unfamiliar trait, about which they did not have strong prior beliefs.

Experiment 3: Manipulating the Desirability of an Unfamiliar Trait

Although the vague BTA measure in Experiment 2 gave participants' more leeway to construe their own introversion in self-flattering ways, the fact that introversion is a well-known trait means that each participant already had some sense of his or her level of introversion. In Experiment 3, we invented a trait, social responsiveness, in order to test the effect of motivation on vague and specific self-assessments for an unfamiliar trait.

Method

Participants

We pre-registered a sample of 200 participants online via Amazon Mechanical Turk based on an estimation of the appropriate sample size. The analyses below include 391 participants (198 women, 198 men, *M*age = 33).[5]

Design

The experiment had a 2 (motivation: low vs. high) × 2 (order of measures: vague first vs. specific first) between-subjects design. The dependent variables included counterbalanced vague (BTA) and specific (overplacement) measures. For this experiment and all following, please see our file drawer for a description of other measures, overestimation and overprecision, and their

[5] Because the study's success depended on the successful manipulation of participants' beliefs about the value of social responsiveness, our pre-registered exclusion criteria would have had us drop participants in the high motivation condition who responded below a 4 on the manipulation check, and participants in the low motivation condition who responded above 2. Because this stringent criterion would have led us to drop so many cases, we wound up collecting complete surveys from 391 participants to reach our planned sample size. Subsequently, we concluded that excluding data from so many participants was problematic. Appendix B reports the same analyses with the smaller sample. The results are not materially different.

results. We posted these online:

https://osf.io/qayhz/?view_only=17e6b3b3969a4f81a8991a36f93a8d7c .

Procedure and Materials

Participants answered a social responsiveness questionnaire that we created. It included 14 short statements which described various behaviors and attributes of people. We asked each to, "rate how much each sentence accurately describes you, as you are today (not as you once were or strive to be)." Items included, among others, "People like to talk to me about various subjects" and "I can sense when a friend is in a bad mood." Participants responded to each item on a scale ranging from 1 = *Not at all* to 7 = *Completely*. For all 14 statements, see the survey materials posted online.

Motivation manipulation (desirability)

Following this questionnaire, participants encountered our manipulation, the perceived desirability of social responsiveness, the invented trait. Those in the high motivation condition read that people high in social responsiveness were more fulfilled, happier, and healthier. Those in the low motivation condition read that those low in social responsiveness were more comfortable with themselves, successful, and happy (the full passages are in Appendix C). Next, to reinforce the manipulation, all participants typed out two examples of how social responsiveness (either higher or lower) "has contributed to success or happiness in your life."

A comprehension check followed the manipulation and tested whether participants and read the manipulation passage. Then, participants answered a second manipulation check, by rating "How desirable do you think it is for a person to have high social responsiveness*?*" on a scale ranging from 1 = *Undesirable* to 5 = *Desirable*.

Measures

All participants assessed themselves and others using both specific and vague measures. The order manipulation varied whether the vague or the specific measures came first.

BTA measure (vague). There were two vague measures: "To what degree do you believe that you are high in social responsiveness?" and "To what degree do you believe that the average participant in this study is high in social responsiveness?" Participants answered each question on a scale ranging from 1 = *Very low* to 7 = *Very high*.

Overplacement measure (specific). Participants reported their beliefs about their own and others' scores on the 14-item social responsiveness questionnaire. As in the previous studies, we measured overplacement using their average item score, as well as their estimates of the mean score of other participants. In order to impress upon participants our interest in accurate responding, we included a header on the page: "Please try to be as accurate as you can in answering these questions." Moreover, we rewarded participants' accuracy in each of the two estimates: "The closer your estimate is to the truth, the better your chances of winning a $25 prize."

Results

Participants completed the social responsiveness questionnaire before they encountered the motivation manipulation, which made it impossible for the manipulation to affect their scores. Indeed, responses on the questionnaire were similar among those in the high motivation (M = 5.11, SD = 1.04) and low motivation (M = 5.13, SD = .94) conditions, $t(389) = .145$, $p = .884$. The manipulation check reveals that those in the desirable condition rated social responsiveness as more desirable (M = 4.47, SD = .79) than did those in the undesirable condition (M = 2.67, SD = 1.17), $t(333.27) = -17.79$, $p < .001$, correcting for unequal variances.

Estimations of Performance (Self-Ratings)

We standardized both specific score estimates and vague ratings of own score and submitted them to a 2 (motivation: low vs. high) × 2 (order of measures: vague first vs. specific first) × 2 (specificity of measure: vague vs. specific) mixed ANOVA with specificity of measures as a repeated measure. The analysis yields an effect of desirability, $F(1, 387) = 41.49$, $p < .001$, *partial* $\eta^2 = .097$, but not a significant interaction between desirability and specificity, $F(1, 387) = 3.07$, $p = .081$, *partial* $\eta^2 = .008$. This was qualified by a three-way interaction, $F(1, 387) = 5.66$, $p = .018$, *partial* $\eta^2 = .014$. This three-way interaction suggests that the effect of desirability is strongest on vague measures, especially when the vague measures come first. None of the other main effects or interactions is significant.

Overplacement and BTA

We measured overplacement by subtracting an individual's belief that his or her score is better than others' average scores, correcting for the degree to which they actually are better than average: $(estimated\ own\ score - estimated\ others'mean\ score) - (actual\ own\ score - actual\ others'mean\ score)$. In order to compare overplacement with the BTA measure, we standardized and submitted them to a 2 (motivation: low vs. high) × 2 (order of measures: vague first vs. specific first) × 2 (specificity of measure: overplacement vs. BTA) mixed ANOVA. The results reveal a significant main effect of desirability, $F(1, 387) = 45.69$, $p < .001$, *partial* $\eta^2 = .077$. The desirability × specificity interaction is not significant, $F(1, 387) = 3.68$, $p = .06$, *partial* $\eta^2 = .09$, but the effect of motivation is directionally stronger for the vague than the specific measure. None of the other main effects or interaction effects are significant, $ps > .23$.

Discussion

The desirability manipulation in Experiment 3 seemed to have had a stronger effect on overplacement than in Experiment 2. We believe that this is in large part because of the unfamiliar (made-up) trait on which we elicited self-assessments; peoples' beliefs were more pliable and thus more susceptible to influence by motivation. We still have yet to test a cognitive account of overconfidence and there, explicitly compares cognitive and motivational effects on overconfidence in Experiment 4.

Experiment 4: Cognitive vs. Motivational Processes

Experiment 4 compares cognitive and motivational accounts for overconfidence. Cognitive accounts for the bias highlight a key component of task difficulty. Therefore, in Experiment 4, we manipulated both motivation and difficulty. Moore and Healy (2008) show that overplacement and BTA beliefs are highest on very easy tasks, but reverse on hard tasks. In Experiment 4, we tested the effects of desirability and task difficulty on BTA measures and overplacement. As in Experiment 3, we manipulated the desirability of social responsiveness; additionally, we manipulated how difficult it was for participants to claim they were socially responsive.

Method

Participants

We obtained completed questionnaires from 426 participants via Amazon's Mechanical Turk, each paid $.50. Our pre-registered exclusion criteria led us drop data from 47 participants who failed the attention check and another 20 who completed the survey in under 5 minutes or more than 25 minutes. That left us with 359 participants (165 women, 194 men, *M*age = 34),

just over our planned sample size of 356. We estimated that sample size ex-ante to detect an estimated small interaction (d = 0.20) between desirability and specificity with 80% power.

Design. The experiment had a 2 (motivation: low vs. high) × 2 (criterion difficulty: low vs. high) × 2 (order of measures: vague first vs. specific first) between-subjects design. As in Experiment 3, we manipulated motivation by varying the desirability of social responsiveness, an unfamiliar trait and measured participants' overconfidence using vague (BTA) and specific (overplacement) measures, counterbalancing the order of specificity. Additionally, we manipulated the difficulty of the criterion for possessing the trait.

Procedure and Materials

The procedure and materials were similar to Experiment 3 except for one major difference. We altered the social responsiveness questionnaire in order to manipulate how difficult it was for participants to claim that they were socially responsive.

Manipulations and measures. We varied motivation using the same desirability manipulation as Experiment 3, and then administered a 13-item social responsiveness questionnaire. Participants answered "yes," "no," or "unsure" as to whether each of the statements described them (instead of answering how much each described them from a 1 to a 7). The questionnaire items varied between subjects; we manipulated difficulty by varying the stringency of the threshold for being able to answer "yes" to each item on the questionnaire. Half the participants were presented with a difficult threshold. For example, "In the past *day*, there have been *at least five times* where I have told a white lie to avoid hurting someone else's feelings." The other half had to meet a lower bar for claiming they were socially responsive, "In the past *year*, there has been *a time* where I have told a white lie to avoid hurting someone else's

feelings" (italics used for emphasis here and not in the materials). We used the same vague and specific measures of overconfidence as in Experiment 3.

Results

Manipulation checks

All participants completed the social responsiveness questionnaire before the desirability manipulation, so that it could not affect their responses to the questionnaire. As expected, a 2 (motivation: high vs. low) × 2 (difficulty: high vs. low) ANOVA found a main effect of *difficulty* on participants' scores, $F(1, 355) = 461$, $p < .001$, *partial* $\eta^2 = .06$, but no effect of desirability, $F(1, 355) = .02$, $p = .90$, *partial* $\eta^2 < .001$, or an interaction, $F(1,335) = 0.50$, $p = .48$, *partial* $\eta^2 = .001$.

As expected, participants in the difficult condition estimated lower scores ($M = 5.68$, $SD = 3.14$) than participants in the easy condition ($M = 8.36$, $SD = 2.73$), $t(347.62) = 8.64$, $p < .001$, $d = -.91$, adjusting for unequal variances. The desirability manipulation worked as well: participants in the high motivation condition thought social responsiveness was more desirable ($M = 6.24$, $SD = 0.87$) than participants in the low motivation condition ($M = 2.84$, $SD = 1.56$), $t(247.13) = 24.88$, $p < .001$, $d = 2.69$, adjusting for unequal variances. To compare the effect sizes of each manipulation, we converted Cohen's *d* effect sizes to Pearson correlations and compared them using a Fisher test. This analysis suggests that the *desirability* manipulation was stronger than the difficulty manipulation, $z = 4.78$, $p < .001$.

Estimations of Performance (Self-Ratings)

We standardized both specific score estimates and vague self-ratings and submitted them to a 2 (motivation: low vs. high) × 2 (difficulty: high vs. low) × 2 (order of measures: vague first vs. specific first) × 2 (specificity of measure: vague vs. specific) mixed ANOVA with specificity

of measures as a repeated measure. The results reveal three significant two-way interactions. As in Experiment 3, there was a significant interaction between specificity and desirability, $F(1, 355) = 29.51$, $p < .001$, *partial* $\eta^2 = .078$. This finding suggests that desirability had a more powerful effect on vague than specific measures. The interaction between specificity and order, $F(1, 355) = 5.95$, $p = .02$, *partial* $\eta^2 = .017$, suggests that although self-assessments were lower when the specific measures came first, this effect was particularly dramatic for the vague measures. Specificity and difficulty also displayed a significant interaction, $F(1, 355) = 46.07$, $p < .001$, *partial* $\eta^2 = .116$, suggesting that difficulty had a larger effect on specific than vague measures. Figure 3 presents these three interactions.

Overplacement and BTA

We standardized participants' (vague) BTA and (specific) overplacement measures and submitted them a 2 (motivation: low vs. high) × 2 (difficulty: high vs. low) × 2 (specificity of measure: vague vs. specific) mixed ANOVA. Specificity interacted with both difficulty, $F(1, 351) = 7.60$, $p = .006$, *partial* $\eta^2 = .021$, and desirability, $F(1, 351) = 7.30$, $p = .007$, *partial* $\eta^2 = .020$, as shown in Figure 4. We neither observes an interaction between desirability and difficulty nor a three-way interaction, suggesting that motivation and difficulty did not differentially affect the measures. Figure 4 presents these interactions and shows that, directionally, desirability influenced the vague measures more than the specific measures and difficulty influenced the specific measures more than the vague measures.[6]

[6] Our specific measures also afford analyses of participant's accuracy in their self-assessments. We conducted a 2 (motivation: low vs. high) × 2 (difficulty: high vs. low) × 2 (order of measures: vague first vs. specific first) mixed ANOVA separately for overplacement. As expected, there were main effects of desirability, $F(1, 351) = 11.12$, $p = .001$, $\eta^2 = .03$, and difficulty, $F(1, 351) = 23.08$, $p < .001$, $\eta^2 = .06$. Participants underplaced their scores in the difficult, ($M = -1.19$), $t(176) = -4.89$, $p < .001$, but not the easy condition, ($M = .29$), $t(181) = 1.43$, $p = .15$. Participants underplaced their scores in the low motivation condition, ($M = -1.02$), $t(164) = -4.04$, $p < .001$, but not in the high motivation condition, ($M = .05$), $t(193) = .24$, $p = .81$. Difficulty influenced overplacement more than desirability did, but the difference is not significant, $z = 1.04$, $p = .30$.

Discussion

This experiment's findings replicated Experiment 3's effect of desirability on self-assessments; when respondents viewed social responsiveness as desirable, BTA measures inflated more than overplacement did. These results are not due to a floor effect of desirability on specific self-ratings; instead, it appears that desirability had a weaker influence than difficulty did. In fact, we found that difficulty had a larger effect on overplacement than BTA measures.

This is especially surprising because the manipulation check revealed a strong effect of desirability, suggesting our test of overconfidence was a rather conservative one. Desirability and difficulty both affected vague (BTA) and specific measures (overplacement). Our findings that difficulty produced *less* overplacement is consistent with the cognitive account for overconfidence (Moore & Healy, 2008; see footnotes 10 and 11).

Whereas Experiments 1 and 2 documented the influence of motivation on vague measures (BTA beliefs) for a vague trait, Experiments 3 and 4 found that the effect on specific (overplacement) measures is weaker. The next experiment sought to test whether the effect of motivation endures even for performance that is specifically measurable and verifiable.

Experiment 5: Estimating One's Own Intelligence

Results of the first four experiments suggest evidence for a causal effect of motivation on overconfidence, but that this effect is limited to vaguely-defined assessments of vaguely-defined trait domains. In the previous experiments, we purposefully focused on vague traits in order to create an environment that would prove most amenable to finding motivational effects. In Experiment 5, we tested whether the same patterns might extend to objectively verifiable performance. Unlike judgments of personality traits, to which interpretation and subjectivity are inherent, answering scorable knowledge questions is objectively verifiable and measureable. Participants in this experiment answered math and logic questions and assessed their

performance using both vague and specific measures. Building on the patterns we observed in Experiments 1 and 2, we predicted that the relationship between motivation and overconfidence would weaken within a more objective domain. We varied intrinsic motivation by describing the implications of correctly answering questions.[7]

Method

Participants

One hundred and eleven students and staff at an Eastern university (51 women, 60 men; *M*age = 27, *SD* = 11.23) completed this experiment. We determined the sample size, prior to data analysis, on the number of participants we expected to realistically recruit in five experimental sessions. Experiment 4's motivation manipulation produced an effect size of d = 1.2 on the vague BTA measure, which a sample size of 111 should allow us to detect with 99% probability.

Design

The experiment had a 3-cell (Importance: low vs. medium vs. high) between-subjects design. Participants made assessments on both vague (BTA) and specific (overplacement) measures.

Procedure and Materials

Participants completed a ten-item test of math and logic puzzles with items taken from online IQ tests. We described the task differently in order to manipulate motivation.

[7] For this experiment, we manipulated both intrinsic and extrinsic forms of motivation to succeed at this task but for the sake of clarity, we report the results of the intrinsic manipulation of motivation. The original design was a 3 (Importance: low vs. medium vs. high) × 2 (Monetary incentive: present vs. absent) between-subjects design. For all results and a discussion of the monetary (extrinsic) incentives, see our file drawer on OSF.

Participants estimated their own and others' performance before and after answering the questions.

Motivation manipulation.

We orthogonally manipulated intrinsic motivation by manipulating how we described the task, thus manipulating its perceived importance and participants' motivation to perform well.

Participants in the high motivation condition read:

> In this experiment, you will be taking an intelligence test. Intelligence, as you know, is an important dimension on which people differ. There are many positive things associated with higher intelligence, including the fact that more intelligent people are more likely to get better grades and advance farther in their schooling. It may not be surprising to you that more intelligent people also tend to earn more money professionally. Indeed, according to research by Beaton (1975) ten IQ points are worth about four thousand dollars in annual salary. Children's intelligence is a good predictor of their future economic success according to Herrnstein and Murray (1994). Of course, this is partly because, as documented in research by Lord, DeVader, & Alliger (1986) intelligent people are perceived to have greater leadership potential and are given greater professional opportunities. But what may be surprising to you is that intelligent people also tend to have significantly better health and longer life expectancies (see research by Gottfredson & Deary, 2004).

Participants in the medium motivation condition read:

> In this experiment, you will complete a short version of an IQ test, which is known to be a good indicator of one's intelligence.

Those in the low motivation condition read:

> You will complete a series of questions we are testing to see whether or not they can be used as a quiz in another study.

Manipulation check. In order to assess their motivation to perform well, we asked participants, prior to the test, to rate: how motivated they were, how important it was for them to perform well, and how hard they expected to work. Participants responded on a scale ranging from 1 = *Not at all* to 5 = *Extremely*.

Overconfidence measures. We elicited BTA (vague) measures by asking participants to assess their own and others' performance on the questions, both before and after taking it, on a scale ranging from 1 = *Very badly* to 5 = *Very well*. Overplacement (specific) measures included participants' estimates of their own and others' scores on the task from 0 to 10. We counterbalanced the order in which participants responded to the BTA and overplacement measures**.**

Results

Manipulation checks. We excluded no participants from the analyses. We averaged the three manipulation check items together to form one measure of motivation (*alpha* = .83) and submitted it to a 3-cell (importance: low vs. medium vs. high) ANOVA. There is an effect of importance on participants' ratings of their motivation to succeed on the task, $F(2,108) = 5.46$, $p = .006$, *partial* $\eta^2 = .09$, with mean group ratings corresponding to the level of motivation (high: $M = 3.98$, $SD = 0.68$; medium: $M = 3.50$, $SD = 0.89$; low: $M = 3.42$, $SD = 0.89$).

Effect of Motivation on BTA Beliefs and Overplacement

We submitted participants' BTA beliefs to a 3-cell (motivation: *low* vs. *medium* vs. *high*) between-subjects ANOVA. Motivation neither affected BTA beliefs individually before, $F(2, 108) = .17$, $p = .846$, nor after the task, $F(2, 108) = .20$, $p = .817$. Results hold when ratings of self and others were analyzed as a repeated measure (interaction of self-other and motivation before: p = .846; after: $p = 817$). The objectivity and verifiability of performance assessment appears to have suppressed the effect of motivation on BTA beliefs altogether.

We submitted the overplacement measure to the same 3-cell ANOVA. Again, the motivation manipulation did not affect overplacement before, $F(2, 108) = .05$, $p = .950$, or after

the task, $F(2, 107) = .16$, $p = .853$. Results hold when we analyzed estimated and actual differences as a repeated measure.

Note that in aggregate, participants did not display BTA beliefs. Before the task, participants predicted that they would perform no better than others, paired $t(110) = 0.70$, $p = .49$, and afterwards believed that they had performed worse ($M = 3.05$, $SD = 1.00$) than others ($M = 3.32$, $SD = 0.75$), paired $t(110) = -2.77$, $p = .007$, $d = -0.31$. There was only minimal evidence of overplacement before the task and no evidence of it after. People predicted that they would perform better than others ($M = 0.44$, $SD = 1.58$) more so than they actually did ($M = 0.00$, $SD = 1.84$) before the task, paired $t(110) = 2.26$, $p = .03$, $d = 0.26$, but not after (reported: $M = -0.19$, $SD = 2.07$), paired $t(109) = -0.98$, $p = .35$.

Effects of Motivation on Actual Performance

We submitted participants' actual performance to a 3-cell (motivation: low vs. medium vs. high) ANOVA. The results revealed no effect of motivation, $F(2,108) = .51$, $p = .603$, *partial* $\eta^2 = .009$.

Discussion

Motivated overconfidence appears to depend heavily on subjectivity and vagueness. Experiments 1-4 found that motivation affected overconfidence on vaguely defined traits and vague measures. Experiment 5 tested a domain where we could measure performance more objectively and found, accordingly, no effect of motivation on overconfidence.

In fact, we found little evidence of overconfidence whatsoever, regardless of motivation. Even on vague BTA measures, before the task, people expected to perform no differently than others and believed they had performed *worse* afterwards. Overplacement measures likewise

revealed modesty. On average, people only slightly overplaced their scores relative to others prior to the task and did not overplace at all afterwards.

These results are consistent with cognitive theories of overconfidence (Moore & Healy, 2008), which have gained considerable empirical support. This account predicts that different forms of overconfidence, either in absolute self-evaluations or in relative self-judgments like the ones elicited in this experiment, can disappear, and even reverse, when the level of difficulty changes. Specifically, easy tasks produce overplacement but underplacement can result when difficulty is high. The fact that motivation had no effect on either BTA or overplacement in Experiment 5, given the initially low levels of overplacement and relative high difficulty of the task, provide further support for the cognitive explanation. Furthermore, the result showing that motivation did not affect performance itself is consistent with past work showing that people's mindset (their optimism) does not affect performance as much as people think it will (Tenney, Logg, & Moore, 2015).

Experiment 6: Idiosyncratic Construals

Why are overly favorable self-perceptions inflated when assessments are vague? One possibility is that people take advantage of vague standards to engage in self-enhancement. Alternatively, people may differ in how they interpret the meaning of the trait they assess. In the absence of specifically-defined criteria for possessing the trait, they might construct their own criteria in a way that emphasizes their relative strengths. Based on these criteria, people's assessments of themselves would naturally skew positive.

It is possible that people's criteria weightings are driven by self-serving motives to sustain flattering beliefs about the self (Brownstein, 2003). However, it is easiest to maintain the illusion when the self-deception is subtle enough to provide plausible deniability (Kunda, 1990;

Shepperd, Ouellette, & Fernandez, 1996). Becoming aware of the artifice undermines the value and credibility of the self-enhancement. Experiment 6 employed a within-subjects research design which allowed us to examine people's awareness of applying idiosyncratic criteria. If individuals are aware of the idiosyncrasies in their own self-assessments, then overconfidence does not result from self-deception.

Experiment 6 examines the relationship between the BTA effect and the specificity of a trait—honesty. We chose honesty because, unlike introversion, it is more universally considered by people as an important, desirable trait; therefore we could expect all participants to be highly motivated to possess it. We tested whether people rely on idiosyncratic criteria for honesty or whether they agree about what honesty means but indulge in rosy self-perceptions. Participants assessed their own and others' honesty before and after specifying what honesty meant to them; they rated the relevance of different dimensions of honesty to their definitions of the trait.[8]

While a vague attribute may allow for self-serving definitions and flattering self-perceptions, specifying the dimensions of the attribute should reduce idiosyncratic construals. Still, honesty is a complex trait and we measured whether people weighted the specific criteria differently from each other when given the opportunity to construct their own conception of honesty. Therefore, we expected that defining honesty through specific behaviors (in phase 2), rather than as a vague trait (in phase 1), would attenuate BTA beliefs, we predicted that stronger BTA beliefs reemerge when participants can independently adjust their criterion weights for the honesty-related behaviors (in phase 3).

[8] A pre-test identified the ten behaviors most strongly associated with honesty. We surveyed eighty-seven students on the campus of a West Coast University and thanked them with candy. Participants read thirty-three behaviors and rated the five which were most relevant to honesty on a scale from 1 = captures my idea of honesty the best to 5. Ten of these behaviors were rated within the top five for more than 50% of participants and thus comprised the list of behaviors we employed in the experiment.

Method

Participants

One hundred and forty-one undergraduate students at a West Coast university completed one fifteen-minute session for either credit or pay. We recruited as many participants as the end of the semester allowed prior to analyzing the data. We needed to drop data from five participants whose ratings included no variance, leaving a sample size of 136.

Design

The experiment had a 2 (target: self vs. others) × 3 (assessment type: vague vs. specific vs. relevance) within-subject design. The assessment type manipulation involved assessing BTA beliefs in three different formats: (1) their own and others' honesty, without clarifying what honesty meant; (2) the frequency at which they display eleven specific honesty-related behaviors; and (3) the relevance of each of these behaviors to their own definition of the trait. We measured BTA beliefs in each of the three phases and then compared them with each other. To determine whether people defined honesty in a self-serving manner, we measured the correlation between (1) how frequently people rated enacting each behavior in the second phase with (2) how relevant that behavior was to their definition of honesty in the third phase.

Procedure and Materials

Vaguely-presented traits. In phase 1, participants rated how well each of the following ten traits described them and how well they described the average participant in the study: honest, kind, responsible, intelligent, competent, secure, conscientious, agreeable, imaginative, and outgoing. They rated each trait on a scale from 1 = *Does not describe me at all* to 9 = *Describes me very well*. We assessed indirect BTA beliefs for each phase by comparing self and other ratings.

Specific behaviors. In phase 2, participants read eleven statements pertaining to behaviors considered as honest, e.g., "When I make a promise, I keep it," "I do not say things I know to be untrue." The full list of statements is in Appendix D. For each statement, they rated how often it accurately describes them, on a scale ranging from 0% = *I never do this* to 100% = *I always do this*. Next, participants estimated their own and others' overall honesty, as defined by the specific behaviors on the list. This judgment explicitly encouraged participants to treat the behaviors as equally important by asking them to average across the eleven items on the list.

Relevance of behaviors to honesty. In the third and final phase of the experiment, we explicitly re-introduced the opportunity to define honesty in a more personal way. Participants reported how relevant each of the same specific behaviors were to their interpretation of honesty on a scale from 0 = *Not at all important* to 100 = *Most important*. After rating each of the behaviors in terms of their relevance to honesty, participants used the weights to assess their own and others' honesty on a scale from 0% to 100%. They rated both other students at the school in general and the average participant in the study.

Results

BTA Beliefs

We re-scaled the vague ratings of honesty from a 9-point to a 0 to 100 scale in order to compare them with the other ratings from the other phases. We submitted all ratings to a 2 (target: self vs. others) × 3 (assessment type: vague vs. specific vs. relevance) repeated measures ANOVA. There are main effects of target, $F(1, 135) = 72.06$, $p < .001$, *partial* $\eta^2 = .35$, and assessment type, $F(2, 134) = 10.47$, $p < .001$, *partial* $\eta^2 = .14$, and, importantly, a significant interaction between the two factors, $F(2, 134) = 10.11$, $p < .001$ *partial* $\eta^2 = .13$.

Consistent with the main effect of target, participants displayed BTA beliefs, rating themselves as more honest than others in each assessment phase, *p*s < .001. These BTA beliefs weakened when participants considered specific behaviors than when they assessed honesty as a vague trait. As Figure 5 shows, participants rated themselves as more honest than others, even when making assessments on a specific scale, (*M*difference = 3.99, *SD* = 12.65), *t*(135) = 3.68, *p* < .001, *d* = 0.63, but this effect was weaker than the one observed in their vague ratings. When participants applied their own idiosyncratic weights to the various honest behaviors, the BTA effect strengthened again (*M* = 7.49, *SD* = 12.02), *t*(135) = 7.27, *p* < .001, *d* = 1.25.[9]

Idiosyncratic Definitions of Honesty

When allowed to assess their honesty based on their own definitions, participants' BTA beliefs became stronger relative to the specific assessments and more similar to the initial vague assessments. If each person considered the 11 specific behaviors related to honesty in a different, distinct way, then, according to their own definitions, each person could correctly believe they were more honest than others.

For each behavior, we computed a correlation between how frequently people claimed to display it and how relevant they thought it was. The frequency and relevance ratings correlated positively for every behavior, *r*s > .24, *p*s < .01, see Table 4. We are cautious to conclude from this correlational result that people weighted the relevance of behaviors in a self-serving manner; we cannot rule out the possibility that the more relevant people thought behaviors were to honesty, the more frequently they displayed them, and that people were aware that others had their own idiosyncratic construals.

[9] Specificity affected ratings of others, such that in phase three, participants considered the average experiment participant more honest (*M* = 74.90, *SD* = 13.87) than their fellow students in general (*M* = 70.99, *SD* = 14.00), *t*(135) = 4.87, *p* < .001, *d* = 0.28. This kinder assessment of the average participant implies a conservative test of BTA beliefs in phases 1 and 2.

We examined whether these seemingly self-serving definitions corresponded with self-perceptions of honesty. We multiplied frequency ratings by relevance ratings for each behavior, summed the product across behaviors, and measured the correlation of the product with participants' final self-assessments vis-à-vis the different behaviors. It appears that as definitions became more flattering, so did self-assessments, $r = .68$, $p < .001$. We also tested how similar participants' definitions of honesty were to each other and measured the correlation of each participant's relevance ratings with every other participants' ratings. The average of these correlations was low, $r = .09$, which suggests that people did not converge on one definition of honesty.

Discussion

The results of Experiment 6 suggest that a vague definition of a trait allows people to produce more positive self-evaluations by relying on idiosyncratic criteria for what it means to possess the trait. Utilizing specific criteria, people employed idiosyncratic definitions of honesty. Focusing on honesty allowed us to use a domain where people were highly motivated to possess the focal trait. BTA beliefs were strong both when assessments were based on a vague scale and on personal definitions of the trait. This result suggests that when the desirable trait was originally presented in vague terms, people may have used idiosyncratic interpretations of the trait to assess themselves.

Second, our results suggest that people appear capable of moderating their own BTA beliefs when the domain is clarified. This result implies that specifying definitions can help people reduce BTA biases. People exhibited less extreme BTA beliefs in light of specific definitions, implying that they knew that others might not share their definitions of honesty (see Roy and Liersch, 2014).

Finding an effect of vagueness *within*-subject, and *within*-trait is important because it shows that people are aware of, and do not try to hide, the degree to which self-enhancing beliefs emerge in the presence of vagueness. The idiosyncratic trait definitions that drive this effect may not be motivated self-delusions. They are conscious and may even be rationally justifiable. The clear implication is that beliefs that appear self-serving are not driven by an unrealistic self-aggrandizing, but instead by self-consciously idiosyncratic standards of assessment.

General Discussion

Is overconfidence motivated? Our results suggest that motivation affects overconfidence less than the prior literature might suggest. It is most certainly not the case that the desire to possess a trait or ability always leads people to self-enhance. For example, the desire to see oneself as intelligent did not lead our participants to delude themselves into believing they had aced an IQ test. When motivation increases self-enhancement, its effect is strongest for ambiguous traits assessed using vague BTA measures. The striking limitation of these vague measures is that they lack an objective accuracy standard. Getting specific reduces the effect of motivation, and so overconfidence appears less pervasive than the prior literature implies. Our results help identify both when motivation contributes to self-enhancing beliefs and how people construct these beliefs.

Our results build on prior research that has examined either the vagueness of measurement (Epley & Dunning, 2000; Preuss & Alicke, 2009) or the ambiguity of the trait (Alicke et al., 1995; Dunning et al., 1989). We replicate the main effects of motivation and specificity but more importantly, show that that specificity interacts with motivation to affect overconfidence.

Table 5 summarizes our results. In Experiments 1 and 2, people displayed stronger BTA beliefs when they were motivated to view themselves as introverted, but only when the definition of what it means to be introverted was specific; the effect weakened when assessments were made on specific measures that captured overplacement. When we elicited judgments in an unfamiliar, vague domain (in Experiment 3), overconfidence emerged again, and this time was not limited to vague measures. Experiment 4 compared the effects of motivational and cognitive factors on overconfidence within the same unfamiliar, vague domain; both affected vague (BTA) and specific measures (overplacement). When the domain itself was objective and verifiable, as was test performance in Experiment 5, motivation's effect on overconfidence again disappeared. In fact, the unambiguous domain appears to have suppressed overly flattering self-assessments altogether.

Experiment 6 provided new insight into the psychological mechanisms behind the construction of subjective self-perceptions. Although people were capable of decreasing their BTA biases when criteria were made specific, idiosyncratic definitions also contributed to BTA beliefs. We cannot distinguish the degree to which these idiosyncratic trait definitions are the result of self-serving definitions or whether people simply work to enact those honesty-relevant behaviors they regard as most important. However, our results suggest that a reduction in biased beliefs about one's introversion were due to clarifying not only the trait's measurement, but also what it means to be introverted.

Theoretical Implications

This paper contributes to the research on motivation, social comparison and self-perception by providing an empirical examination of a widespread assumption in the field: that the motivation to possess a certain quality drives the degree to which people are biased in their

assessment relative to others (Brown, 2011; Dunning, 1995; Dunning et al., 1989; Alicke, 1985; Mazar, Amir, & Ariely, 2008). Our results suggest that motivation affects overconfidence mostly in subjective, vaguely-framed contexts. These results help us better understand past correlational work on the relationship. Another contribution of this work is the measurement of overplacement, which compares people's beliefs about themselves relative to others with their actual relative standings. Measuring overplacement allowed us to increase the resolution of our tests, and capture individual bias rather than bias at the level of the sample.

Examining the relationship between motivation and overconfidence also contributes to work on self-enhancement. The term self-enhancement is regularly used to describe flattering self-perceptions regardless of their accuracy. Some work has offered a motivational explanation for self-enhancement, similar to the proposed effects on overconfidence (Dunning, 2005; Gosling, John, Craik, & Robbins, 1998; Greenwald, 1980; Kunda, 1990; Sedikedes & Gregg, 2008). We expand these findings by directly measuring the extent to which people's self-ratings are consistent with reality, as well as how they are affected by motivation.

We should note that a motivational account of overconfidence differs from how researchers have measured wishful thinking. Wishful thinking has often been studied by manipulating desirability (motivation) and measuring the perceived likelihood of future events (e.g., Lench & Ditto, 2008; Marks, 1951; Windschitl, Scherer, Smith, & Rose, 2013; Windschitl, Smith, Rose, & Krizan, 2010). Neither self-enhancement nor wishful thinking require the benchmark of accuracy, in contrast to overconfidence. Furthermore, experimental evidence suggests that motivation does a poor job explaining empirical evidence of wishful thinking (Bar-Hillel & Budescu, 1995; Bar-Hillel, Budescu, & Amar, 2008; Krizan & Windschitl, 2007; Vosgerau, 2010).

Practical Implications

Are overconfident beliefs self-serving? For them to qualify as such, holding overconfident beliefs would have to benefit the individual holding them. However, it is easy to identify risks of overconfidence. Overconfidence, after all, can impair both performance and well-being. Overconfident people risk too much (Camerer & Lovallo, 1999; Odean, 1998). And while we may experience pleasure in savoring a bright future (Loewenstein & Prelec, 1993), those who are most confident in their performance, and who therefore believe they need not try hard, can actually perform worse (Cain, Moore, & Haran, 2015; Stone, 1994; Vancouver, Thompson, Tischner, & Putka, 2002). For instance, the student who is overconfident about his performance and thus does not believe he needs to study is unlikely to outperform his peers.

Overconfidence in one's abilities invites disappointment when performance turns out worse than expected (McGraw, Mellers, & Ritov, 2004; van Dijk, Zeelenberg, & van der Pligt, 2003). People seem aware of the disappointment that follows overconfidence when they display defensive pessimism. In fact, people who lower their expectations through defensive pessimism enjoy their success as much as optimists but are not as distraught by failure (Norem & Cantor, 1986). If self-flattering beliefs are self-interested, then people should display overconfidence in all of the domains they value. Yet, people often display underconfidence in domains they think are important (Blanton, Axsom, McClive, & Price, 2001; Kruger, 1999; Moore, 2007; Windschitl, Kruger, Simms, 2003), including social status, respect, and influence (Anderson, Srivastava, Beer, Spataro, & Chatman, 2006); for more evidence, just talk to any graduate student.

Our results suggest that one should not always expect greater motivation to beget greater overconfidence. When performance standards are quantitative and objective, our results imply

that motivational effects on overconfidence are limited. Some domains, then, are more suitable for presenting such conditions than are others. Within the workplace, success often depends on numbers. A company must turn a baseline profit to continue functioning, which often depends on the number of clients secured or products sold. Athletic champions are determined by the number on the scoreboard and records for speed are based on the clock. However, even if motivation only affects overconfidence in purely subjective contexts, those contexts are not entirely uncommon. Obviously, objective criteria are not always readily available for some consequential outcomes. Mates are rarely chosen based on objective, verifiable or measureable criteria. Assessment of academic papers depends on subjective assessments made by readers and reviewers. Employee evaluations are, to a great extent, driven by the subject assessment of the manager. Under these circumstances, we expect wider latitude for subjective construal of performance and stronger effects of motivation on overconfident beliefs.

High levels of overconfidence become more likely when one's goals are not specifically defined, which holds important implications for individuals, managers, and organizations for whom overconfidence may contribute to unmet expectations. Yet, our results are hopeful in that they suggest a path to more accurate self-assessments. Even within ambiguous domains, providing clearly defined criteria for what makes a productive employee, an effective leader, and an efficient team, may help people better calibrate their self-perceptions with reality.

Conclusion

We have sought to test a widely held belief that overconfident beliefs are driven, in part, by the motivation to view oneself positively. We directly manipulated motivation and measured overconfidence, examining what inflates and deflates positive self-perceptions. We found limited evidence for motivational influence on overconfidence. The most important implication

of our findings is the insight into when motivation has an effect and when it does not. Objective, verifiable domains appeared to suppress overconfidence, even on a vague measure. Within a subjective performance domain, motivation had influence when assessments were made on vague but less so on specific measures, and mostly when the trait was vague and less so when it was clearly defined. Indeed, whenever clarity of criteria, measures, and domains allowed for us to compare self-reports with verifiable truth, we found little evidence of a motivational influence on overconfidence.

Origin Story

This paper owes its genesis to questions about the durability and prevalence of overconfidence. Psychologists were routinely taken aback by evidence showing how common it is for people to be underconfident (Moore & Small, 2007). For instance, on difficult trivia quizzes, the majority of people believe that they are worse than others (Moore & Healy, 2008). When presenting this work, we routinely encountered the objection that studying such trivial tasks neglects the powerful role of motivation in driving people's beliefs about consequential performance domains in everyday life. We began this research project with the goal of identifying the role of motivation. Although numerous papers claimed that self-enhancement motivations drove people to believe they were better than others, the evidence for this claim was largely correlational and lacked clean experimental tests. We set out to provide such a test.

When our early results failed to find an effect of our manipulations of motivation on any form of overconfidence, we were stunned. These results made us more skeptical that motivation played the powerful and pervasive role so many had assumed it did. The story of this research project is the story of our search to find a context—any context—in which we could identify an effect of motivation on overconfidence. After a set of results failing to find any effect of motivation on overconfidence, we finally were able to identify when it mattered—when both the

performance domain and its method of assessment were sufficiently vague to allow individuals to apply idiosyncratic construals of performance.

References

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology*, *49*(6), 1621-1630. Doi:10.1037/0022-3514.49.6.1621

Alicke, M. D., & Govorun, O. (2005). The better-than-average effect. In M. Alicke, D. Dunning, & J. Krueger, *The self in social judgement* (pp. 85-106) Hove, UK: Psychology Press.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, *68*(5), 804–825.

Anderson, C., Srivastava, S., Beer, J. S., Spataro, S. E., & Chatman, J. A. (2006). Knowing your place: Self-perceptions of status in face-to-face groups. *Journal of Personality and Social Psychology*, *91*(6), 1094. Doi:10.1037/0022-3514.91.6.1094

Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2005). Large stakes and big mistakes. *Review of Economic Studies, 76*(2), 451–469.

Armor, D. A., & Sackett, A. M. (2006). Accuracy, error, and bias in predictions for real versus hypothetical events. *Journal of Personality and Social Psychology*, *91*(4), 583–600.

Bar-Hillel, M., & Budescu, D. (1995). The elusive wishful thinking effect. *Thinking & Reasoning, 1*(1), 71-103.

Bar-Hillel, M., Budescu, D. V., & Amar, M. (2008). Predicting world cup results: do goals seem more likely when they pay off?. *Psychonomic Bulletin Review*, *15*(2), 278-283. Doi: 10.3758/PBR.15.2.278

Barber, B., & Odean, T. (1999). Do investors trade too much?. *American Economic Review*, 89(5), 262.

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly journal of Economics*, 261-292.

Benoît, J. P., Dubra, J., & Moore, D. A. (2015). Does the better-than-average effect show that people are overconfident?: two experiments. *Journal of the European Economic Association*. *13*(2), 293–329. Doi: 10.1111/jeea.12116

Blanton, H., Axsom, D., McClive, K. P., & Price, S. (2001). Pessimistic bias in comparative evaluations: A case of perceived vulnerability to the effects of negative life events. *Personality and Social Psychology Bulletin*, *27*(12), 1627–1636. Doi: 10.1177/01461672012712006

Brown, J. D. (2011). Understanding the better than average effect: motives (still) matter. *Personality and Social Psychology Bulletin*, *38*(2), 209–219. Doi:10.1177/0146167211432763

Brownstein, A. L. (2003). Biased predecision processing. *Psychological Bulletin*, *129*(4), 545.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons. *Journal of personality and social psychology*, *90*(1), 60.

Cain, D. M., Moore, D. A., & Haran, U. (2015). Making sense of overconfidence in market entry. *Strategic Management Journal, 36*, 1-18. Doi:10.1002/smj.2196

Cain, S. (2013). *Quiet: The power of introverts in a world that can't stop talking*. New York: Random House.

Camerer, C. F., & Lovallo, D. (1999). Overconfidence and excess entry: an experimental approach. *American Economic Review*, *89*(1), 306-318.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, *130*(5), 813-838. Doi:10.1037/0033-2909.130.5.813

Chambers, J., Windschitl, P., & Suls, J. (2003). Egocentrism, Event Frequency, and Comparative Optimism: When what Happens Frequently is "More Likely to Happen to Me". Personality and Social Psychology Bulletin, 1343-1356.

Chamorro-Premuzic, T. (2013). *Confidence: Overcoming Low Self-Esteem, Insecurity, and Self-Doubt*. London: Penguin.

Cooper, A. C., Woo, C. Y., & Dunkelberg, W. C. (1988). Entrepreneurs' perceived chances for success. *Journal of Business Venturing*, *3*(2), 97-108. Doi:10.1016/0883-9026(88)90020-1

Deci, E. L., & Ryan, R. M.(1985). Cognitive evaluation theory: Perceived causality and perceived competence. In Deci, E. L., & Ryan, R. M. (Eds.) *Intrinsic Motivation and Self-Determination in Human Behavior.* New York: Plenum. 43-85.

Dunning, D. (1995). Trait importance and modifiability as factors influencing self-assessment and self-enhancement motives. *Personality and Social Psychology Bulletin*, *21*(12), 1297-1306. Doi: 10.1177/01461672952112007

Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry*, *10*(1), 1-11.

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York, NY: Psychology Press.

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106. Doi:10.1111/j.1529-1006.2004.00018.

Dunning, D., Leuenberger, A., & Sherman, D. A. (1995). A new look at motivated inference: are self-serving theories of success a product of motivational forces?. *Journal of Personality and Social Psychology*, *69*(1), 58. Doi:10.1037/0022-3514.69.1.58

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: the role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology*, *57*(6), 1082 -1090. Doi:10.1037/0022-3514.57.6.1082

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*(6), 861–875.

Fabricius, G., & Büttgen, M. (2013). The influence of knowledge on overconfidence: consequences for management and project planning. *International Journal of Business and Management*, *8*(11), 1-12. Doi:10.5539/ijbm.v8n11p1

Gilovich, T., Kerr, M., & Medvec, V. H. (1993). Effect of temporal perspective on subjective confidence. *Journal of Personality and Social Psychology*, *64*(4), 552–560.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of personality and social psychology*, *74*(5), 1337 -1349. Doi:10.1037/0022-3514.74.5.1337

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*(1), 1-4.

Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased: motivated reasoning and rationality. *Psychology of Learning and Motivation*, 41.

Harris, A. J. L., & Hahn, U. (2011). Unrealistic optimism about future life events: a cautionary note. *Psychological Review*, *118*(1), 135–154. Doi:10.1037/a0020997

Herrnstein, R. J., & Murray, C. (1994). The bell curve: The reshaping of American life by differences in intelligence. *New York: Free*.

Huff, C., & Tingley, D. (2015). "Who are these people?" Evaluating the demographic characteristics and political preferences of MTurk survey respondents. *Research & Politics*, 2(3). DOI: 10.1177/2053168015604648

Kahneman, D. (2011b). Thinking fast and slow. New York: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, 103(3), 582-591.

Klein, W. M., & Kunda, Z. (1992). Motivated person perception: constructing justifications for desired beliefs. *Journal of Experimental Social Psychology*, *28*(2), 145-168. Doi:10.1016/0022-1031(92)90036-J

Krizan, Z., & Windschitl, P. D. (2007). The influence of outcome desirability on optimism. *Psychological Bulletin*, *133*(1), 95. Doi:10.1037/0033-2909.133.1.95

Krueger, J. I. (1998). Enhancement bias in descriptions of self and others. *Personality and Social Psychology Bulletin, 24*(5), 505–516.

Krueger, J. I., & Wright, J. C. (2011). Measurement of self-enhancement (and self-protection). In M. D. Alicke & C. Sedikides (Eds.), *Handbook of Self-Enhancement and Self-Protection* (pp. 472–494). New York: Guilford.

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*(2), 221-232. Doi:10.1037/0022-3514.77.2.22

Kruger, J., & Burrus, J. (2004). Egocentrism and focalism in unrealistic optimism (and pessimism). *Journal of Experimental Social Psychology, 40*(3), 332–340. Doi:10.1016/j.jesp.2003.06.002

Kruger, J., & Savitsky, K. (2009). On the genesis of inflated (and deflated) judgments of responsibility: Egocentrism revisited. *Organizational Behavior & Human Decision Processes, 108*(1), 972–989. Doi:10.1016/j.obhdp.2008.06.002

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498. Doi:10.1037/0033-2909.108.3.4800

Kunda, Z., & Sanitioso, R. (1989). Motivated changes in the self-concept. *Journal of Experimental Social Psychology*, *25*(3), 272-285. Doi:10.1016/0022-1031(89)90023-1

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*(4), 636-647. Doi:10.1037/0022-3514.53.4.636

Larrick, R. P., Burson, K. A., & Soll, J. B. (2007). Social comparison and confidence: When thinking you're better than average predicts overconfidence (and when it does not). *Organizational Behavior and Human Decision Processes*, *102*(1), 76-94. Doi:10.1016/j.obhdp.2006.10.002

Larwood, L., & Whittaker, W. (1977). Managerial myopia: self-serving biases in organizational planning. *Journal of Applied Psychology*, *62*(2), 194-198.

Lench, H. C., & Ditto, P. H. (2008). Automatic optimism: biased use of base rate information for positive and negative events. *Journal of Experimental Social Psychology*, *44*(3), 631-639. Doi:10.1016/j.jesp.2007.02.011

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological Review*, *100*(1), 91-108. Doi:10.1037/0033-295X.100.1.91

Marks, R. W. (1951). The effect of probability, desirability, and "privilege" on the stated expectations of children*. *Journal of Personality*, *19*(3), 332-351. Doi:10.1111/j.1467-6494.1951.tb01107.x

Markman, A. B., Maddox, W. T., & Worthy, D. A. (2006). Choking and excelling under pressure. *Psychological Science*, *17*(11), 944–948.

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: a theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633-644. Doi:10.1509/jmkr.45.6.63

McCroskey, J. C. (1997). Self-report measurement. In J. Daly, J. McCroskey, J. Ayres, T. Hopf, & D. Ayres (Eds.), *Avoiding communication: Shyness, reticence, and communication apprehension* (pp. 191-216). Cresskill, NJ: Hampton Press.

McGraw, A. P., Mellers, B. A., & Ritov, I. (2004). The affective costs of overconfidence. *Journal of Behavioral Decision Making*, *17*(4), 281–295. Doi:10.1002/bdm.472

Middleton, W., Harris, P., & Surman, M. (1996). Give'em enough rope: Perception of health and safety risks in bungee jumpers. *Journal of Social and Clinical Psychology*, *15*(1), 68-79. Doi:10.1521/jscp.1996.15.1.68

Moore, D. A. (2007). Not so above average after all: when people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*, *102*(1), 42-58. Doi:10.1016/j.obhdp.2006.09.005

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. Doi:10.1037/0033-295X.115.2.502

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative social judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92*(6), 972–989. Doi: 10.1037/0022-3514.92.6.972

Moore, D. A., Tenney, E. R., & Haran, U. (2016). Overprecision in judgment. In G. Wu & G. Keren (Eds). *The Wiley Blackwell Handbook of Judgment and Decision* Making, Vol. 1, 182-209. Wiley-Blackwell.

Norem, J. K., & Cantor, N. (1986). Defensive pessimism: harnessing anxiety as motivation. *Journal of Personality and Social Psychology, 51*(6), 1208-1217. Doi:10.1037/0022-3514.51.6.1208

O'Brien, E. (2013). Easy to retrieve but hard to believe: metacognitive discounting of the unpleasantly possible. *Psychological Science*, *24*(6), 844–851. Doi:10.1177/0956797612461359

Odean, T. (1998) Are investors reluctant to realize their losses? *The Journal of Finance*, *53*(5), 1775-1798. Retrieved from http://www.jstor.org/stable/117424

Otten, W., & Van Der Pligt, J. (1996). Context effects in the measurement of comparative optimism in probability judgments. *Journal of Social and Clinical Psychology*, *15*(1), 80-101. Doi: 10.1521/jscp.1996.15.1.80

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, 23(3), 184-188.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5, 411–419

Preuss, G. S., & Alicke, M. D. (2009). Everybody loves me: Self-evaluations and metaperceptions of dating popularity. *Personality and Social Psychology Bulletin*, *35*(7), 937–950.

Radhakrishnan, P., Arrow, H., & Sniezek, J. A. (1996). Hoping, performing, learning, and predicting: changes in the accuracy of self-evaluations of performance. *Human Performance*, *9*(1), 23-49. Doi**:**10.1207/s15327043hup0901_2

Roy, M. M., & Liersch, M. J. (2014). I am a better driver than you think: examining self-enhancement for driving ability. *Journal of Applied Social Psychology, 43*(8), 1648–1659.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68–78.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-assessment. *Journal of Personality and Social Psychology*, *84*(1), 60-79. Doi:10.1037/0022-3514.84.1.60

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science*, *3*(2), 102-116. Doi:10.1111/j.1745-6916.2008.00068.x

Shepperd, J. A., Klein, W. M., Waters, E. A., & Weinstein, N. D. (2013). Taking stock of unrealistic optimism. *Perspectives on Psychological Science*, *8*(4), 395-411. Doi: 10.1177/1745691613485247

Shepperd, J. A., Ouellette, J. A., & Fernandez, J. K. (1996). Abandoning unrealistic optimism: Performance estimates and the temporal proximity of self-relevant feedback. *Journal of Personality and Social Psychology*, *70*(4), 844–855.

Simon, M., Houghton, S. M., & Aquino, K. (2000). Cognitive biases, risk perception, and venture formation: *How individuals decide to start companies. Journal of business venturing*, *15*(2), 113-134.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366. Doi: 10.1177/0956797611417632

Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self-deception requires specificity. *Cognition*, *115*(2), 268–281. Doi:10.1016/j.cognition.2009.12.017

Stone, D. (1994). Overconfidence in initial self-efficacy judgements: effect on decision processes and performance. *Organizational Behavior and Human Decision Processes*, *59*(3), 452-474. Doi:10.1006/obhd.1994.1069

Svenson, O. (1981). Are we all less risky and more skillful fellow drivers?. *Acta Psychologica, 47*(2), 143–148. Doi:10.1016/0001-6918(81)90005-6

Taylor, S. E., & Brown, J. D. (1994). Positive illusions and well-being revisited: Separating fact from fiction. *Psychological Bulletin, 116*(1), 21-27. Doi:10.1037/0033-2909.116.1.21

Tenney, E. R., Logg, J. M., & Moore, D. A. (2015). (Too) optimistic about optimism: The belief that optimism improves performance. *Journal of personality and social psychology*, 108(3), 377. Doi: 10.1037/pspa0000018

Tversky, A, & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–31. Doi:10.1126/science.185.4157.1124

Vancouver, J. B., Thompson, C. M., Tischner, E. C., & Putka, D. J. (2002). Two Experiments examining the negative effect of self-efficacy on performance. *Journal of Applied Psychology*, *87*(3), 506-516. Doi:10.1037//0021-9010.873.3.506

Van den Steen, E. (2004). Rational overoptimism (and other biases). *American Economic Review*, *94*(4), 1141–1151. Doi:10.1257/0002828042002697

van Dijk, W.W., Zeelenberg, M., & Van Der Pligt, J. (2003). Blessed are they who expect nothing: lowering expectations as a way of avoiding disappointment. *Journal of Economic Psychology*, *24*(4), 505–516. Doi:10.1016/S0167-4870(02)00211-8

von Hippel, W., & Trivers, R. L. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, *34*, 1–56.

Vosgerau, J. (2010). How prevalent is wishful thinking? Misattribution of arousal causes optimism and pessimism in subjective probabilities. *Journal of Experimental Psychology: General, 139*(1), 32-48. Doi:10.1037/a0018144

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*(5), 806–820. Retrieved from http://search.proquest.com/docview/1295906909?accountid=14496

Weinstein, N. D. & Lachendro, E. (1982). Egocentrism as a source of unrealistic optimism. *Personality and Social Psychology Bulletin*, *8*(2), 195–200. Doi:10.1177/0146167282082002

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The influence of egocentrism and focalism on people's optimism in competitions: when what affects us equally affects me more. *Journal of Personality and Social Psychology*, *85*(3), 389-408. Doi:10.1037/0022-3514.85.3.389

Windschitl, P. D., Scherer, A. M., Smith, A. R., & Rose, J. P. (2013). Why so confident? The influence of outcome desirability on selective exposure and likelihood judgment. *Organizational Behavior and Human Decision Processes*, *120*(1), 73–86. Doi:10.1016/j.obhdp.2012.10.002

Windschitl, P. D., Smith, A. R., Rose, J. P., & Krizan, Z. (2010). The desirability bias in predictions: going optimistic without leaving realism. *Organizational Behavior and Human Decision Processes*, *111*(1), 33–47. Doi:10.1016/j.obhdp.2009.08.003

Zacharakis, A. L., & Shepherd, D. A. (2001). The nature of information and overconfidence on venture capitalists' decision making. *Journal of Business Venturing, 16*(4), 311–332.

Zenger, T. R. (1992). Why do employers only reward extreme performance? Examining the relationships among performance, pay, and turnover. *Administrative Science Quarterly*, *37*(2), 198-219. Retrieved from http://www.jstor.org/stable/2393221

Zuckerman, E. W., & Jost, J. T. (2001). What makes you think you're so popular? Self-evaluation maintenance and the subjective side of the "friendship paradox". *Social Psychology Quarterly*, *64*(3) 207-223. Retrieved from http://www.jstor.org/stable/3090112

Tables

Table 1

Subjectivity Systematically Varies Across Experiments

Experiment	Domain	Measure
1	Subjective (Introversion)	Vague (with vague & specific criteria)
2	Subjective (Introversion)	Vague & Specific (consensus definition)
3	Subjective (Social Responsiveness)	Vague & Specific
4	Subjective (Social Responsiveness)	Vague & Specific
5	Objective (Test)	Vague & Specific
6	Subjective (Honesty)	Vague & Specific (idiosyncratic definitions)

Table 2

Behaviors by Trait in Experiment 1

Trait	Behavior
Introverted	I do not express my happiness outwardly.
	I work alone when I can rather than with a group.
	I am comfortable with silence.
	I am quiet in large groups of people.
	I think about what I am going to say before I say it.
Agreeable	I cooperate in most situations.
	I get along well with others.
	I avoid arguments.
	I think about other people's issues.
	I allow people the chance to explain themselves.
Conscientious	I pay attention to details.
	I am careful when I make decisions.
	I create goals for myself.
	I plan ahead.
	I check my work.
Imaginative	I find inspiration easily.
	I have a lot of ideas to share.
	I find it easy to think of lots of different kinds of ideas for a project.
	I approach problems differently from most people.
	I am curious about alternate outcomes for everyday situations.
Outgoing	I seek out social situations.
	I try to connect and develop relationships with most people I meet.
	I put myself in situations where I am likely to meet new people.
	I introduce myself to people I don't know.
	I initiate conversations.
Honest	If I make a mistake, I own up to it.
	When I make a promise, I keep it.
	I do not say things I know to be untrue.
	I do not purposely deceive others.
	I fulfill my obligations and do what I say I will do.

Table 3

Frequency of Enacting Behavior in Experiment 6

Behavior	Frequency *M*
I do not cheat on my boyfriend/girlfriend. (5)	89.43
I fulfill my obligations and do what I say I will do. (4)	86.99
I do not steal. (8)	86.10
When I make a promise, I keep it. (1)	86.04
If I find something of value I do my best to return it to the owner. (10)	83.89
I live according to my own values. (7)	83.57
If I make a mistake, I own up to it. (6)	82.50
(Other) I am honest in ways that the above statements fail to capture. (11)	82.04
I do not pretend to be something I am not. (9)	76.04
I do not purposely deceive others. (3)	75.78
I do not say things I know to be untrue. (2)	74.59

Note: Behaviors are listed in order of the magnitude of the BTA effect. The number next to the trait is the order in which the behavior was presented to participants.

Table 4

Correlations between the Frequency and Relevance Weights in Experiment 6

	Frequency and Relevance Ratings	
Behaviors	*Correlation*	*p*
I do not steal. (8)	0.597	< .001
I do not cheat on my boyfriend/girlfriend. (5)	0.587	< .001
If I find something of value I do my best to return it to the owner. (10)	0.576	< .001
I do not say things I know to be untrue. (2)	0.546	< .001
(Other) I am honest in ways that the above statements fail to capture. (11)	0.517	< .001
When I make a promise, I keep it. (1)	0.455	< .001
I do not pretend to be something I am not. (9)	0.433	< .001
I do not purposely deceive others. (3)	0.417	< .001
I fulfill my obligations and do what I say I will do. (4)	0.393	< .001
If I make a mistake, I own up to it. (6)	0.348	< .001
I live according to my own values. (7)	0.242	< .01

Note: Behaviors are listed in order of the magnitude of their frequency and relevance rating correlation. Numbers in parentheses correspond with the order in which behaviors were presented.

Table 5

Summary of the effect of motivation and difficulty on overconfidence across experiments

Experiment	Results
1	Motivation affects BTA for introversion when criteria is *vague* but not when it is specific.
2	Motivation affects BTA for introversion *more than it affects overplacement*.
3	Motivation affects BTA and overplacement for social responsiveness.
4	Motivation *and difficulty* affect BTA and overplacement for social responsiveness.
5	Little evidence of BTA beliefs or overplacement for intelligence overall. Motivation neither affects BTA nor overplacement.
6	A vague trait on which people are motivated to see themselves possessing the trait, allows for idiosyncratic definitions of what it means to possess the trait, even using specific criteria, which produces more positive self-evaluations.

Figures

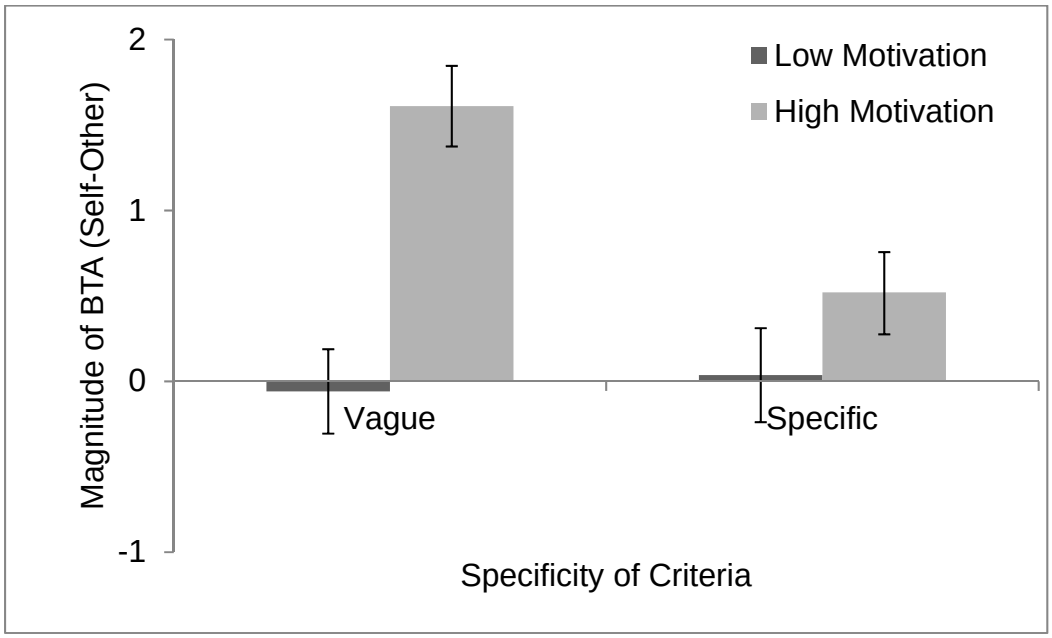

Figure 1. The magnitude of BTA beliefs as a function of experimental motivation (low: extroversion is important vs. high: introversion is important) and specificity conditions in Experiment 1. Error bars represent standard errors.

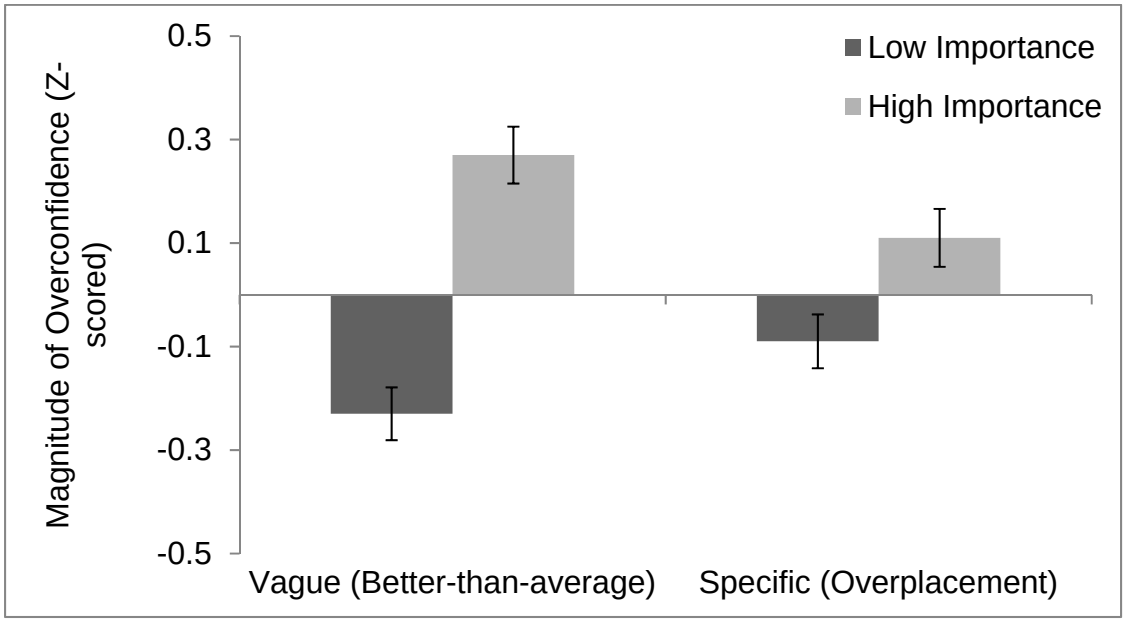

Figure 2. The magnitude of overconfidence for introversion as a function of experimental motivation condition (low: extroversion is important vs. high: introversion is important) and specificity of measure (vague: BTA vs. specific: overplacement), for participants who saw the manipulation before making self-assessments in Experiment 2.

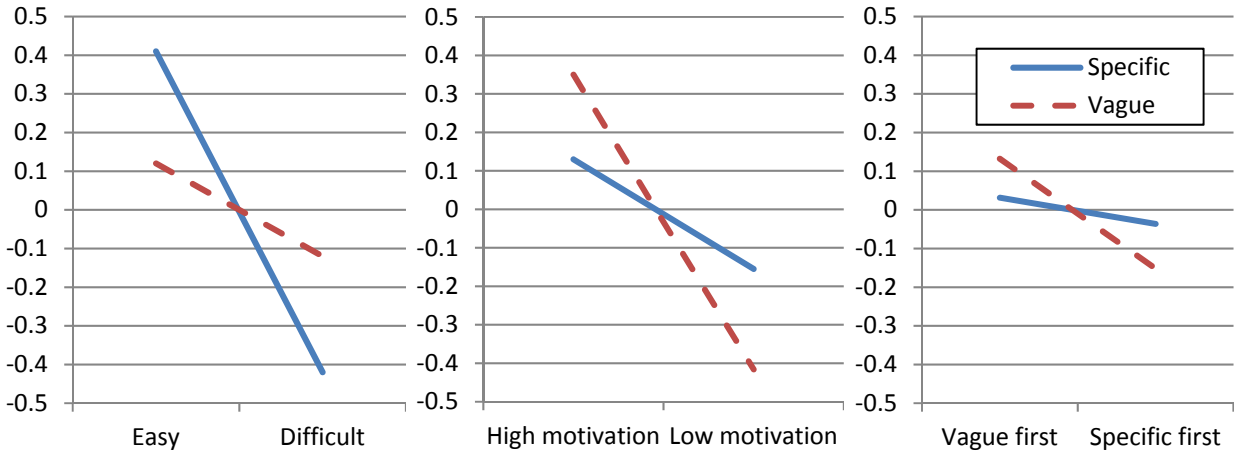

Figure 3. Specific and vague self-reports of participants' scores on the social responsiveness questionnaire (standardized), as a function of the three between-subjects manipulations, Experiment 4.

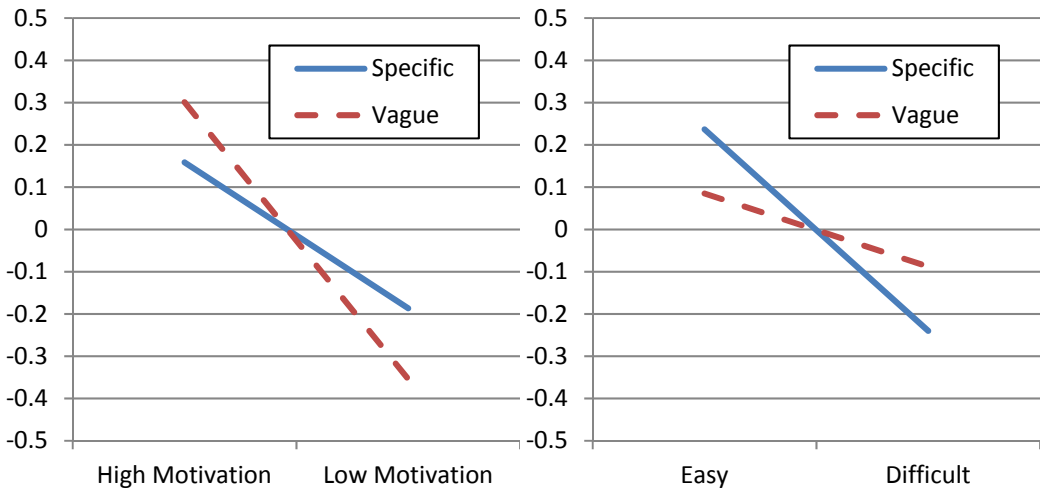

Figure 4. Overplacement of social responsiveness, by motivation, difficulty, and measure specificity in Experiment 4.

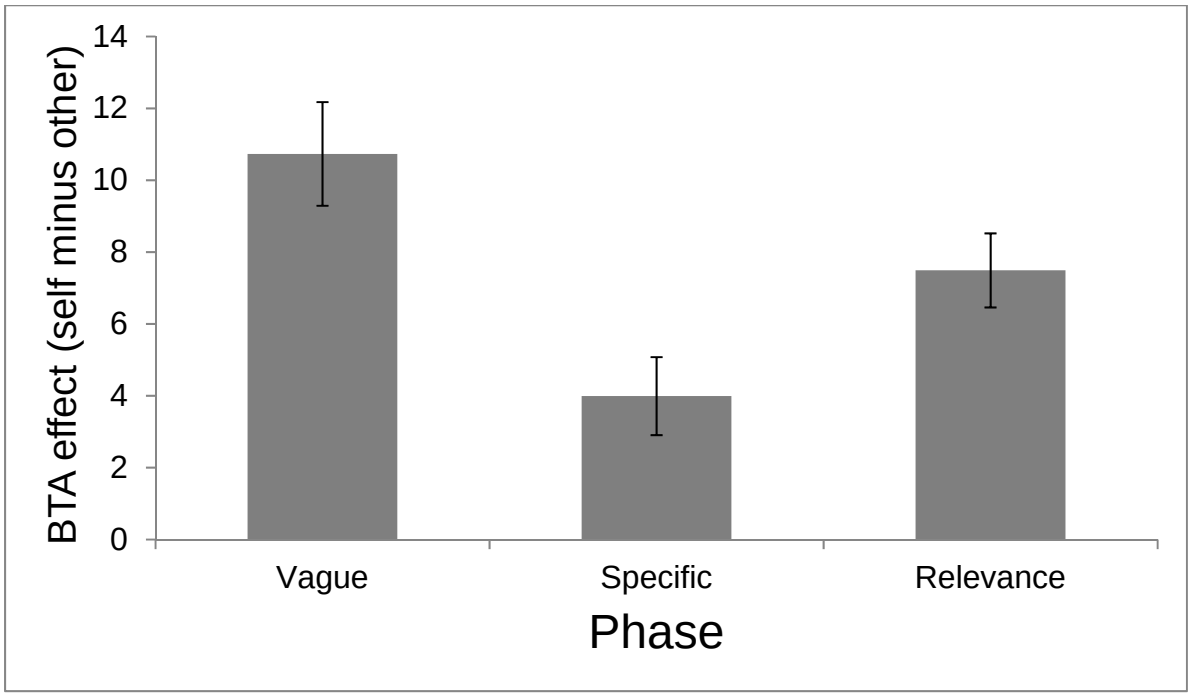

Figure 5. The BTA effect (self and other difference) within each phase in Experiment 6.

Appendix A: Pre-test information and Questionnaire Scoring for Experiment 1

Specificity Pre-Test:

We separately pretested the relevance of specific behaviors for each trait and listed them in the experiment with the first rated as most relevant in the pretest. One hundred and thirty-two online and lab participants pre-tested the specific behaviors following an unrelated experiment. For each trait, participants read a list of ten behaviors (eleven for introversion) and rated five that best captured their idea of the trait. We used the five behaviors most frequently rated either "1" or "2" in terms of highest relevance.

Appendix B: Experiment 3 manipulation text and supplementary analyses

Results

Analyses below exclude the 156 participants who our pre-registered exclusion criteria would have had us drop from analyses: those who completed the survey too quickly (under 5 minutes) or who failed the manipulation check, leaving us with a sample of 235.

Analyses using the full sample without data exclusions

As participants completed the social responsiveness questionnaire before they encountered the motivation manipulation, it could not have affected questionnaire responses. Indeed, responses were similar among those in the high motivation (M = 5.28, SD = .90) and low motivation (M = 5.24, SD = .88) conditions, $F(1, 233) = 0.13$, $p = .72$.

Estimations of Performance (Self-Ratings)

We standardized both specific score estimates and vague ratings of own score and submitted them to a 2 (motivation: *low* vs. *high*) × 2 (order of measures: *vague first* vs. *specific first*) × 2 (specificity of measure: *vague* vs. *specific*) mixed ANOVA with specificity of measures as a repeated measure. The results reveal, predictably, a strong effect of desirability, $F(1, 231) = 101.4$, $p < .001$. They also reveal an interaction between desirability and specificity, $F(1, 231) = 6.68$, $p = .01$, qualified by a three-way interaction, $F(1, 231) = 10.27$, $p = .002$. This three-way interaction, suggests that the effect of desirability is strongest on vague measures, especially when the vague measures come first. None of the other main effects or interactions are significant.

Overestimation

Overestimation is the difference between estimated and actual score. A 2 (motivation: low vs. high) × 2 (order of measures: vague first vs. specific first) ANOVA reveals a main effect for desirability, $F(1, 231) = 58.2$, $p < .001$. Note, however, actual scores on the questionnaire were higher than estimated scores, implying underestimation. The main effect of desirability arises because those in the high motivation condition underestimated their scores less (M = -.15) than did those in the low motivation condition (M = -1.43). Neither main effect for order, $F(1, 231) = 2.06$, p = .15, nor the desirability × order interaction attained significance, $F(1, 231) = 0.02$, $p = .90$.

Overplacement and BTA

We measured overplacement by subtracting an individual's belief that his or her score is better than others' average scores, corrected for the degree to which they actually are better than average: $(estimated\ own\ score - estimated\ others'mean\ score) - (actual\ own\ score - actual\ others'mean\ score)$. In order to compare this measure with our BTA measure, we standardized both of them and submitted them to a 2 (motivation: *low* vs. *high*) × 2 (order of measures: *vague first* vs. *specific first*) ×2 (overplacement vs. BTA) mixed ANOVA. The result reveals a significant main effects of desirability, $F(1, 218) = 79.93$, $p < .001$. This main effect is qualified by a between-subjects interaction between desirability and order, $F(1, 218) = 5.11$, $p = .02$. There is also a within-subjects interaction between desirability and specificity of the measure, F (1, 218) = 5.26, p = .02. This last interaction describes the fact that the effect of the motivation manipulation is stronger for the vague than the specific measure.

Overprecision

We measured overprecision by taking the variance in the reported distribution of others' scores and subtracting it from the actual variance in others' scores. A 2 × 2 ANOVA reveals no significant main or interaction effects, Fs(1, 231) < 2.30, ps > .13. Average variance in the

distributions participants reported for others was 2.81, whereas the actual variance in scores was 1.14, meaning that participants reported distributions that were, on average, less precise (more dispersed) than the true distribution.

Alternative measures of Overplacement

The manuscript reports analyses using participants' point estimates for their scores. We conducted the analyses using

Appendix C: Questionnaire and Passages Used in Experiment 3

Social Responsiveness Questionnaire: How accurately does each statement describe you? 1-7 scale

1. People like to talk with me about various subjects.
2. I am good at identifying the needs of others.
3. When arriving at a new place, I seek out other people.
4. I notice when people are different from me.
5. I notice how people conduct themselves in social situations.
6. I am careful not to hurt other people's feelings.
7. My friends say that I am a good listener.
8. When someone else is under stress, I know what to do to make him or her feel better.
9. I know how to say the right thing at the right time.
10. I am attuned to the thoughts and opinions of others.
11. I know how to handle awkward moments.
12. I often get invited to social events.
13. I like to be included in conversation.
14. I can sense when a friend is in a bad mood.

High Motivation condition: "What is social responsiveness? It is a trait that makes people good listeners and empathetic friends. These skills help people build trusting relationships and successful cooperation with different kinds of people. People with high social responsiveness tend to have more meaningful relationships, a stronger network of social support, and more friends they can rely on. Consequently, they are better able to sense when others need their help, whether that is help finding a job or finding someone to talk to. Because these social relationships are so important for life satisfaction, friends of people with high social responsiveness report being more fulfilled, happier, and even healthier."

Low Motivation condition: "What is social responsiveness? It is a trait makes people attuned to the opinions of others at the expense of their personal values. Being too sensitive to others' opinions leads to stress, social anxiety, and unhappiness. People with low social responsiveness are more independent and decisive. They are less likely to constantly seek the approval of others and less likely to be neurotic. Those with high social responsiveness depend on frequent praise and affirmation from other people. Ironically, this makes them socially awkward and can often impede the development of healthy, trusting relationships built on mutual respect. The result is that those with low in social responsiveness report being more comfortable with themselves, successful, and happy."

Appendix D: Behaviors considered as honest in Experiment 6

- When I make a promise, I keep it
- I do not say things I know to be untrue.
- I do not purposely deceive others.
- I fulfill my obligations and do what I say I will do.
- I do not cheat on my boyfriend/girlfriend.
- If I make a mistake, I own up to it.
- I live according to my own values.
- I do not steal.
- I do not pretend to be something I am not.
- If I find something of value I do my best to return it to the owner.
- I am honest in ways that the above statements fail to capture.