Self-Enhancement Bias and Error: Measurement, Perception, and Motivation

By Patrick R. Heck

B.A., Willamette University, 2010

Sc.M, Brown University, 2013

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in the Department of Cognitive, Linguistic, and Psychological

Sciences at Brown University

Providence, Rhode Island

May 2017

This dissertation by Patrick R. Heck is accepted in its present form by the Department of Cognitive, Linguistic, and Psychological Sciences as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____   _____

Dr. Joachim Krueger, Advisor

Date _____   _____

Dr. Bertram Malle, Reader

Date _____   _____

Dr. Fiery Cushman, Reader

Approved by the Graduate Council

Date _____   _____

Dr. Andrew G. Campbell, Dean of the Graduate School

Curriculum Vitae
Patrick R. Heck
Department of Cognitive, Linguistic, and Psychological Sciences at Brown University
190 Thayer St., Providence, RI 02912

Education

Psychology – Ph.D. Advisor: Joachim I. Krueger (defense date: December 8th, 2016).
Brown University Department of Cognitive, Linguistic, and Psychological Sciences.

Psychology – Sc.M (May, 2013)
Brown University Department of Cognitive, Linguistic, and Psychological Sciences.

Psychology – B.A. (May, 2010)
Willamette University Department of Psychology – *Cum Laude*

Professional Appointments

Visiting Assistant Professor (Spring, 2017)
Brown University Department of Cognitive, Linguistic, and Psychological Sciences

Visiting Instructor (Spring, 2015)
Connecticut College Department of Psychology

Academic Scholarships and Awards

2016 - Deans' Faculty Fellowship: Professorship at Brown University
2015 - Max Planck Research Group: Summer institute participant and travel award recipient
2015 - Peter D. Eimas Research Award: Brown University CLPS Department
2014 - Research and Mentorship Grant: Brown University Sheridan Center for Teaching
2011 - Graduate Fellowship: Brown University CLPS Department
2010 - Paul M. Evans Award: for excellence in Psychology - Willamette University
2010 - Senior Certificate: awarded for institutional service - Willamette University
2009 - Summer Science Fellowship: American Psychological Association
2008 - Research Internship: Macquarie University Department of Cognitive Science, Sydney

Scholarship

Journal Publications

Heck, P.R., & Krueger, J.I. (2016). Social perception of self-enhancement bias and error. *Social Psychology*. Advance online publication. doi: 10.1027/1864-9335/a000287

Heck, P.R., & Krueger, J.I. (2015). Self-enhancement diminished. *Journal of Experimental Psychology: General, 144,* 1003-1020. doi: http://dx.doi.org/10.1037/xge0000105

Heck, P.R., & Krueger, J.I. (2015). Happiness: A theory of relativity. Review of 'The myths of happiness' by Lyubomirsky, S. *American Journal of Psychology*, 128, 126-128. doi: 10.5406/amerjpsyc.128.1.0125

Submitted Manuscripts

Heck, P.R., & Krueger, J.I. (under revision at *Social Cognition*). Social perception in the volunteer's dilemma: Role of choice, outcome, and expectation.

Heck, P.R., & Krueger, J.I. (under review). Go P!: Evaluating (some) critiques of null hypothesis significance testing.

Krueger, J.I., Heck, P.R., & Wagner, D. (under revision at *Journal of Experimental Social Psychology*). Egocentrism in the volunteer's dilemma.

Chapters

Krueger, J.I., Evans, A. M., & Heck, P. R. (2016). Let me help you help me: Trust between profit and prosociality. In Van Lange, P. A. M., Rockenbach, B., & Yamagishi, T. (Eds.). *Social dilemmas: New perspectives on trust.* New York: Oxford University Press.

Krueger, J.I., & Heck, P.R. (in press). The search for the self. In Nelson, T. (Ed.), Getting grounded in social psychology.

Peer-Reviewed Conference Symposium Presentations

Krueger, J.I., & Heck, P.R. (2016, September). *Vicissitudes of self-enhancement.* Symposium session talk presented by Krueger, J.I. at the Society for Experimental Social Psychology (SESP), Santa Monica, CA.

Krueger, J.I., Ullrich, J., & Heck, P.R. (2016, January). *Egocentrism and prosocial behavior in the volunteer's dilemma*. Symposium session talk presented by Heck, P.R. at the Society for Personality and Social Psychology (SPSP), San Diego, CA.

Krueger, J.I., Freestone, D., & Heck, P.R. (2014, July). *Social projection in the inductive reasoning model*. Symposium session talk presented by Krueger, J.I. at the European Association for Social Psychology (EASP), Amsterdam, NL.

Teaching Experience

Courses Taught

CLPS 1782: *The Social Self* (Spring Term, 2017, Brown University).
PSY 493E: Social Judgment and Decision Making (Spring Term, 2015, Connecticut College).
CEPY 0944: Game Theory in Social Decision-Making. (Summer Session 2014, Brown University).

Teaching Assistantship – Brown University

CLPS *2908* (Graduate): *Multivariate Statistics*
CLPS 0900: Quantitative Methods in Psychology (Lectures: Linear Regression)
CLPS 0700: Social Psychology
CLPS 1791: Laboratory in Personality and Clinical Assessment (Section leader)
CLPS 0700: Social Psychology (Lectures: Self-Enhancement, Self-Positivity, and Egocentrism.)
CLPS 1790: Laboratory in Social Cognition (Section leader)
CLPS 0900: Quantitative Methods in Psychology (Lectures: The Sampling Distribution; Correlation & Covariance; Correlation Coefficients.)

Certificates in Teaching - The Sheridan Center for Teaching and Learning

Certificate V – Principles and Practice in Reflective Mentorship (May, 2015)
Certificate III – The Professional Development Seminar (May, 2015)
Certificate II – The Course Design Seminar: Principles and Practice (May, 2014)
Certificate I – Reflective Teaching (May, 2013)

Psychology Department Tutor – Willamette University

PSY 252/253: Research Methods and Statistics I & II (4 semesters)
PSY 210: Introduction to Psychology (3 semesters)

PREFACE AND ACKNOWLEDGEMENTS

I would like to acknowledge the resources, time, and effort spent on this dissertation by others involved in the process. To Joachim Krueger, my PhD advisor, mentor, and friend, I am grateful for six years of close mentorship and collaboration. I hope that we can continue to write and explore together. I am grateful to Bertram Malle and Fiery Cushman, who have been outstanding committee members and mentors in all parts of my graduate education, including my writing, theorizing, teaching, and service. The CLPS department faculty, graduate students, and administrative staff at Brown University have been instrumental in their continued support, debate, and collegiality. The Sheridan Center for Teaching and Learning, alongside the Graduate School at Brown University, has shown me unyielding support and opportunity in pursing my passions for research, teaching, and higher education. To all of those involved in my time spent at Brown: thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Section I: Measurement of Self-Enhancement Bias and Error

The first section of this dissertation proposes, develops, and validates a novel theoretical framework for self-enhancement. This measure, informed by decision theory, contributes to research on self-enhancement and the larger domain of social-cognitive 'biases' by distinguishing between *bias*, or the tendency to claim superiority over others, and *error*, or a situation where this self-superiority claim is proven wrong. This approach is face-valid and theoretically relevant in its simple decomposition of the claim to be better than average into separate categories of bias and error. Similarly, the decision-theoretic measure yields four mutually exclusive categories of decision type, each of which can be used to generate novel predictions within theories of cognition, social judgment, and decision-making. In this dissertation, self-enhancement error (referred to throughout as a False Alarm) will be the target decision or population of interest across seven studies in three major sections. Section I specifically introduces the measure, reports two validation studies, and explores a large pooled dataset of self-estimates, estimates of the average person, and performance on an objective task. Because Sections II and III of this dissertation borrow from the theory and history presented in Section I, I proceed by providing a thorough review of the fraught history of self-enhancement's many conceived measures.

Introduction to Self-Enhancement

Making social judgments compels self-enhancement. Indeed, self-enhancement biases dominate the literature on accuracy in self-judgment, and a recent influential paper cited a dramatic assessment of the costs of self-superiority, stating: "No problem in judgment and decision making is more prevalent and more potentially catastrophic than

overconfidence" (Plous, 1993, p. 217, as cited in Moore & Healy, 2008). The prevalence of self-enhancement is recognized as universal, with current cultural debates on the topic asking not *whether* the bias exists, but *how* different types of people demonstrate it (Alicke & Sedikides, 2011; Gaertner, Sedikides, & Chang, 2008; Sedikides, Gaertner, & Toguchi, 2003). Svenson (1981) famously reported that most people think they are better drivers than average, most college professors think they are better teachers than their colleagues and most surveyed college students report themselves to be well-above the 50th percentile in leadership ability (Dunning, 2012). Self-enhancement (and the similarly inescapable better-than-average effect) has held its ground amidst a crisis of replication in social psychology (Ebersole et al., 2016; Varnum, 2015). These effects are pervasive and robust.

However, in order to satisfy laws of sampling and mathematics, it is necessary that some of these self-enhancers must be wrong. Not everyone can be above average. Research often treats the intra- and interpersonal costs of self-enhancement as self-evident: being wrong *must* be irrational. Inflated and inaccurate self-perceptions often lead to poor decisions (Larrick, Burson, & Soll, 2007; Moore, Kurtzberg, Fox, & Bazerman, 1999) or damaged interpersonal relationships (Hoorens, Pandelaere, Oldersma, & Sedikides, 2012; Paulhus, 1998; Robins & Beer, 2001). Still others, however, argue that self-enhancement has direct hedonic benefits (self-enhancing *feels good*) and is argued to be positively associated with well-being (Taylor & Brown, 1988). A host of moderators have been introduced that speak to the debates over prevalence and adaptiveness. Cultural researchers argue that individualists self-enhance more than collectivists (Heine & Hamamura, 2007), although the evidence suggests that

individualists are simply more likely to *express* self-enhancement (Sedikides, Gaertner, & Toguchi, 2003). Self-enhancement is stronger in the domains of morality, warmth and communion than in orthogonal domains of agency or competence (Allison, Messick, & Goethals, 1989; Brambilla & Leach, 2014; Paulhus & John, 1998). Easy tasks produce the most self-enhancement, though difficult tasks often produce self-effacement (Kruger, 1999). A longstanding debate over self-enhancement processes asks whether the phenomenon is a result of a motivational drive to see oneself in a positive light (Alicke, 1985; Brown, 1986; Brown, 2012), or is simply a byproduct of cognitive systems and statistical thinking (Chambers & Windschitl, 2004; Galesic, Olsson, & Rieskamp, 2012; Moore & Healy, 2008).

Finally, and at the heart of the research presented in this dissertation, is the debate over how self-enhancement should be conceptualized (in broad terms) and measured (in specifics). Does 'true' self-enhancement reflect an individual's comparison between himself and others? Or perhaps a comparison of the self-image to reality? Is self-enhancement, by definition, a social phenomenon? These types of questions have been repeatedly affirmed and denounced, each attempting to conclude whether self-enhancement is beneficial or detrimental, and which types of groups self-enhance more than others, because consensus over measurement and a broader framework for self-enhancement as a social psychological phenomenon is absent.

This section is tasked with reviewing measurement perspectives, proposing an integrative and face-valid measure, and diagnosing self-enhancement error in two studies. The results of these studies show that previous measures conflate perceived self-superiority with genuine superiority. Baumeister, Bratslavsky, Finkenauer, and Vohs

(2001) proselytized that the proportion of self-enhancers in the population is "implausibly high" (p. 348), though there is no rational standard on which to evaluate how much bias might be too much bias. The technique presented in Section I (and expanded in Sections II and III) integrates research on behavioral prediction (Epley & Dunning, 2001; Epley & Dunning, 2006), knowledge claiming (Paulhus, Harms, Bruce, & Lysy 2003), and decision theory (Swets, Dawes, & Monahan, 2000). On a simple performance task, I develop this new measure with the goal of discriminating between irrational self-enhancement *error* and rationally defensible *bias*.

Measurement History

Self-enhancement was originally conceptualized according to Festinger's (1954) theory of social comparison. In order to situate oneself among others, an individual must first construct a notion of own and others' standing. By using (often limited) self-knowledge and an error-prone notion of the 'average' person as a referent, it is easy to conclude that social comparison errors will occur in self-favoring ways. Here, self-enhancement consists primarily of the better-than-average effect (Alicke, 1985; Brown, 1986), where individual rate themselves in direct relation to their notion of the average other. Direct comparison is difficult for several reasons, however, including the fact that providing a single comparative measure obscures information contained within its constituent parts (Krueger, Freestone, & McInnis, 2013; Krueger & Wright, 2011). For example, an individual who praised the self will have the same score on this measure than an individual who diminishes the average person. Though this measure is still used for its simplicity (Eriksson & Funcke, 2014; Guenther & Timberlake, 2012), it is problematic and can in many cases be replaced by a simple self-judgment (Klar & Giladi, 1999). Expanding this direct measure of the better than average effect requires an indirect

approach where individuals rate themselves and others separately. It is the additional job of the individual (or the researcher), then, to subtract the latter from the former, resulting in an enticing single digit estimate of self-enhancement.

This is, however, where accuracy is obscured. A positive score on this measure of Self (S) relative to the average Other (O) indicates self-enhancement (S > O), while a negative score indicates self-effacement (S < O). Only when S is exactly equal to O is a social perceiver cleared from the conviction of bias. The systematic development of measures and terms designed to identify bias is, of course, a symptom of a psychological *zeitgeist* privileging cognitive biases over rational, accurate behavior (Krueger & Funder, 2004). This measure parsimoniously demonstrates the argued conflation of self-perception and reality: those who claim to be better than others are labeled as self-enhancers when their claim may in fact be accurate. Taylor and Brown's (1988) influential article was originally written to target positive *illusions*, and so these individuals who have an accurate self-image should not be grouped with those who misperceive or distort reality.

Continued research developed a new perspective on self-enhancement concerned with measuring social reality (Colvin, Block, & Funder, 1995; John & Robins, 1994). By replacing criterion judgments of the abstracted 'average person' with objective measures of performance or personality, this program took an encouraging step forward in conceptualizing self-enhancement. However, the fundamental concern remained unaddressed. According to the social reality perspective, a self-enhancer is argued to see oneself (S) as exceeding a criterion measure of true score (T). This individual may overestimate her performance on a test (Moore & Small, 2007), or see herself as

friendlier than she is perceived by her peers (Colvin et al., 1995). Self-enhancement was once again reduced to a single numerical estimate, where the only opportunity to *not* be diagnosed with bias is to predict the criterion measure exactly ($S = T$). This approach has been used to study personality, performance, and even prosocial behavior. For example, Epley and Dunning (2000) asked participants to predict how much money they would donate to charity. When later compared to how much these individuals actually donated, a difference as small as $0.01 was enough to demonstrate 'irrational' self-enhancement.

Incorporating accuracy into measures of self-enhancement has been an encouraging (but slow) process. Krueger and Mueller (2002) leveraged the accuracy correlation between self-judgment and performance to explain self-enhancement as a statistical necessity resulting from regression to the mean. Those with high true scores can rationally use this correlation, though imperfect, to estimate their own performance. Because estimates of others tend to be self-projective (Robbins & Krueger, 2005), it is unsurprising that individuals predict others will perform similarly to them. These two results alone are enough to explain the commonly observed pattern where low performers overestimate their own performance but claim to be worse than average, while high performers to the opposite (Moore & Small, 2007; Moore & Healy; 2008). Other approaches have attempted to use the logic of statistical regression to develop self-enhancement measures independent of performance or accurate self-perception to varying degrees of success (Krueger, 1998; Leising, 2015). Still, these (occasionally complex) approaches have been unsuccessful in overtaking the problematic simple difference score measures introduced in the social comparison and social reality perspectives.

It follows, then, that the social comparison and social reality approaches produce conflicting arguments and results regarding the adaptiveness of self-enhancement. Taylor and Brown (1988) classically labeled self-enhancement as a 'positive illusion,' linking social-comparative self-enhancement to positive well-being outcomes. The social reality perspective, however, argues that self-enhancement is harmful, demonstrating a negative link to likeability and social relationships (Colvin et al., 1995; Paulhus, 1998). It is difficult to resolve the adaptiveness debate specifically because the conflicting evidence presented by both camps appears valid yet was obtained under separate conceptualization and measurement assumptions. The ability to address the Taylor and Brown hypothesis must come from a measure that considers both perspectives (Krueger & Wright, 2011).

Several attempts have been made to integrate or move beyond the theoretically limiting social comparison and social reality perspectives. Perhaps the most successful research developed Differential Information Theory (DIT) (Moore & Small, 2007; Moore & Healy; 2008). Here, self-judgments (S) and other-judgments (O) are referenced in relation to true performance (T). This ordering yields a replicable pattern where low scorers overestimate their own performance (S > T) but claim to be worse than average (S < O). Conversely, high performers underestimate their own performance (S < T), but claim to be above average (S > O). Thus, low- and high-performers are at once accurate and inaccurate in their self-judgments. DIT makes the useful prediction that judgments of the average other contain more error than self-judgments, leading to the regressive pattern described above.

Kwan, John, Kenny, Bond, and Robins (2004) reconceptualized self-enhancement according to the Social Relations Model (Kenny, 1994). They treated the social

7

comparison measure as a perceiver effect, because O is estimated by the self-enhancing (or -effacing) agent. The social reality index is a target effect, or, how others view the agent. Self-enhancement, then, is argued to arise as the interaction of perceiver and target effects, independent of the separate effects themselves. This interaction term requires subtracting both the respondent's average judgment of others (O) and the average judgment others make of the respondent (T) from his or her self-judgment (S). Adding the grand mean of all judgments to this term then centers the numerical result around 0. This hybrid index, $S - O - T + M$, (shortened to $S - O - T$ because M is a constant), is discouragingly similar to the sum of the two conventional indices, $(S - O) + (S - T)$, or $2*S - O - T$ (Krueger & Wright, 2011). Again, self-enhancement reduces to a difference score measure predisposed to label individuals as self-enhancers ($S - O - T + M > 0$) or self-effacers ($S - O - T + M < 0$), with no room for accurate self-perceivers.

The novel decision-theoretic approach proposed in this dissertation follows from Moore and Healy (2008) and Kwan et al. (2004), integrating the social comparison and social reality perspectives. Indeed, in order to allow for accuracy in a comparative self-judgment (such as the infamous better than average effect), a measure must consider where the self-perceiver actually stands relative to others. This approach uses performance data to identify those who inaccurately claimed to be better than average, and, importantly, operationalizes the difference between self-enhancement bias and self-enhancement error. By separating false from truthful self-superiority claims, we gain access to two psychologically distinct individuals. This measure is necessarily more conservative than traditional indices because it restricts the criteria for a diagnosis of self-enhancement error. This allows for a more realistic assessment of accuracy and error in

the landscape of self-perception across domains of performance, personality, strategic

interaction, and interpersonal dynamics. Similarly, independent decision categories

beyond self-enhancement error present a useful framework for continued study of less

popular effects (for example, self-effacement and humility).

Self-enhancement bias is defined as claiming to be or have performed better than

the average person. This is a simple adaptation of the better-than-average effect. Self-

enhancement error is defined specifically as *inaccurately* believing to be or have

performed better than average. These distinctions draw a novel distinction between bias

and error (Swets, Dawes, & Monahan, 2000), clearing biased individuals who accurately

claim to be better than average from the charge of error. Three inputs are necessary to

diagnose self-enhancement error: a self-judgment, S, a judgment of the average other, O,

and a criterion representing true score, T. As in the social comparison approach,

individuals provide separate estimates of self and other. These individuals also provide T

having completed some performance task. From these inputs, respondents can be

categorized into four groups according to a full crossing of comparative self-perception

and accuracy (see Figure 1). Following the nomenclature of Signal Detection Theory and

the Neyman-Pearson lemma, those who claim to have performed better than average, S >

O, and did, T > $\overline{T}$, are Hits, H. It should be noted that $\overline{T}$ represents the median rather

than the arithmetic mean. This is because the median avoids problematic cases of

distributional skew: 50% of scores fall above the median, and 50% fall below.

Respondents who believe they scored better than average but did not, T ≤ $\overline{T}$, are errant

self-enhancers, or, False Alarms (FA). Respondents who inaccurately claim to score

below (or equal to) average are Misses, M. Finally, those who claim to score below (or equal to) average are Correct Rejections, CR, if their score fell below the median.

Figure 1. The decision-theoretic approach to self-enhancement bias and error. S represents self-judgement, O represents judgment of the average other, T represents individual total score, and $\overline{\text{T}}$ represents the sample median total score.

Reality

	$T > \overline{T}$	$T \leq \overline{T}$
$S > O$	**Hit**	**False Alarm** *Self-Enhancement Error* ★
$S \leq O$	**Miss** *Self-Effacement Error*	**Correct Rejection**

Perception

Swets, Dawes, and Monahan (2000) encouraged the application of decision theory to decision problems in psychological research. These four independent categories conceptualize the ordering of self- and other-judgments as such a problem. If accuracy is a motivating or desirable force, then decision-makers will seek to make accurate types of decisions (H, CR). However, achieving accuracy is difficult because T, and to a greater extent $\overline{\text{T}}$, is noisy and unclear to self-perceivers. Thus, when choosing to seek accurate category types, individuals must engage in error-management processing to weigh all possible outcomes in order to determine whether the benefit of making a Hit outweighs the detriment of committing a False Alarm (and again for CR and M) (Haselton et al., 2009; Lynn & Feldman Barrett, 2014; Pascal, 1669/1962). In cases where missing a

positive result (a Type II error) is more costly than falsely claiming to be better than average (Type I error) a bias should emerge in the self-favoring direction. Sections II and III of this dissertation continue to explore the consequences and perceptions of decisions categorized in this way. The present section proceeds by exploring these categories on noncomplex, unmotivated performance tasks.

Though the decision-theoretic approach is unique to social judgments, a similar application has been developed for strictly intrapersonal knowledge or familiarity judgments (Paulhus, Harms, Bruce, and Lysy, 2003). Here, individuals who claim impossible knowledge are said to be self-enhancing by "overclaiming." The over-claiming questionnaire (OCQ) comprises 150 items (e.g., "Manhattan Project"), each rated by participants on a continuous familiarity scale ranging from 0 ("Never heard of it") to 6 ("Very familiar"). In fact, one in five items does not exist (e.g., "cholarine;" "plates of parallax). Any familiarity claim with a nonexistent item contributes to an individuals' overall measure of error. However, this technique is limited by its inherently asocial design and is not sensitive to social comparison or self-evaluation. The index proposed in Section I considers both social (comparative) and nonsocial (performance) inputs.

In Study 1, the decision-theoretic approach to self-enhancement bias and error is used to categorize individuals who complete a performance task. Collecting S, O, and T allows participants to be categorized according to conventional scores (S – O, S – T, and S – O – T), and the novel index (S compared to O; T compared to $\overline{T}$). The primary hypothesis, which persists through the entirety of this dissertation, is that conventional measures will overdiagnose self-enhancement error by conflating Hits and False Alarms,

therefore committing a logically invalid reverse inference fallacy (Krueger, 2014; Wason, 1960). Conventional measures treat both Hits and False Alarms as cases of self-enhancement, labeling both types of individual as symptomatic of an error. By removing Hit participants from this large pool of self-enhancers, however, it becomes clear that the total number of errors will be smaller than previously concluded. Heck & Krueger (2015) identify the reverse inference fallacy succinctly: "Error implies bias, but bias does not entail error" (p. 1007). Study 2 replicates the results of Study 1 and extends the validity of decision-theoretic measure by providing some respondents with accurate feedback on their performance. By doing so, the self-enhancement options available to a self-perceiver are constrained as self-perception can no longer be manipulated. Similarly, providing feedback on an individual's performance will reduce the error in noisy self-judgments, and should lead to a population-level increase in accuracy.

Study 1: Self-Judgment and Performance

In Study 1, participants completed a performance task and provided estimates of their own and the average other's score. By combining self-estimates with estimates of the average person, individual performance scores and the average (median) score, a decision-theoretic framework allows those who mistakenly believe to be above average (False Alarms) to be separated from those correctly identify an above average performance (Hits). Here, the tendency to claim self-superiority represents self-enhancement *bias*, while having this claim proven wrong represents self-enhancement *error*. For those who do not make this claim, the decision framework distinguishes between accurate and inaccurate self-effacers. Analyses will be conducted on the prevalence of these four types of decision category, and the prediction that conventional

12

measures of self-enhancement overdiagnose error. These analyses are then developed

using correlational approaches and computer simulations to better understand the

properties of self-enhancement bias and error.

Method

Participants. Participants ($N = 201$) were recruited using Amazon Mechanical

Turk. Ten participants, who failed to complete the task or admitted to using outside help,

were excluded from analysis. Two additional cases were excluded for reporting more

than a single value in response to any of the performance assessment questions. The data

of 189 participants remained for analysis (53 female, median age = 25).

Procedure. An online survey was prepared using Qualtrics (2013) and eligibility

for participation was limited to residents of the United States. The survey comprised a

20-item sports quiz with 10 items of medium difficulty, 5 easy items, and 5 difficult

items adapted from Moore and Small (2007). Answers to each item were entered into a

free-response text box. Inter-item reliability was moderate to high ($\alpha = .75$). After

completing the quiz, participants were asked to provide performance estimates for

themselves, S ("How many questions do you estimate you answered correctly?") and the

average other person, O ("How many questions do you estimate the average other person

answered correctly?") in counterbalanced order using a dropdown menu. Participants'

total scores, T, were tallied after data collection as an index of their performance.

To provide an additional measure of the conventional better-than-average effect,

participants rated the statement "I am more knowledgeable than the average person" on a

scale ranging from 0 (strongly disagree) to 10 (strongly agree). This variable is labeled K

for *Knowledge*. The trivia task and the question for K were presented in counterbalanced

order. One random order for the quiz items was created and used throughout. Finally, participants reported their gender and age. After assuring participants that their answer would not affect their completion approval, an exclusion criterion question asked whether they used any materials outside of their own knowledge to answer the questions (Goodman, Cryder, & Cheema, 2012). Participants were then debriefed and given a confirmation code to enter into the survey client to receive compensation for completion.

Results

Conventional analyses. Table 1 shows descriptive statistics for the three primary variables, S, O, and T, the two simple difference scores of self-enhancement, S – O, S – T, the BTAE index K, as well as all intercorrelations. S judgments were higher than O judgments, $t(188) = 5.02$, $p < .01$, $d = .47$, indicating favorable self-other comparisons. There was, however, no evidence for a social-reality enhancement effect, as S judgments did not differ statistically from T scores, $t(188) < 1$. This lack of a mean-level effect is not an unusual result (Colvin et al., 1995; John & Robins, 1994). Variable K also produced a BTAE; its mean lay above the midpoint of the scale, $t(188) = 5.47$, $p < .01$, $d = .40$.

Most intercorrelations among the primary and the derived measures were positive, as one should expect on empirical and logical grounds. I proceed by noting exceptions. O judgments were unrelated to test scores, suggesting that neither high scorers nor low scorers were particularly biased against the average other. Furthermore, the correlation between T and the social reality index S – T was negative, as should be expected on mathematical grounds (Krueger & Wright, 2011; McNemar, 1969), but the effect was surprisingly small.

Table 1

Descriptive Statistics and Intercorrelations (Study 1)

Measure		Correlation				
	M (*SD*)	O	T	S – O	S – T	K
Self-judgment (S)	12.39 (4.01)	.20	.69	.79	.67	.53
Other-judgment (O)	10.78 (2.79)	-	.04	.32	.23	-.15
Total Score (T)	12.20 (2.98)		-	.60	-.10	.43
Social Comparison (S – O)	1.61 (4.41)			-	.47	.58
Social Reality (S – T)	0.20 (2.98)				-	.29
Direct BTAE (K)	6.39 (2.24)					-

Note. Measures and total scores range from 0-20. The direct measure of the better-than-average effect (BTAE) ranges from $0 - 10$. For all variables, $N = 189$.
If $r \geq .14$, $p < .05$
If $r \geq .18$, $p < .01$

The data fell into the pattern predicted by Moore and Small's (2007) differential-regression model (see also Fiedler & Krueger, 2012). Figure 2 shows that S judgments were slightly regressive with respect to T, b = .93, $r(187) = .69$. This finding attests to the overall accuracy of self-perception. In contrast, O judgments were almost perfectly regressive, b = .04, $r(187) = .04$. Taken together, these two regressions reveal an important and lawful divergence between the social comparison and the social reality approach. As Moore and Small pointed out, low scorers (low T) will likely overestimate their own performance (social reality), while accurately believing that they scored worse than others (social comparison), whereas high scorers show the reverse pattern. Finally, S judgments predicted O judgments, b = .28, $r(187) = .20$, $p < .01$, suggesting social projection. Because projection is imperfect, regression to the mean demands the better-than-average effect.
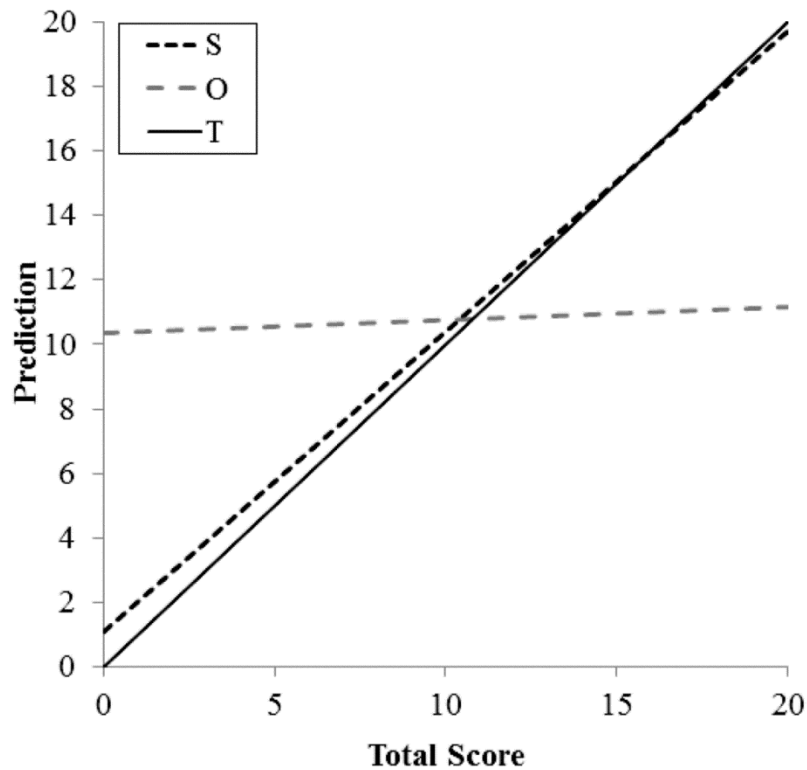
Figure 2. Predicted values generated by regressing self-judgments S, and judgments of the average person O, on total score T (Study 1).

Decision-theoretic categories. Applying the decision-theoretic classification scheme, respondents were sorted into mutually-exclusive categories. Those who thought they scored better than average, S > O, and did score higher than average (median), T > $\overline{T}$, were Hits, H ($N = 79$). Those who thought they scored better than average, but did not, S > O and T ≤ $\overline{T}$, were False Alarms, FA, representing self-enhancement error ($N = 44$). Those who thought they did worse than average and did, S ≤ O and T ≤ $\overline{T}$, were Correct Rejections, CR ($N = 52$), and those who thought they did worse than average, but did not, S ≤ O and T > $\overline{T}$, were Misses, M, or accurate self-effacers ($N = 14$). The correlation between categories of perception and reality was $\Phi = .41$, suggesting a fair

16

degree of accuracy. Notice that the aggregate of the H and FA reflects the self-enhancement bias as seen from social-comparison perspective, whereas only the FA may be said to have committed an error of judgment.

Self-enhancement error. To explore how well the conventional indices of self-enhancement predict error, correlations were computed between their respective difference score measures and a dummy variable, in which a FA (i.e., $S > O \cap T \leq \bar{T}$) was scored as 1 and all other types as 0. The correlations were $r(187) = .22, .41,$ and $.52$ respectively for the social comparison measure $S - O$, the social reality measure $S - T$, and the hybrid measure $S - O - T$ (all $p < .05$).

Conventional difference-score indices of self-enhancement overdiagnosed error. For the social comparison measure, it is easy to see that overdiagnosis is inevitable. If $S > O$, the person is a conventional self-enhancer, but could be either a False Alarm or a Hit in the decision-analytic context. The same is true for the hybrid index. For any False Alarm, the hybrid score will be positive ($S - O - T + \bar{T} > 0$) because $S > O$ and because T, which is subtracted, is smaller than $\bar{T}$, which is added. The reverse is not true. The hybrid score is also positive for a Hit if $S - O > T - \bar{T}$. For the social reality measure, overdiagnosis cannot be shown deductively. A False Alarm ($S > O \cap T \leq \bar{T}$) or a Hit ($S > O \cap T > \bar{T}$) can occur if $S < T$ or if $S > T$.

All three conventional measures overdiagnosed self-enhancement (see Tables 2a-2c for frequencies underlying the computed probabilities). For the social comparison measure, the probability of $S - O > 0$ given a FA was 1, whereas the inverse conditional probability was .36. For the social reality index, the probability of $S - T > 0$ given FA was .77, whereas the inverse conditional probability was .38. For the hybrid index, the

probability of $S - O - T + \bar{T} > 0$ given a FA was 1, whereas the inverse conditional

probability was .37.

Table 2a
Frequencies of Social Comparison Bias and Self-Enhancement Error

	False Alarm	~(False Alarm)	Total
$S - O > 0$	44	79	123
$S - O \leq 0$	0	66	66
Total	44	145	$N = 189$

Table 2b
Frequencies of Social Reality Bias and Self-Enhancement Error

	False Alarm	~(False Alarm)	Total
$S - T > 0$	34	55	89
$S - T \leq 0$	10	90	100
Total	44	145	$N = 189$

Table 2c
Frequencies of Hybrid Index Bias and Self-Enhancement Error

	False Alarm	~(False Alarm)	Total
$S - O - T + \bar{T} > 0$	44	75	119
$S - O - T + \bar{T} \leq 0$	0	70	70
Total	44	145	$N = 189$

Simulating Categorization

Having opened an empirical window into how self-judgments, other judgments,

and performance data can be combined to identify four types of respondent within a

general decision-theoretic framework, the data not only recovered familiar patterns (e.g.,

Moore's differential regression), but also revealed the hypothesized confound between

18

self-enhancement and inaccuracy in conventional measures. These measures

systematically overdiagnosed self-enhancement. Using any of the three conventional

measures, a person supposedly showing self-enhancement was more likely to be accurate

than inaccurate.

To explore the environment of these input variables and resulting indices, a

computer simulation was developed with the goal of understanding how a measure of

False Alarms compares specifically with the hybrid measure of self-enhancement. This

measure is perhaps the most sophisticated recent attempt to measure self-enhancement

bias, and in theory, provides the most conservative test of the decision-theoretic

measure's relative validity. Recall that the hybrid measure also represents an effort to

integrate the social-comparison approach with the social-reality approach, but suffers

from similar limitations by reporting a simple difference score.

Eleven simulations generated 20,000 individuals each. In each simulation, the

accuracy correlation between self-judgments S and total score T was varied from 0 to .99

in steps of .1. All other input parameters remained constant and in alignment with the

results of our empirical study and past findings. Means (standard deviations) were as

follows: S = 14.0 (4.0), O = 12.0 (2.75), and T = 10.0 (3.0).[1] To recreate the relationships

among variables found in Experiment 2, the correlation between S and O judgments

(social projection) was set to .2 and the correlation between O judgments and T scores

was set to 0.

Panel a of Figure 3 shows the overall effect of judgmental accuracy, r_{ST}, on the

relative proportions of the four decision-theoretic categories without regard to the hybrid

---

[1] Two additional simulations were run with the opposite pattern of means (S < O < T), and simulations in which S, O, and T are all equal. The results were analytically predictable, consistent with the empirical conclusions, and did not warrant separate reporting.

19

measure. Predictably, H and CR become more prevalent as accuracy increases, whereas FA and M become less prevalent.

Panel b shows the classifications of respondents with positive hybrid scores (S – O – T + $\overline{\text{T}}$) generated by the underlying input distributions. There are now more False Alarms, as S is on average higher and O is on average lower. When accuracy is low, the hybrid measure is more likely associated with FA than with H. This was an encouraging result for the measure. As accuracy increases, however, a positive hybrid score begins to contain more observed Hits than False Alarms. In other words, judgmental accuracy moderates the degree to which the hybrid measure correctly detects self-enhancement error. Recall from the discussion of reverse inference that a positive hybrid index score is necessary for a False Alarm. The simulations show that even when there is no correlation between self-judgments and test scores, that is, even in the case in which the hybrid score performs best, ~35% of the simulated individuals are correct in their favorable comparative self-judgments. An experiment using only this hybrid measure would fail to detect any of them.

Panel c shows the results for negative hybrid scores. Only a few individuals commit a self-effacement error (M). Even in the absence of accuracy, many of the individuals with a negative hybrid score (~22%) are correct in their comparatively unfavorable self-assessment. As accuracy increases, the hybrid score performs increasingly worse, capturing more CRs and fewer Ms. Once again, the hybrid measure obscured accurate individuals.
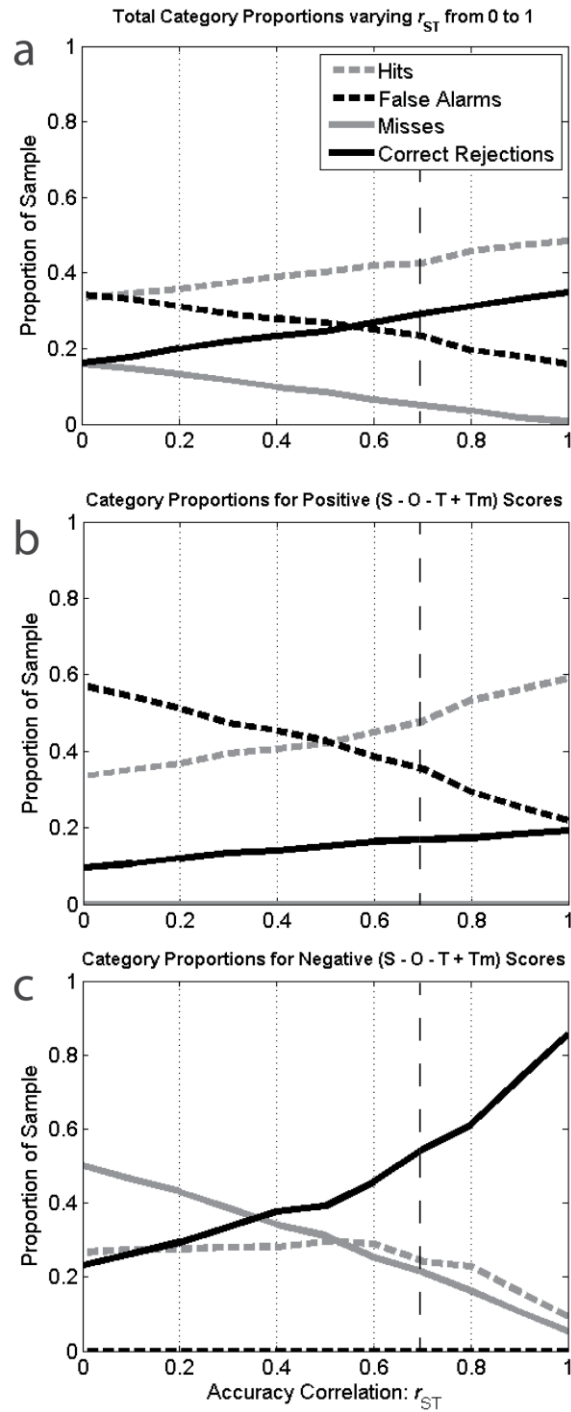
Figure 3. Simulation extending study 1, where accuracy, r_{ST}, is varied from 0 to 1. The hybrid index is computed as $S - O - T + \overline{T}$. Panel a represents the proportion of each category in the simulated population. Panel b displays percentages of each category for those with a positive hybrid index score. Panel c displays these proportions for those with a negative hybrid index score. The vertical dashed line represents the simulation resulting from study 1's empirically observed accuracy correlation ($r_{S,T} = .69$).

21

Study 2: Self-Judgment, Performance, and Feedback

Study 2 had two goals. The first was to replicate the results of Study 1. The second was to add an experimental intervention in order to ensure that the decision-theoretic approach would behave lawfully in responding to changes in the information given to participants. This took the form of a feedback manipulation. Half of participants received their actual test scores after they had estimated them, constraining their decision-making. Knowing their own performance scores, participants could only vary their estimates of the average person. In line with social projection, it was predicted that respondents would use their known T score to predict the score of the average person, O. By doing so, all error is removed from the self-judgment. In sum, respondents should project their own scores to construct O in absence of the noisy and modifiable self-judgment, S. Critically, replacing S with T changes how a False Alarm occurs. Now, a self-enhancement error (FA) arises when $T > O \cap T \leq \bar{T}$, that is, if an individual with an average or worse-than-average score places the average person below his own (known) performance.

How might feedback change the incidence of self-enhancement *error*? The outcome must depend on the difference between S and T, and the projective relationship between T and O. Among respondents who initially overestimated their performance, feedback would reduce self-enhancement error unless O estimates were to decrease so much as to offset the drop from S to T. Conversely, among respondents who initially underestimated their performance, feedback would *increase* error unless O estimates offset the rise from S to T.

A traditional difference-score measure of self-enhancement was included for the dimension of life contentment. To begin to explore the adaptiveness question, a measure of self-esteem was also included. The contentment-related self-enhancement measure should track the difference-score measure, S – O, as both represent self-enhancement bias in judgment. The prediction is less clear for how this measures will predict self-enhancement error. Self-esteem is predicted to correlate positively with the three conventional measures of self-enhancement because each of these difference-scores shares method variance (obtaining self-judgments) (Krueger & Wright, 2011). Though uncertain of this outcome, it is possible that self-esteem correlates positively with self-enhancement error (Taylor & Brown, 1988).

Method

Participants. Undergraduate participants were recruited at Brown University (N = 202). The data of two participants were excluded from analysis because they predicted the average person would only answer 1 out of 30 questions correctly (an extreme outlier of O judgments). The data of 98 women and 102 men remained for analysis. Exact age information was not collected.

Procedure. All materials were formatted using the Qualtrics survey tool (Qualtrics, 2013). Participants first completed the performance and social self-esteem scales of the State Self-Esteem Scale (SSES) (Heatherton & Polivy, 1991). They then completed a 30-item trivia quiz, randomized for each participant, containing 4 easy items, 14 medium-difficulty items, and 12 difficult items (Moore & Small, 2007). Item difficulties were selected to increase score variance while avoiding extreme results. In order to provide feedback to participants in real time, a multiple-choice format was

adopted with one correct answer and three foils presented in a counterbalanced, dropdown menu for each question. After completing the quiz, all participants were asked to estimate how many questions they answered correctly. Participants in the experimental condition were then told their actual score and asked to estimate how many questions the average test-taker answered correctly. No feedback was given in the control condition: participants simply estimated how many questions the average person answered correctly. Finally, participants were asked to rate how content they and the average person are with their life overall on a Likert scale ranging from very discontented (0) to very contented (6). The variable name K was retained in reference to *Kontentment*. The survey concluded with a standard debriefing form.

Results

The general pattern resembled the findings of Study 1 (see Tables 3a and 3b). Difference scores were correlated with their individual components and with one another as one should expect on mathematical grounds. In the control condition, the pattern of differential regression emerged again, such that S judgments tracked T scores more closely than did O judgments. The K variable showed statistically significant self-enhancement both in the control, $t(99) = 4.43$, $p < .01$, $d = .52$, and the feedback condition, $t(99) = 5.84$, $p < .01$, $d = .67$.

Effects of feedback. Tables 3a and 3b show the means and standard deviations for the primary variables and the difference scores, as well as their intercorrelations, respectively for the control (no-feedback) and the feedback condition. The projection hypothesis stated that in the absence of feedback, respondents would base their O judgments on their own S judgments. Consequently, the correlation between O and S

24

should be greater than the correlation between O and T. This was the case, $r_{SO} = .63$, $>$

$r_{OT} = .11$, $Z = 4.88$, $p < .01$ (Steiger, 1980). Conversely, when information regarding test

scores was fed back, projection implies that the correlation between O and T would be

greater than the correlation between O and S. This part of the projection hypothesis was

only marginally supported, $r_{SO} = .13$, $< r_{TO} = .29$, $Z = 1.69$, $p < .09$.

Table 3a

Descriptive Statistics and Intercorrelations (Study 2 Control Condition)

Measure		Correlation				
	M (SD)	O	T	S – O	S – T	K
Self-judgment (S)	17.03 (5.18)	.63	.27	.51	.85	.06
Other-judgment (O)	16.79 (4.78)	-	.11	-.35	.57	-.19
Total Score (T)	21.0 (3.41)		-	.21	-.28	.00
Social Comparison (S – O)	0.24 (4.07)			-	.39	.29
Social Reality (S – T)	-3.97 (4.92)					.07
Direct BTAE (K)	0.49 (1.11)					-

Note. Measures and total scores range from 0-30. The direct measure of the BTAE
from 0-6. For all variables, $N = 100$.
If $r \geq .20$, $p < .05$
If $r \geq .25$, $p < .01$

Table 3b

Descriptive Statistics and Intercorrelations (Study 2 Feedback Condition)

Measure		Correlation				
	M (SD)	O	T	S – O	S – T	K
Self-judgment (S)	17.06 (5.17)	.13	.54	.82	.80	.00
Other-judgment (O)	18.90 (3.63)	-	.29	-.45	-.05	-.03
Total Score (T)	21.15 (3.59)		-	.32	-.06	.00
Social Comparison (S – O)	-1.80 (5.46)			-	.75	.02
Social Reality (S – T)	-4.09 (4.14)				-	.00
Direct BTAE (K)	.64 (1.10)					-

Note. Measures and total scores range from 0-30. The direct measure of the BTAE
ranges from 0-6. For all variables, $N = 100$.

Consistent with the projection hypothesis, O judgments were higher in the feedback condition ($M = 18.86$) than in the control condition, ($M = 16.79$), $t(198) = 3.77$, $p < 01$, whereas the means of S and T did not differ between conditions, $t < 1$. This mean-level effect in O reflected the predicted shift from the use of S judgments to the use of T scores as the preferred basis for estimates of O. In both conditions, S judgments were significantly lower than T scores (control: $t(99) = 8.10$, $p < .01$; feedback: $t(99) = 9.89$, $p < .01$). The difference between S and T in the feedback condition was greater than the difference in O judgments between conditions. This arrangement of input variables suggested an *increase* in self-enhancement error after feedback. This result may appear puzzling, but is logically explained. When respondents had underestimated their performance (S < T), feedback raised the number of individuals who think they scored higher than average. This change increases both Hits and False Alarms unless O estimates increase as much as S estimates do. Because projection to others is not perfect, O estimates can rise only moderately after receiving feedback about S. This was indeed the case. In the control condition, the proportions for the four classifications were 20% H, 26% FA, 21% M, and 33% CR. In the feedback condition, the proportions were 42% H, 35% FA, 6% M, and 17% CR. Feedback thus boosted both Hits and False Alarms. Because the increase in H was greater than the increase in FA, there was a nominal increase in overall accuracy. Whereas in the control condition, the proportion of correct classifications (H + CR) was 53%, ($\Phi = .05$), the corresponding proportion after feedback was 59% ($\Phi = .24$), $p < .19$.

These findings show that the decision-theoretic framework is not only suited to separate bias from accuracy, but is also sensitive to changes in both bias and accuracy

following the introduction of more information. This supports its validity as a measurement tool. The analysis of error overdiagnosis is replicated below.

Diagnosis of error. In the control condition, all three conventional measures were correlated with FA (vs. the other three categories), $r(98)$ = .41, .34, and .57 respectively for the social comparison measure S – O, the social reality measure S – T, and the hybrid measure S – O – T (all p < .05). The probability of a positive social comparison or a positive hybrid score given a FA was again 1, thus making overdiagnosis and the reverse inference fallacy virtually inevitable. Given self-enhancement error (FA), the probability of S > O was .58 and the probability of a positive hybrid score was .57. There was no overdiagnosis for the social reality measure. The probability of S > T given a FA was .46, whereas the inverse probability (FA given S > T) was .60.

In the feedback condition, T scores replaced S judgments so that social comparison reduced to T – O and the hybrid measure reduced to $\overline{T}$ – O. The social reality measure reduced to T – T, or a constant of 0. Self-enhancement error was given by a False Alarm = T > O $\cap$ T $\leq$ $\overline{T}$. Over respondents, False Alarms (vs. all other categories) were not correlated with T – O, $r(98)$ = .12, but they were correlated with the hybrid measure, $r(98)$ = .49. Given a FA, the probability of a positive conventional index, T – O or $\overline{T}$ – O, was 1. The inverse conditional probability, however, was merely .46 in either case. Once again, the data reveal a reverse inference problem where bias is not necessarily symptomatic of error. Individuals charged with self-enhancement by conventional measures were more likely accurate in their self-perception than self-aggrandizing.

Self-esteem. Self-esteem was hypothesized to associate with measures of self-enhancement. The State Self-Esteem Scale (SSES) yielded separate scores for performance and social self-esteem. These subscales were averaged because they were sufficiently correlated (Control: $r = .89$; Feedback: $r = .62$). As has been previously shown, self-esteem was correlated with S, $r = .31$, $p < .01$, but not with O, $r = -.01$ or T, $r = .04$, in the control condition. As a consequence, self-esteem lawfully predicted all difference-score indices of self-enhancement: $r = .38$ (S – O: social comparison), .29 (S – T: social reality), and .33 (hybrid). Critically, self-esteem also predicted self-enhancement error, FA, $r = .26$ (all $p < .01$, all df = 98) but not correct enhancement, H, $r = .18$. The correlation between self-esteem and self-enhancement error remained when correlations involving H were controlled, $r = .33$. The feedback condition yielded a similar pattern. Self-esteem was correlated with S, $r = .40$, but not with O, $r = -.10$ or T, $r = .01$. Yet, self-esteem predicted the modified social comparison measure T – O, $r = .42$, and it was marginally correlated with FA, $r = .16$, $p = .06$, and $r = .19$, $p < .19$, when correlations with H were controlled. There was no correlation with correct enhancement, H, $r = -.01$. In both conditions, self-esteem predicted the difference score variable K (*Kontentment*), $r = .40$ and .20 in the control and the feedback condition, respectively (both $p < .05$, all df = 98). In short, the measure of self-esteem behaved as predicted. The most intriguing result was the positive partial correlation with self-enhancement error. This finding suggests that self-enhancement error specifically may indeed be associated with positive well-being, consistent with the Taylor-and-Brown hypothesis.

Categorizing Pooled Samples

Studies 1 and 2 provided initial evidence that the decision theoretic approach to self-enhancement bias and error could validly discriminate between four categories of individual. Study 1 was conducted using an online population and Study 2 was conducted on university students. Since these studies were initially run and reported (see Heck & Krueger, 2015), additional samples of individuals' performances and comparative self-judgments have been collected in contexts of pilot testing, pretesting, and theoretical exploration. Including studies 1, 2, 5, 6, and 7 of this dissertation, a total of 1,779 participants over 12 independent samples have completed Moore and Small (2008) style trivia quizzes of varying composition and provided self- and other-estimates using similar procedures as the ones reported in Studies 1 and 2. In the spirit of replicability, meta-analysis, and effect size estimation, this large pool allows for a conclusion to be drawn regarding the prevalence of accuracy in the population and the general composition of social-perceivers in the domain of trivia knowledge. These data, and a coding sheet identifying the various contexts of their collection, are hosted at http://www.patrickrheck.com/data--materials.html. A decision-theoretic classification scheme was applied to each sample. The resulting category proportions are presented in Table 4. Tables 5a – 5d present input variable correlations computed over all participants (5a) and separately within each category type (5b – 5d).

Over all 12 samples, 49.5% of participants claimed to have performed better than the average person (S > O). This is remarkably close to the benchmark of rationality (50% claim to be above average, 50% claim to be below average) prescribed by the better-than-average-effect. On simple trivia tasks, it appears to be the case that most individuals do *not* claim to be better than average. It is worth noting that 45% of

participants actually performed above average, despite the decision to use the median as a

measure of central tendency. This is because the alternative category includes those who

performed exactly as the median. As such, it can be argued that this measure does

commit a small violation of distributional assumptions. However, because self-

enhancement is the phenomenon of interest, choosing to sort participants into either

'better than average' or 'no better than average' was a theoretically justified choice.

Turning to category proportions, accuracy again appeared to dominate. 28.7% of

participants correctly claimed to be better than average (H), while 20.9% made a similar,

but erroneous, claim (FA). 34.2% claimed to be worse than (or equal to) average and

were proven correct (CR). The smallest group of individuals, only 16.3% of participants

claimed to be worse than (or equal to) average when they in fact exceeded the average

score. This arrangement yielded categorical accuracy over all participants ($\Phi = .26$). This

classification approach provided strong evidence for a measurable difference between

self-enhancement bias and error, where of the 881 participants who claimed to be above

average, only 371 (42.11%) of them were in error. Indeed, claiming to score better than

average was more likely to be a true statement than a false one.

Table 4

Proportions of Self-Perception, Performance, and Decision-Theoretic Classifications

	Perception (rows) by reality (columns)		
	$T > T_{median}$	$T \leq T_{median}$	
$S > O$	510 (28.7%)	371 (20.9%)	881 (49.5%)
$S \leq O$	290 (16.3%)	608 (34.2%)	898 (50.5%)
	800 (45%)	979 (55%)	Total N = 1,779

Turning to correlations between input variables, it was unsurprising to observe

accuracy (r_{ST} = .72), and social projection (r_{SO} = .72) over all participants. These results

are characteristic of self- and social-perceivers (Robbins & Krueger, 2005; Zell & Krizan,

2014). Indeed, their magnitude can be said to be large in this task domain. The correlation

between O and T (r_{OT} = .70) is less understood. Much of this relationship can be

explained by the high covariant nature of all three input measures: accuracy and social

projection alone are enough to predict a relationship between their two unique input

terms (here, O and T), (Krueger, Freestone, & McInnis, 2013; Krueger, Freestone, &

Heck, unpublished). This relationship diminished substantially, though it remained

positive, when controlling for self-judgments, $r_{OT.S}$ = .36. Thus, performing well on the

trivia task was associated with increased perceptions of the average person.

The present discussion of self-enhancement bias and error warrants targeted tests

of input correlations between Hits and False Alarms. Given that Hits had to correctly

identify their standing relative to the average person to be categorized as such, it may be

the case that self-judgment accuracy is greater for Hits than for False Alarms. This

appeared to be the case, Hit r_{ST} = .83 > FA r_{ST} = .73, Z = 3.79, p < .001. It is

unsurprising to conclude that Hits were better calibrated than False Alarms, and this

result stands as a validity check of the decision-theoretic approach. False Alarms,

however, appeared to exhibit greater social projection than Hits, FA r_{SO} = .93 > Hit r_{SO} =

.89, Z = 3.45, p < .001. This result was intriguing as it relates to Differential Information

Theory (Moore & Small, 2008) and the claim that low performers are especially ignorant

(meta-ignorance; Kruger & Dunning, 1999). DIT assumes that information discrepancies

of oneself self and others behave symmetrically in order to produce the pattern of

overestimation and underplacement for low performers. If it is the case that accuracy is

lower for these individuals (FA, in present terms), then this group in particular may be

forced to rely on other information-gathering practices. Given that social projection is

stronger in uncertain environments (Krueger, 2007), False Alarm targets may rely more

on this process than their more accurate counterparts (Hits). This, however, poses a

challenge for DIT, which parsimoniously relies on symmetrical regression effects alone.

Table 5a

Pooled Sample Intercorrelations over all Observations

Total N = 1,779	S	O
O	0.72	-
T	0.72	0.70

Table 5b

Pooled Sample Intercorrelations over Hits

H (n = 510)	S	O
O	0.89	-
T	0.83	0.77

Table 5c

Pooled Sample Intercorrelations over False Alarms

FA (n = 371)	S	O
O	0.93	-
T	0.73	0.72

Table 5d

Pooled Sample Intercorrelations over Misses

M (n = 290)	S	O
O	0.87	-
T	0.68	0.74

Table 5e

Pooled Sample Intercorrelations over Correct Rejections

CR (n = 608)	S	O
O	0.81	-
T	0.65	0.69

General Discussion

Theoretically prevalent measures of self-enhancement overdiagnose self-enhancement error. By claiming that more error exists than is demonstrated by a fair, face-valid measure, the inferences drawn according to these indices were similar in nature to the very error they seek to detect. Social comparison, social reality, and hybrid social relations measure restricted individuals to be labeled as self-enhancing, self-effacing, or in rare cases, accurate (S = T) or self-assimilating (S = O). The decision-theoretic approach included accuracy as a primary determinant of decision categorization, without losing any of the predictive or theoretical strengths of the social comparison and social reality approaches. This measure detected a reasonable amount of self-enhancement bias in the population, though detected error was substantially lower. Among two studies, exploratory simulations, and a larger pooling of self-judgment and performance data, accuracy was found to dominate in the population when measuring the truthfulness of a self-enhancing or self-effacing claim. This was true over several units of

analysis, including input measures at the mean level, an accuracy correlation r(S,T) at the population level, and category membership within unique samples.

One of studies reported in the pooled analysis is worth detailing (Heck, Krueger, & Sachs, unpublished). Here, Brown university students were surveyed and their reference group in constructing the average person was specified as "the average Brown student." Similar results under these circumstances would strengthen the validity of the technique by replicating the results in a different population and by constraining participants to an ingroup community before collecting their input data. In a between-subjects manipulation, half of participants were told that the average Brown student took the same test and performed either very well or very poorly. Much as Study 2 constrained self-enhancement variation to estimates of the average other, this method solidifies the average performance (O) and allows only for self-enhancing (or -effacing) variation to occur in self-estimates. Interestingly, False Alarm occurrences relative to Hits decreased as the average other was thought to perform well. It may be the case that respondents perceived an increased costliness of claiming (but failing) to be above average as this category became less likely for their (high-performing) peers. Similarly, student perceivers run the risk of committing hubris if claiming to perform even better than high performing others (Van Damme, Hoorens, & Sedikides, 2015). Though merely illustrative in this example, varying input feedback type, directionality, and magnitude offers a useful manipulation tool for future study. The decision-theoretic approach is well-suited to explore the unique contributions of self-judgments and other judgments to self-enhancement bias and error, and may be adapted to explore the prevalence or role of motivated reasoning (see Section III).

Studies 1 and 2 also allowed accuracy to be explored as a criterion variable (Epley and Dunning, 2006). A unique accuracy correlation can be computed over each set of four category types of respondents. In Study 1, self-judgment calibration was substantially higher for Hit respondents, $r_{ST} = .80$, than for respondents who committed a False Alarms, $r_{ST} = .095$. Among self-effacers (S ≤ O), these accuracy correlations were both moderate, $r_{ST} = .38$ and .58, though the pattern similarly suggested that the sample of decision errors (Misses) were less calibrated than the sample of accuracy self-perceivers (Correct Rejections). This pattern was similar in Study 2. This was in line with Kruger and Dunning's (1999) hypothesis that low-performers (FA and CR) are worse at predicting their own performance than high-scorers (H, M). Importantly, FA respondents appeared to be lowest in overall calibration, although this was not true when tested over a pooling of 12 samples of similar tasks.

Bias, error, and rationality. Self-enhancement is a flagship bias of irrational and failed reasoning. It is so prevalent as to invite cheeky demonstrations of anecdotal better-than-average offshoots including college students rating themselves as better lovers than average ("the Good-in-Bed Effect;" Beggan, Vencill, & Garos, 2013), and a sample of prisoners rating themselves as no less ethical, moral, or trustworthy than nonincarcerated others ("Behind Bars but Above the Bar;" Sedikides, Meek, Alicke, & Taylor, 2014). This type of research is intuitive beyond humorous and clever titles because the assumed irrationality of self-enhancement is so strong as to be considered ubiquitous. Taylor and Brown (1988)'s proposals, alongside modern theoretical paradigms seeking to detail all possible errors in human reasoning (Krueger & Funder, 2004), argued that self-enhancement reveals "deeply rooted design flaws of the social mind" (Heck & Krueger,

2015, p. 1017). This type of thinking suggests that if over 50% of individuals claim to be better than average in a given domain, then judgment and decision-making in that domain is unreliable. When attention focuses on group-level effect, similar research may commit an ecological fallacy by assuming that all (or a majority of) individuals in that group behave in the same flawed way. The decision-theoretic approach refutes this conclusion.

Though not a primary goal of the research presented in this dissertation, the inclusion of a self-enhancement scale allowed for commentary on Taylor and Brown's adaptiveness debate. They famously showed that a measure of self-enhancement bias (S – O) correlated positively with self-esteem. Our own analysis demonstrated that self-enhancement *error* was also associated with higher self-esteem. Although this is one point of evidence for Taylor and Brown's claim that ignorance can be blissful, it is important to note that shared method variance may account for this result as self-esteem and self-estimates were collected using similar methods from within the same individual (Krueger & Wright, 2011). For conclusive evidence pointing toward positive or negative outcomes, these outcomes must be measured independently of performance and self-estimates.

A decision-theoretic approach distinguishes bias from error without weighting the costliness of errors. Self-enhancement bias can be thought of as a threshold that must be surpassed before an individual will claim to be better than average. This threshold is restricted by S, O, and the relationship between the two. In absence of any social context, these two inputs should vary equally and independently. However, social projection and the privileging of self-relevant information, or egocentrism, quickly violate this assumption (Krueger, Freestone, & McInnis, 2013; Robbins & Krueger, 2005; Krueger,

36

Freestone, & Heck, unpublished). It is no surprise, then, that self-judgments do more of the 'heavy lifting' in self-enhancement than judgments of others (Klar & Giladi, 1999). Still, the arrangement of S relative to O invites a discussion of decision strategy. Those who think that they outperformed others make a risky bet: they know that they will be either a Hit or a False Alarm and necessarily forego the more neutral possibilities of making a Miss or a Correct Rejection[2]. The social comparison measure, S – O, captures this bias. The unique category of self-enhancement error can be applied using an error management framework to better understand the risks of claiming to be better (or worse) than average.

Error Management Theory can lend credence to the rational side of self-enhancement bias and error. A person who commits a self-enhancement error must be labeled as inaccurate, but is not condemned to irrationality (Einhorn, 1986; Haselton, 2007). This person was merely attempting to manage two possible errors: Type I errors (False Alarms) and Type II errors (Misses). If one error is more costly than another, then biased behavior is rational behavior (Haselton, Bryant et al., 2009; Lynn & Feldman Barrett, 2014). On an anonymous, low desirability task, the psychological risks of failing to outperform the average person are low. So, what reasons do individuals have to *not* claim self-superiority? In the absence of anticipated costs, private self-enhancing claims can be made without concession. Brown (2011) disparaged the use of trivia tasks when attempting to elicit self-enhancing judgments, claiming that such tasks are unimportant ("Use of the word *trivia* denotes it," Brown, 2012, p. 210). Indeed, on a more desirable or important task (for example, performance during a competitive job interview, or on the

---

[2] The costs and benefits of each category are thoroughly explored from a reputational perspective in Section II.

future-deciding MCAT), it would be unusual for an individual to claim above average performance but not actually believe this to be true. In uncertain situations, self-identifying as better than average requires a certain amount of psychological commitment. Similarly, there is little to be gained by claiming inferior performance when one seeks to perform well (but, see Section III for a discussion on self-protection). From a strategic perspective, it is clear that many reasons exist to bias comparative performance estimates in favor of the self.

Section Conclusion

In two studies, computer simulations, and analyses conducted over 12 independent samples, a new approach to measuring self-enhancement bias and error distinguished between those who accurately claimed to be above average (Hits) and those who did so in error (False Alarms). Accuracy was shown to dominate in the population regardless of how it was measured. Conventional and still popular measures of self-enhancement were shown to obscure accuracy in comparative self-judgment, leading to an overdetection of bias in the population. The distinction between bias and error was shown to be theoretically important and to be associated with a common individual difference outcome measure: self-esteem. Section I proposed and validated the framework that Sections II and III apply to social perception, reputation, and motivated reasoning. Having demonstrated that the decision-theoretical measurement approach can distinguish between bias and error, I proceed by asking in Section II whether lay social perceivers are similarly sensitive to this distinction.

Section II: Social Perception of Self-Enhancement Bias and Error

In Section I, the decision-theoretic approach to self-enhancement bias and error was shown to be an effective measurement tool able to distinguish between accurate and inaccurate self-enhancers. The goal of Section II is to demonstrate that lay social perceivers are similarly sensitive to this distinction and that claiming to be better or worse than average can be cast as a decision strategy that may incur reputational benefits and costs. By showing that observers differ predictably in their responses to the four category types, the decision-theoretic approach gains ecological and construct validity. Furthermore, it is argued that this measurement tool can be utilized to better understand social perception and reputational processes. I proceed by introducing social perception and reputation cast according to this framework, followed by two experiments on how individuals perceive self-enhancement bias and error.

Introduction

Social perceivers can detect explicit self-enhancement bias. Those who see themselves as better than others are viewed as disagreeable, poorly adjusted, and narcissistic (Paulhus, 1998), particularly when such self-superiority claims are made public (Hoorens, 2012). Negative perceptions of self-enhancers are associated with relational difficulties, including interpersonal maladjustment and poor social skills (Colvin, Block, & Funder, 1995). Indeed, the fact that self-enhancement is often seen as a negative behavior may partially explain why most individuals ascribe less of it to themselves than to others (Pronin, Lin, & Ross, 2002). Nevertheless, people tend to assume that self-enhancement is a common phenomenon in the social world (Kruger & Gilovich, 1999). Debate over the prevalence of self-enhancement is ongoing. Motivational explanations argue that self-enhancement occurs because it helps to create and maintain a positive self-image (Alicke & Sedikides, 2011; Taylor & Brown, 1988).

39

Cognitive accounts clarify that social comparison biases can occur without a motivational component: a positive self-image and social projection processes alone are enough to produce the better-than-average effect according simply to statistical regression (Heck & Krueger, 2015; Krueger, Freestone, & MacInnis, 2013).

The present section sets aside the prevalence of self-enhancement bias and instead asks about its consequences. It has been argued that the consequences of self-enhancement can be measured according to a reputational perspective, where social perceivers can praise or disparage an individual's reputation (Taylor & Brown, 1988; Hoorens, Pandelaere, Oldersma, & Sedikides, 2012; Van Damme, Hoorens, & Sedikides, 2015). The assumption that this type of research makes is that if social perceivers praise a target, then that target should experience greater well-being. Targets who are disparaged, however, experience an aversive detriment. To date, this line of research has struggled with how to appropriately conceptualize and measure self-enhancement, and in some cases overlooks or confounds orthogonal judgment domains. For example, individuals who claim to be better than average have been shown to be rated as more competent than similar, but self-effacing, targets (Anderson, Brion, Moore, & Kennedy, 2012). Conversely, Anderson, Ames, and Gosling (2008) demonstrated that individuals who self-enhanced in the status domain were punished and rated as unlikable by observers. To address this gap in the literature, I proceed by introducing accuracy into the conceptualization of self-enhancement, in order to distinguish between target self-enhancement bias and error (Heck & Krueger, 2015; Swets, Dawes, & Monahan, 2000). By adopting this approach, it becomes possible to systematically vary the claim an

individual makes, whether evidence exists that support or refutes this claim, and the

domain of observer judgment.

Applying a decision-theoretic framework (as in Section I) allows target individuals

to be categorized into four decision types according to criterion dimensions. The first

dimension is what we label self-enhancement *bias*: does the target self-enhance or not? The

second dimension captures self-perception accuracy. Knowing their comparative self-

judgment, we can identify whether that claim was accurate or not. Crossing these two

dimensions yields four unique types of targets: A Hit (H) claims to better than average and

actually is. A False Alarm (FA) similarly claims to be better than average but is not. This

category represents the target phenomenon: self-enhancement error. A Miss (M) occurs

when a person who claims to be worse than average is actually found to be above average.

Finally, a Correct Rejection (CR) is a person who correctly claims to be worse than average.

In Section I, it was shown that a reasonable proportion of a large sample of participants

revealed self-enhancement bias, with a much smaller proportion of these committing a self-

enhancement error (Heck & Krueger, 2015).

Separating targets along these two dimensions, into four disparate and face-valid

categories, allows us to ask whether observer judgments track comparative self-judgment,

judgment accuracy, performance on an objective task, or any unique combination.

Specifically, the present section aims to explore the reputational consequences of self-

enhancement error (False Alarms) and whether these consequences exceed those of self-

enhancement bias. Two studies seek to obtain observer judgments of each of the four

categories described above. These results can be directly compared with perceptions of

truncated target persons who were only described by their self-perception (claiming to be

better than others or not) or only by their performance (better than others or not). This combination of the unique elements of the decision-theoretic framework allows us to test specific hypotheses and situate our own results in the context of previous work. Before providing these hypotheses, I proceed by reviewing the often-conflicting literature on how observers perceive self-judgment, performance, and accuracy.

Research on the interpersonal benefits of self-enhancement is mixed. The better-than-average effect has been linked to desirable outcomes including increased self-esteem, adjustment, and self-efficacy (Taylor and Brown, 1988; Taylor, Lerner, Sherman, Sage, & McDowell, 2003). However, a host of research concludes that self-enhancers as identified by the social reality approach (a target's self-judgment compared to others' perception of the target) are less liked, less well-adjusted, and more narcissistic than those who do not express inflated self-views (Colvin, Block, & Funder, 1994; John & Robbins, 1994; Lafrenière, Sedikides, Van Tongeren, & Davis, 2015; Paulhus, 1998; Robins & Beer, 2001; Schroeder-Abé, Rentzsch, Asendorpf, & Penke, 2015; Tenney & Spellman, 2011). Because these conclusions are drawn from measurement approaches that differ in theory and formalization, it is difficult to determine when and how self-enhancement is good or bad for an individual. Kwan, John, Robins, Bond, & Kenny, 2004 proposed a resolution to the conflict of measurement, but their measure remains problematic as it still relies on a single difference score that obscures judgmental accuracy. To date, no research has been conducted on how observers perceive self-enhancement as defined by Kwan et al.'s proposed integrative index.

Perhaps the most recent and informative work on how individuals perceive self-enhancers proposed the hubris hypothesis (Hoorens, Pandelaere, Oldersma, & Sedikides, 2012; Exline & Geyer, 2004). Here, individuals who claim to be better than others (or better

42

than average) are argued to express hubristic self-views that necessarily diminish those around them. By doing so, these individuals invite dislike from their peers. Indeed, explicit self-enhancers have been shown to be liked less and viewed as less warm than self-enhancers who do not make their claims explicitly, or who self-enhance in absolute terms ("I am a good friend") rather than comparative terms ("I am a better friend than most others") (Van Damme, Hoorens, & Sedikides, 2015). This work informs future predictions for perceptions of self-enhancement bias but has little to say about judgment accuracy or the distinction between self-enhancement bias and error.

The decision-theoretic approach to self-enhancement bias and error can address limitations in previous work by asking how observers perceive each unique element of a self-enhancing claim, accuracy in that claim, and objective performance. A hypothetical target individual cast according to this approach can explain both sets of results described above. Tommy claims to be more intelligent than average. He rates himself as 8 out of 10 in intelligence and rates the average person a 5. Everyone who knows Tommy, however, rates his intelligence as a 7. Knowing this, one can observe both the bias in Tommy's judgment (8 > 5), and the error (8 > 7). Additional measures reveal that Tommy scored high in both self-esteem and narcissism, which yields the puzzling conclusion that self-enhancement (broadly defined) is both good (higher self-esteem) and bad (higher narcissism). Tommy is an example of the phenomenon of interest: a False Alarm. By introducing targets who vary in both their comparative self-judgment and their performance, observer judgments of the full set of decision types become available.

To study observer judgments, I proceed using a prevalent two-dimensional model of social perception. The two major dimensions, *competence* and *morality*, can be captured

43

using short scales validated in past research. The two dimensions are known by various names, and there are conceptual differences among available theoretical models, though the two are regularly considered orthogonal regardless of their identifiers (Abele, Cuddy, Judd, & Yzerbyt, 2008; Goodwin, Piazza, & Rozin, 2014).

The goal of Study 3 was to obtain observer judgments of the four types of decision category. Here, observers rated targets who were described as having taken either a test of general or moral intelligence, made either an above- or below-average comparative self-estimate of their performance, and were shown to either score above- or below-average. This design allows for variation in observer judgments to be decomposed into factors of the type of test taken by the target (intelligence, morality), information provided by the target (comparative claim, performance), and observers' dimension of judgment (competence, morality). In Study 4, observers also provided ratings of incomplete targets who only provided one piece of information: either their self-assessment or their performance on an intelligence test. A strength of this type of design is that targets' comparative self-judgment represents the classic measure of the social-comparison paradigm (the target either self-enhanced or did not), whereas the target's performance information mirrors the social-reality paradigm (the target overestimated himself or not). The results of these studies can speak to not only how self-enhancement is perceived, but also to how individuals may be able to strategically manage their reputations (Leary & Baumeister, 2000; Paulhus, 1984).

Hypotheses

There were two primary hypotheses in the competence domain. First, it was predicted that observers would judge accurate self-enhancers (H) as more competent than inaccurate self-enhancers (FA) regardless of the performance or judgment domain. This

44

hypothesis should be supported inasmuch as self-judgment accuracy and an above-average performance are valued. Second, observers are predicted to judge accurate (CR) and inaccurate (M) self-effacers similarly. This is because both of these targets warrant perceived competence in one regard; those who exhibit a Miss perform well and those who commit a Correct Rejection demonstrate accuracy in their self-judgment. These predictions are particularly strong for target persons who took an intelligence test and were judged on competence, because here judgment dimension (competence) and criterion dimension (intelligence test) are congruent. This pattern can be formalized as an interaction between reality (above or below average performance) and self-perception (claiming to be above or below average). A main effect of reality is possible such that those who score above average (H and M) may be rated as more competent than those who score below average (FA and CR). However, when considering the accuracy of comparative self-judgments, observers should value accurate self-perception (Tenney, Vazire, & Mehl, 2013). If this is the case, then observers will judge accurate self-perceivers (H and CR) as more competent than inaccurate targets (FA and M). Taken together, these two predictions generate a predicted ordering of competence judgments (H > CR ≈ M > FA) formalized as a statistical interaction between reality (test performance) and perception (claim to be better/worse than average). For those target persons who were described as taking a morality test, competence ratings should follow the same pattern. Because judgment domain (competence) and performance domain (morality test) are no longer matched, however, these effects may be weaker. To summarize these predictions, both successful test performance and accurate self-perception should be praised in the domain of perceived competence.

For judgments of morality, the predicted judgment order shifted. It was hypothesized here that observers might morally credit self-effacers (CR and M > H and FA), raising a new question of whether observers disparage self-enhancers or morally praise humble self-effacers. This question can be explored in Study 4. When completing a test of intelligence, no main effect of reality is expected. Because morality and competence are argued to be orthogonal, performing poorly or well on an intelligence test should not affect perceived morality. However, this main effect was predicted to emerge when considering target who took a morality test. One of the primary goals of Studies 3 and 4 was to determine whether self-enhancement error (FA) would elicit the harshest morality judgments due to a violation of both humility and accuracy (or as it may be considered in some cases, honesty). This is formalized as an interaction between reality and perception, similar to the predicted effect for judged competence. Finally, it is difficult to make a prediction for one group in particular: those who claim to be more moral than average. This type of claim is paradoxical in nature. Here, a person may gain credit for an above average performance on a morality test but be disparaged for predicting this result by claiming to have performed above average.

Study 3: Judging the Four Decision-Theoretic Categories

Respondents received brief descriptions of hypothetical individuals who had taken either a test of general intelligence or a test of moral aptitude. All participants were presented with descriptions of four target individuals representing a full crossing of claiming to have scored better than average (or not) and actually having scored better than average (or not). Hence, the study design had one between-participants variable (type of test taken by the target) and two within-participant variables (perception and reality of the target person's

self-prediction). Respondents rated each target person on a series of trait adjectives used in previous research of this type (e.g. Krueger & Acevedo, 2007) to capture the two prominent dimensions of social judgment, competence and morality (Abele et al., 2008).[3]

Method

Participants ($N = 200$) were recruited on Amazon Mechanical Turk (MTurk; Amazon, 2014). All participants were screened using TurkGate to ensure that they had not previously participated in our studies on self-enhancement (Turkgate, 2013). Participants received $0.30 as compensation. Average completion time was 3:41. The data of two participants, who selected the scale midpoint for each rating, were excluded. Gender and age information was not collected.

Sample size (with $n = 99$ for each of the between-respondents condition) was set so that small to medium effects could be detected with an acceptable probability. G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) estimated that within-respondent tests of reasonable mean differences of $d = .28$ and $.33$ would respectively be statistically significant with a probability of .80 or .90.

Procedures and design. Survey materials were presented online (Qualtrics, 2014) to be accessed by residents of the United States. All participants provided informed consent and were told that they would read about four target individuals who had completed a test of intelligence and who estimated their own performance on that test before knowing the actual result. The test was either one of "general intelligence" or "moral intelligence." No further information was provided about the nature of these tests. The descriptions merely stated that

---

[3] A special issue edited by Abele et al. (2008) provides an in-depth treatment of theory and findings relevant to the two-dimension framework of social perception.

"those who score high on general (moral) intelligence are thought to have a very high IQ (be very moral people)."

The four target persons were characterized as follows: One target had performed above average and perceived himself as above average. This is the condition of high reality and high perception, or a "Hit," H. Another target had performed below average but had perceived himself as better than average (low reality, high perception, or "False Alarm," FA). A third had performed above average but perceived himself as worse than average (high reality, low perception, or a "Miss," M). Finally, one target had both performed and perceived himself as being below average (low reality, low perception, or "Correct Rejection," CR). For example, the description of the Hit in the general intelligence condition read as follows: *"Harry recently took a test designed to assess his general intelligence. When asked to report how he thought he did, he responded, "better than the average person." In fact, it turns out that he actually did beat the average overall score on the general intelligence test."* Each target was presented on a single page and the order of the four targets was randomized for each participant. All target names were male.

Participants were asked to rate each target on three trait adjectives comprising a scale for the domain of competence (intelligent, rational, and naïve (reverse scored)), and in addition, they rated the person on the adjective 'competent' itself. Respondents also rated each target on three trait adjectives related to the domain of morality (ethical, trustworthy, and selfish (reverse scored)), and in addition, also rated the target on the adjective 'moral' itself. Previous research has shown that these two scales are sufficiently reliable and independent of each other (Krueger & Acevedo, 2007; Krueger & DiDonato, 2010). All ratings were made on a scale from 1 (not at all) to 5 (extremely). Trait adjectives were

presented on a single page below the target description and were randomized for each participant. After completing their ratings, participants were directed to a debriefing page and given a code to enter into MTurk indicating that they completed the task.

Results

Analytic strategy and initial findings. Ratings for each target were aggregated into unweighted averages to represent dimensions of competence (intelligent, rational, naïve (reverse scored)) and morality (ethical, trustworthy, and selfish (reverse scored)). As expected, these scales were correlated with their respective single-rating adjective measures of them (competent, $r = .65$; moral, $r = .70$). Both scales had satisfactory reliability (mean inter-item correlations = .38 [$\alpha = .63$] and .48 [$\alpha = .72$] respectively for competence and morality). The two scales were only modestly correlated with each other over respondents and within- and between-conditions, $r(790) = .31$.

Hypothesis tests recruited a set of four two-way analyses of variance (ANOVA) with repeated measures on both variables (*reality*, or the target's above/below average performance, and *perception*, or the target's claim to be above or below average). To take the correlation between the two judgment dimensions into account, four similar analyses with repeated covariates were performed (ANCOVA, Tabachnik & Fidell, 2007, pp. 214-215). In all resulting figures, raw means are shown as columns and the adjusted means are shown as dashed bars ('ghost columns'). The findings were similar regardless of the analytic approach; ANCOVA results will only be mentioned when they depart from the conventional analysis. To represent effect sizes for main and simple effects, Cohen's index d was computed in addition to the η_p^2 index routinely provided by SPSS software.

The differences between and among the four conditions are apparent in the graphed

means and inferential statistics (see Figure 4).
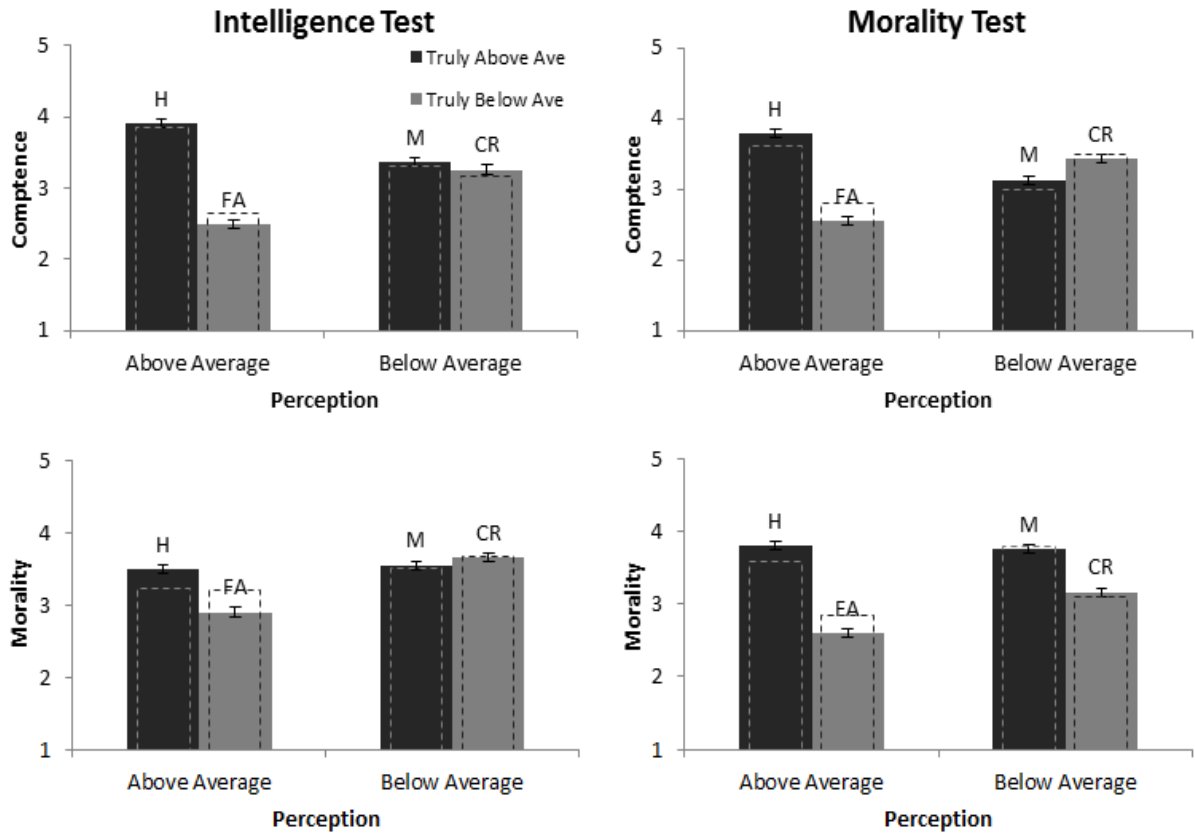
Figure 4. (Study 3) Raw scale means for all four conditions. Dashed bars represent adjusted means after controlling for the rating not displayed (competence controlling for morality and vice versa). Error bars represent one standard error of the mean. H = Hit; FA = False Alarm; M = Miss; CR = Correct Rejection.

Judged competence after intelligence test. For targets who had taken a general

intelligence test, it was predicted that Hits (H) would be perceived as most competent and

that False Alarms (FA, or self-enhancement errors) would be perceived as least

competent. The pattern of means, as seen in the top left panel of Figure 4, fits this

prediction. Observers judged above-average targets (H and M) as more competent than

below-average targets (CR and FA), $F(1, 98) = 197.19$, $p < .01$, $\eta_p^2 = .67$, $d = 2.01$. They

also rewarded accuracy in self-perception (H + CR > FA + M), as shown by the significant interaction between the reality and the perception effect, $F(1, 98) = 125.30$, $p < .01$, $\eta_p^2 = .56$. Simple comparisons revealed that respondents judged those who correctly claimed to be above average (H) as far more competent than those who did so falsely (FA), $F(1, 98) = 274.47$, $p < .01$, $\eta_p^2 = .74$, $d = 2.36$, but they did not differentiate between those who accurately claimed to be below average (CR) and those who falsely did so (M), $F(1,98) = 1.89$, $p < .17$, $\eta_p^2 = .02$, suggesting that self-enhancement error was the greater of two possible decision errors. The main effect of perception, $F (1, 98) = 8.59$, $p < .01$, $\eta_p^2 = .08$, $d = .42$, fell to nonsignificance in the ANCOVA.

Judged competence after morality test. When considering a person who had taken a morality test, a similar pattern emerged (Figure 4, top right panel). As expected, the effect of reality, $F(1, 98) = 54.41$, $p < .01$, $\eta_p^2 = .36$, $d = 1.07$, and its interaction with the target's self-perception, $F(1, 98) = 126.76$, $p < .01$, $\eta_p^2 = .56$, were statistically significant and of medium size. Respondents judged those who made a self-enhancement error (FA) as less competent than those who accurately claimed to be better than avearage (H), $F(1, 98) = 205.03$, $p < .01$, $\eta_p^2 = .699$, $d = 2.05$. They also judged those who correctly perceived themselves to be below average (CR) as more competent than those who falsely claimed to be below average (M), $F(1, 98) = 8.79$, $p < .01$ $\eta_p^2 = .07$, $d = .42$, which was unexpected. The main effect of the target's self-perception was significant, $F(1, 98) = 5.83$, $p < .01$, $\eta_p^2 = .06$, $d = .35$, but this effect again disappeared in the ANCOVA. Taken together, these results suggested that social-perceivers not only rewarded high performance but also accurate self-perception with greater perceptions of competence.

51

Judged morality after intelligence test. For morality judgments about a target person's perceived and actual performance on an intelligence test, it was predicted that respondents would value modesty. Consistent with this hypothesis, the main effect of perception was significant, $F(1, 98) = 59.49$, $p < .01$, and large, $\eta_p^2 = .38$, $d = 1.10$ (see Figure 4, bottom left). Perception and reality together produced an interaction effect, $F(1, 98) = 54.42$, $p < .01$, $\eta_p^2 = .36$, such that targets committing a self-enhancement error (FA) were judged especially harshly. Hits (H) were perceived as more moral than False Alarms (FA), $F(1,98) = 59.68$, $p < .01$, $d = 1.10$, but Correct Rejections (CR) were not perceived as more moral than Misses (M) after correcting for multiple comparisons, $F(1, 98) = 5.19$, $p < .026$, $d = .32$. This interaction effect disappeared in the ANCOVA. Likewise, only the raw data revealed a main effect of reality, $F(1, 98) = 29.49$, $p < .01$, $\eta_p^2 = .23$, $d = .78$.

Judged morality after morality test. For judgments of morality, a target's performance on a morality test was predicted to dominate. This hypothesis was also supported. The bottom right panel of Figure 4 shows that performance on a moral test determined how a person was seen on the dimension of morality, $F(1, 98) = 183.10$, $p < .01$, $\eta_p^2 = .65$, $d = 1.93$. Targets who claimed to be more moral than average were judged as *less* moral than those who claimed to be worse than average on the morality test, $F(1, 98) = 21.56$, $p < .01$, $\eta_p^2 = .18$, $d = .66$. There was also a significant interaction between reality and perception, $F(1, 98) = 33.66$, $p < .01$, $\eta_p^2 = .26$. Interestingly, self-enhancement errors (FA) were judged more harshly than correct modesty (CR), $F(1, 98) = 41.60$, $p < .01$, $\eta_p^2 = .27$, $d = .92$, but respondents did not discriminate between accurate self-enhancement (H) and self-effacement error (M). This suggested that given a worse-than-average performance, the worst claim a target could make was a self-enhancing one.

Study 3 Discussion

Study 3 provided initial evidence that social perceivers are sensitive to the decision-theoretic categorization scheme. Self-enhancement error was shown to be disparaged relative to otherwise poor performance (CR) and similarly self-enhancing claims (H). This study also demonstrated that observers consider the domain of targets' self-judgment (own competence, own morality) and performance (competence test, morality test), producing unique patterns depending on their own domain of judgment (perceived competence, perceived morality). The dimension of judgment appears to be more important than the dimension of self-perception or performance. When observers judged competence, they rewarded both the target's performance and accuracy in self-perception. When they judged morality, observers punished self-enhancement bias but did not reward correct self-enhancement. Finally, observers were more sensitive to the test performance when the test was a moral one.

Study 4 was designed with two primary goals. The first was to replicate the two focal tests: the interactive effect of perception and reality on competence judgments and the negative effect of self-enhancement bias on morality judgments. The second goal was to introduce judgments of the constituents of each category type (H, FA, M, CR) by providing only partially described targets. As such, Study 4 introduces targets for whom only information regarding their self-perception or only information regarding their test performance was available. To maintain power levels in the presence of a new between-subjects manipulation, targets were restricted to having taken only a test of general intelligence.

Study 4: Judging Complete and Incomplete Category Types

The design, procedures, and analytical approach were similar to the ones described in Study 3. In Study 4, however, some observers were asked to rate a different series of four targets. Participants observed targets who thought they scored better (or worse) than average when no performance information was available. To the extent that only social comparison information (a target's claim) is available to observers, this approximates the social comparison approach to self-enhancement (see Section I). Respondents in the baseline condition also judged targets who only provided performance information and whose self-estimate was unknown. These targets approximate the social reality approach to self-enhancement, where performance represents an objective criterion. These extensions generate two additional hypotheses: First, self-enhancement bias in absence of performance information may lower judgments of morality because participants cannot tell whether this claim is accurate or inaccurate. Second, low performance may not only lower judged competence but also judged morality (a halo effect).

Method

Participants. Participants ($N = 200$) were recruited on Amazon Mechanical Turk (MTurk; Amazon, 2014), resulting in 100 participants per sample. As in Study 3, TurkGate (2013) ensured that respondents had not contributed to earlier studies. Participants received $0.30 for completing the task (mean completion time: 3:53). The data of two participants were excluded from analysis because all ratings were the midpoint of the scale, yielding data from 198 participants in total. Power estimates for tests of correlated means were the same as in Study 3. For comparisons of independent samples, mean differences of $d = .40$ and .46 would respectively be significant with a probability of .80 or .90.

Procedures and design. All survey materials were presented online using Qualtrics survey software (Qualtrics, 2014). Eligibility to complete the survey was restricted to individuals residing in the United States. All participants gave informed consent. They were told that they would rate four target individuals who completed a test of general intelligence and subsequently exhibited some judgment or behavior.

Participants assigned to the experimental condition ($n = 99$) received complete information about the targets: the materials and the procedures were the same as in Study 3, although all participants were told that targets had completed a test of general intelligence. Participants in the control condition ($n = 99$) received only one piece of information regarding either the target's performance (reality) or comparative self-judgment (perception). These participants rated four targets who thought they had performed better than average on the test (high perception), thought they had performed worse than average (low perception), had indeed performed better than average (high reality), or had performed worse than average (low reality). Instructions clarified that trimmed targets who perceived themselves as better or worse than average did not know their actual scores. Each target was presented to participants on a single page and the order of these pages was randomized for each participant.

All participants were instructed to rate each target on the six trait adjectives representing facets of competence and morality. After completing their ratings, participants were directed to a debriefing page and given a completion code to enter into Mturk indicating that they completed the task.

Results

The two scales again showed satisfactory reliability (mean inter-item $r = .31$ [$\alpha =$ .50] and .48 [$\alpha = .69$] respectively for competence and morality,[4] and the scale scores were moderately correlated with each other, $r(394) = .51$ and .35 respectively in the experimental and the control condition. ANOVA and ANCOVA tests were performed separately and produced both raw and adjusted means (see Figures 5 and 6; adjusted means presented as dashed ghost columns).

Judgments of fully described targets. The pattern of results for competence judgments in the experimental condition was similar to that found in Study 3 (Figure 5, top). The focal hypothesis was supported by the interaction between perception and reality, $F(1, 98) = 132.21$, $p < .01$, $\eta_p^2 = .57$. Targets correctly claiming to be above average (H) were judged most favorably and targets committing a self-enhancement error (FA) were judged least favorably, $F(1, 98) = 230.64$, $p < .01$, $\eta_p^2 = .70$, $d = 2.17$. There was no difference between those who claimed to be below average, such that those who were incorrect (M) were perceived as no more competent than those who were correct (CR), $F(1, 98) = 1.73$. The main effect of reality demonstrated that above-average targets (H and M) were rated as more competent than below-average targets (FA and CR), $F(1, 98) = 153.95$, $p < .01$, $\eta_p^2 = .61$, $d = 1.77$, while the main effect of perception was not significant, $F(1, 98) = .41$ (although it was in the ANCOVA).

Morality judgments also replicated the pattern found in Study 3 (Figure 5, bottom). The critical finding was that respondents were sensitive to the target's self-perception. Self-enhancers (H, FA) were judged more negatively than self-effacers (M, CR), $F(1, 98) = 47.01$, $p < .01$, $\eta_p^2 = .32$, $d = .98$. They were also sensitive to performance, judging above-

---

[4] The low reliability in competence ratings was due primarily to the adjective 'Naïve.' Removing this item from analyses caused no notable changes in the results or interpretations; all reported results include this item.

average performers more favorably than below-average performers, $F(1, 98) = 20.70$, $p <$ .01, $\eta_p^2 = .18$, $d = .65$. There was an interaction between perception and reality, $F(1, 98) = 58.16$, $p < .01$, $\eta_p^2 = .37$. Respondents judged targets who correctly considered themselves below average (CR) as more moral than targets who did so incorrectly (M), $F(1, 98) = 7.35$, $p < .01$, $\eta_p^2 = .07$, $d = .38$, an effect not observed in Study 3. Respondents appeared to morally penalize false modesty. Finally, there was a large difference between the two targets who claimed to be better than average (H > FA), $F(1, 98) = 69.23$, $p < .01$, $\eta_p^2 = .41$, $d = 1.19$, but interestingly, this effect disappeared in the ANCOVA. This ANCOVA result was the only observation across Studies 3 and 4 where False Alarm targets were *not* perceived as the lead favorable category, though it is unclear why.

Figure 5. (Study 4, complete information condition) Raw scale means for competence (top panel) and morality (bottom panel). Dashed bars represent adjusted means after controlling for the rating not displayed (competence controlling for morality and vice versa). Error bars represent one standard error of the mean. H = Hit; FA = False Alarm; M = Miss; CR = Correct Rejection.

Judgments of trimmed targets. A series of analyses was conducted on the ratings of targets in the partial information (baseline) condition. Targets who scored above average on an intelligence test should be seen as more competent than targets who did not. It was less clear, however, if this should also hold true for targets who merely claimed to be above

average. It might turn out that when true performance is unknown, targets who claim to be better than average are judged as *more* competent than targets who do not (Anderson, Brion, Moore, & Kennedy, 2012; Lamba & Nityananda, 2014). The reason is that respondents may – correctly – assume that reality and perception are positively correlated, recognizing self-judgment as a valid cue for performance. If so, a self-enhancer is more likely to be accurate (H) than inaccurate (FA), and should be rewarded with a high competence rating. The effect of reality was expected to be larger than the effect of perception because correlations between reality and perception are imperfect, thus eliciting a regression effect for variation in ratings of self-perceptions. The findings (see Figure 6, top) confirmed a large effect of reality, $F(1, 98) = 78.74$, $p < .01$, $\eta_p^2 = .45$, $d = 1.27$. There was also an effect of perception favoring the self-enhancers, $F(1, 98) = 16.75$, $p < .01$, $\eta_p^2 = .15$, $d = .59$.[5] To summarize, a person showing a self-enhancement bias was given the benefit of the doubt because the self-prediction was more likely correct than incorrect.

For judgments of morality (see Figure 6, bottom), the critical prediction was that self-enhancers would be judged as less moral than humble self-effacers. This turned out to be the case, $F(1, 98) = 6.90$, $p < .01$, $\eta_p^2 = .07$, $d = .38$. There was no effect of reality, $F(1, 98) = 2.20$, $p < .14$.

---

[5] This conclusion was upheld by a significant interaction term in a 2 (above average vs. not) by 2 (reality vs. perception) ANOVA, $F(1, 98) = 24.40$, $p < .01$, $\eta_p^2 = .199$.
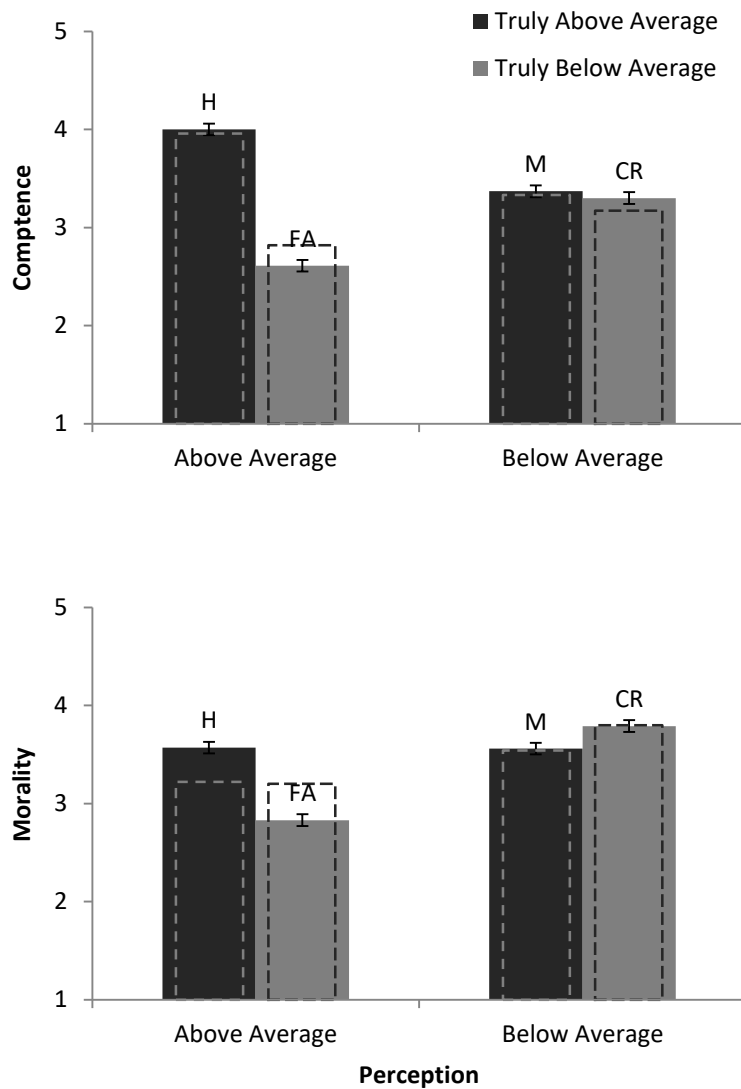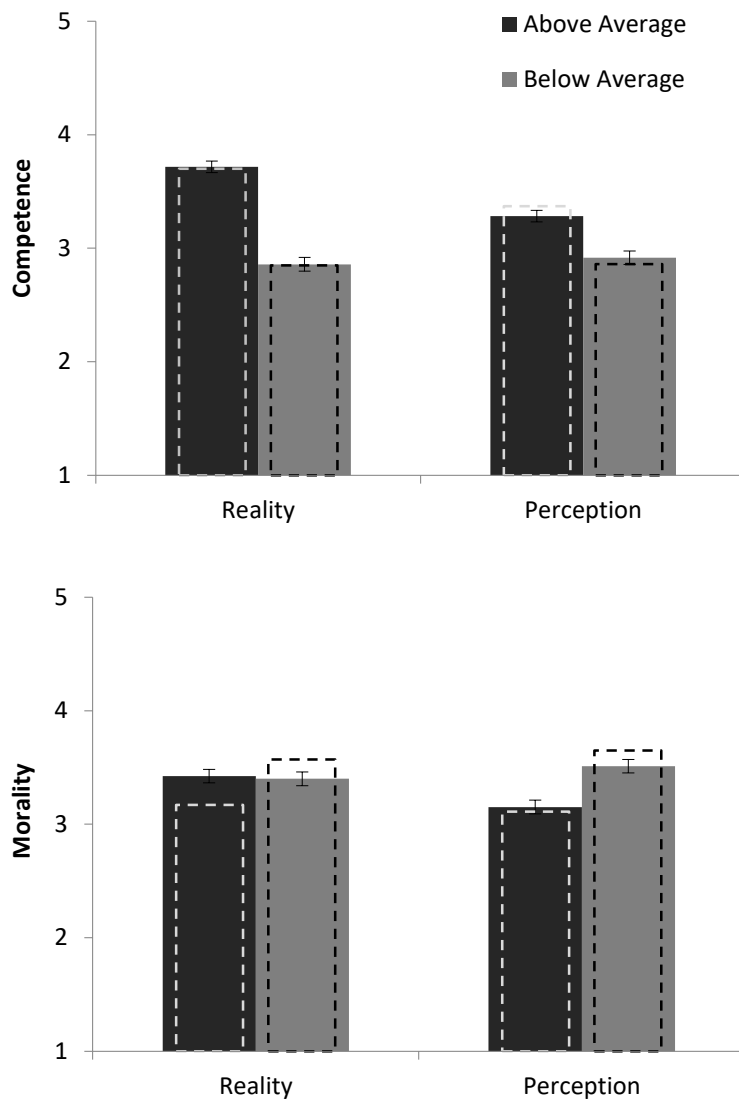
Figure 6. (Study 4, baseline condition) Raw scale means for competence (top panel) and morality (bottom panel) ratings. Dashed bars represent adjusted means after controlling for the rating not displayed (competence controlling for morality and vice versa). Targets were those with only one piece of descriptive information: either their performance or their perception. Error bars represent one standard error of the mean.

Comparing judgments of partially and fully described self-enhancers. The

ecological and construct validity of the decision-theoretic approach to self-enhancement bias

and error requires a psychological difference between those whose self-enhancing

perceptions are justified by reality (H) and those whose errors are revealed (FA). For fully

described targets, judgments of competence revealed an interaction between perception and reality which could not be explained by the social comparison (perception) or social reality (performance) perspectives aalone. To proceed, direct comparisons were tested between judgments of H and FA targets and their individual constituents.

The findings for H targets are displayed in the bottom panel of Figure 7. First, full targets who were accurate are compared with their constituent baselines. Respondents judged a successful self-enhancer (H) as more competent ($M = 4.0$) than a self-enhancer whose performance was still unknown ($M = 3.37$), $t(196) = 8.72$, $p < .01$, $d = 1.08$, or a successful target whose self-perception was unknown ($M = 3.71$), $t(196) = 3.77$, $p < .01$, $d = .54$. This pattern supports the idea that judgments of competence are an additive function of perception (positive self-judgment) and reality (high performance). Judgments of morality showed a similar but attenuated pattern. H targets ($M = 3.57$) were judged as more moral than those who merely claimed to be better than average ($M = 3.15$), $t(196) = 4.37$, $p < .01$, $d = .71$, and as marginally more moral than above-average performers ($M = 3.42$), $t(196) = 1.83$, $p < .07$, $d = .26$.[6]

Additional analyses reveal the unique aspects of self-enhancement error. The findings for FA targets are displayed in the top panel of Figure 7. The FA target was judged as less competent ($M = 2.61$) than someone who only claimed to be above average, ($M = 3.28$), $t(196) = 7.60$, $p < .01$, $d = 1.10$, and also as less competent than a below-average target of unknown self-perception ($M = 2.86$), $t(196) = 3.07$, $p < .05$, $d = .44$.[7] Finally, the FA target was judged as less moral ($M = 2.83$) than a self-enhancer of unknown

---

[6] These last two comparisons were not significant in ANCOVA.
[7] Not significant in ANCOVA

performance ($M = 3.15$), $t(196) = 3.43$, $p < .05$, $d = .49$,[8] and as far less moral than a below-average target of unknown self-perception ($M = 3.40$), $t(196) = 6.74$, $p < .01$, $d = .96$. Social perceivers disparaged self-enhancement error.

Exploring the data post-hoc revealed an unexpected but theoretically coherent difference in the judgments of fully and partially described targets. Targets claiming to be below average were rated as less competent ($M = 2.92$) than similar claimants whose perceptions were known to be either correct, $M = 3.30$, $t(196) = 5.04$, $p < .01$, $d = .72$ or incorrect ($M = 3.37$), $t(196) = 5.97$, $p < .01$, $d = .85$. This pattern suggested a subtle and novel bias: fully described self-effacers were given higher competence ratings regardless of whether their performance validated or violate their self-judgment. If their self-effacement was correct, these individuals were judged as competent because they made an accurate self-judgment. Conversely, if their self-effacing prediction was wrong, these individuals were also judged as competent because they actually performed well on the test. From a position of strategic self-presentation, individuals with low self-confidence would do well to display the results of their performance to others regardless of what these results indicate. It should be noted, however, that judgments of morality did not show this pattern. Respondents only judged CR ($M = 3.79$), but not M ($M = 3.56$), as more moral than those who simply perceived themselves to be worse than average ($M = 3.51$), $t(196) = 3.08$, $p < .01$, $d = .44$; $t(196) = .516$.

---

[8] Not significant in ANCOVA

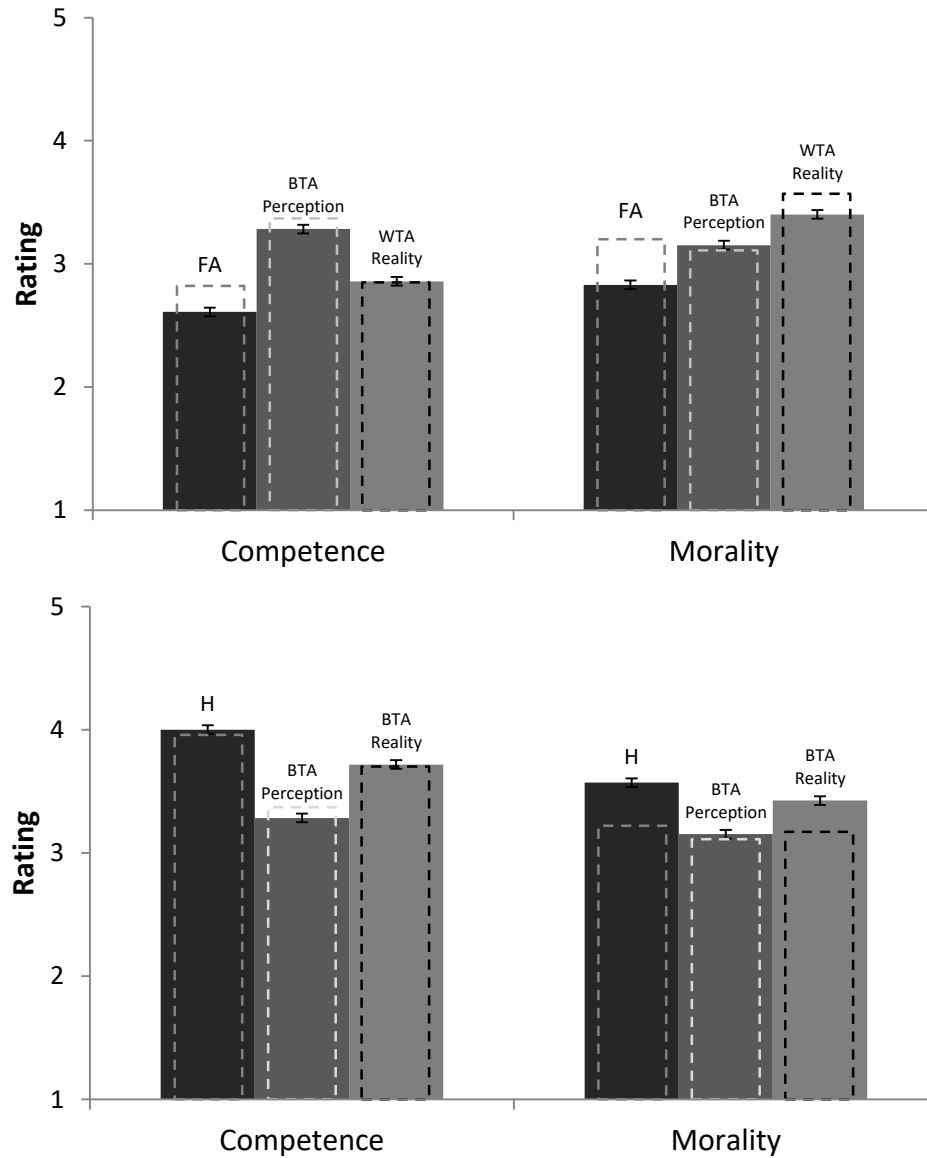*Figure 7.* (Study 4) Complete (full category condition) and incomplete (baseline condition) comparisons between False Alarm and Hit targets to their relevant baselines. Shaded bars display raw scale means. Dashed bars display adjusted means controlling for the rating not displayed. Error bars represent one standard error of the mean. FA = False Alarm, H = Hit, BTA = better than average, WTA = worse than average.

Discussion

      The results of Study 4 tracked those of Study 3. For fully described targets, observer judgments were sensitive to perception and reality information and to the dimension of evaluation. Judgments of competence depended on both test performance and on the accuracy of self-prediction. Observers clearly discriminated between correct and incorrect self-enhancement. In contrast, judgments of morality depended mainly on the direction of the target's self-perception. Observers judged both correct and incorrect self-enhancers negatively.

      Judgments of trimmed targets further supported the idea that in the domain of competence, both the target's self-perception and performance matter. When judging the competence of targets described only by their performance or only by their self-perception, observers seemed mindful of the general accuracy of self-perception (Zell & Krizan, 2014); they assumed that a self-enhancer's claim is more likely to be true than false (see also Anderson et al., 2012; Kennedy, Anderson, & Moore, 2013). Finally, self-enhancement bias was associated with reduced judgments of morality.

<div align="center">

General Discussion

</div>

      The decision-theoretic approach to self-enhancement bias and error was a useful tool in understanding how observers perceive comparative self-judgments, the accuracy of those self-judgments, and objective task performance. Observers judged self-enhancers as highly competent when evidence of their performance supported claims to be better than average, but also when no performance data were available. In other words, participants viewed self-enhancing claims as more competent than self-effacing ones so long as contradictory evidence was unavailable. Importantly, and of critical importance to these studies and

dissertation project, participants viewed inaccurate self-enhancers (False Alarms) as especially low in competence. Indeed, a self-enhancement error was the worst possible decision a target could make. Performance had less influence over perceived morality. Here, observers viewed any claim to be above average as less moral than the humble alternative, regardless of task performance. These patterns help resolve the theoretical inconsistency between Anderson et al. (2012), who found evidence for greater perceived competence of self-enhancing targets, and Hoorens et al. (2012), whose results showed that self-enhancers were viewed less likable and less interpersonally skilled. These previously conflicting results, demonstrating that self-enhancement is both beneficial and detrimental to a target's reputation, are in line with our own when considering the distinction between social perceptual domains of competence and morality.

Implications and limitations

Given the demonstrated reputational risks of expressing both self-enhancement bias and error, one can ask why so many people claim to be better than average. It could be the case that the self-positivity garnered from self-enhancing outweighs the potential reputational risks. Another possibility is that individuals are unaware of how others perceive their claims and behaviors, or that self-perceivers suffer from a reputational bias blindspot (Pronin, Lin, and Ross, 2002). Though unlikely in absolute terms, it could also be the case that individuals care more about their perceived competence, thus justifying their self-inflating claims (Wojciszke, 2005). Here, individuals who bet on being better than average improve the impressions they make by explicitly claiming self-superiority. If these individuals are not obligated to provide confirmatory evidence for their better-than-average claim, this is a fruitful strategic approach.

It is important to note that task context matters. When people consider how they compare with others on some performance task or personality domain, they necessarily enter a relative and perhaps even competitive frame of mind (Festinger, 1954; Tesser, 1988). This context aligns with measures and judgments of competence, where performance is strictly comparative, but does not apply as well to moral domain, which is experienced more often as an absolute (i.e., right or wrong) (Baron, 2012). This creates an interesting paradox where individuals tend to see themselves as more moral than average yet are inexperienced in directly comparing morality between others. It raises the theoretical question of what it means to construct a notion of 'average' morality to begin with. Yet, the results showed that observers were willing to denigrate these 'morality-enhancing' targets despite lacking a clear image of what such enhancement might look like or indicate in an interpersonal environment.

The results of Study 3 and 4 suggest that the decision to self-enhance (or –efface) may be motivated by motivational or reputational concerns. On one hand, claiming relative superiority improved perceptions of targets' competence. Conversely, a similar claim caused a decrease in perceive morality. This pattern was the exact opposite for humility claims. Experiment 2 illustrates this most clearly (control condition; Figure 6, both panels, rightmost bars), where claiming to be better than others was seen as more competent, and less moral, than choosing to remain humble by claiming to be worse than average. What drives this difference between perceptual domains, and how should individuals approach the decision of whether or not to self-enhance? Morality is often described as part of the 'essential' self and is a critical determining factor in lay perceptions of 'what a person is really like' (Hartley et al., 2016; Strohminger & Nichols, 2014). Morality is also more

important to observers than competence when judging others (Goodwin, et al., 2014). Here, it is reasonable to suggest that more can be lost or gained in the moral domain when making social comparative claims. Another consideration individuals may make is that hubristic or arrogant individuals may be seen as morally corrupt or tainted (Rozin, Millman, & Nemeroff, 1986), which can serve as a cue that these individuals will continue to behave in unappealing, self-superior ways. A simpler possibility is that domains of perceived morality and competence are weighted differently in how observers respond to social comparative claims. If it is the case that negative information looms larger in the moral domain and positive information is more important in the competence domain (i.e., Klein & Epley, 2016; Reeder & Brewer, 1979; Skowronski & Carlston, 1989), then one negative observation (a better-than-average claim) and one positive observation (performing better than average) should result in perceptions of low morality and high competence. However, this explanation also assumes that self-perception and reality are also equally weighted. Though intuitive, this assumptions has yet to be verified given the lack of research on perceptions of self-judgment accuracy.

It is unclear whether there will ever be a single answer to the adaptiveness question of self-enhancement. However, it is clear that disparate methodologies and frameworks should be integrated in order to better understand their unique contributing effects. Self-perception, the dimension of judgment, and the presence of outcome information are all relevant to social perceivers, as demonstrated in Studies 3 and 4. Indeed, future research would do well to include these three factors in future exploration, rather than each in isolation. This three-dimensional space makes the individual's decision to self-enhance a complex one that, when implemented successfully, relies on keen social awareness, high

self-knowledge, and several unknown or uncertain factors including task domain, observers and target characteristics, and past behavior. An omniscient self-enhancer would know the results of this long line of research before making a decision, but instead, self-perceivers must rely on heuristics and their own experiences.

Studies 3 and 4 were faced by several limitations. For parsimony and more exploratory power, Study 4 focused only on targets who completed tests of general (and not moral) intelligence. Similarly, the present findings are limited to perceptions of male targets. Self-enhancing (or -effacing) women may be perceived differently, and gender may interact with other environmental factors such as status, workplace vs. home, and culture. For example, it has been shown that dominant and task-motivated women are seen as less likable, less competent, and more threatening than similar men (Carli, LaFleur, & Loeber, 1995). Because all data were collected from individuals residing in the United States, the results similarly generalize only to largely western, independent culture. Ongoing work on the cultural implications of self-enhancement makes unclear predictions for the patterns of interest in Eastern societies (Sedikides, Gaertner, & Toguchi, 2003). Finally, again following from Tesser (1988), accurate self-enhancement claims may instead be disparaged by particularly competitive observers who value or covet the criterion achieved by the target. If two individuals are in competition for a single resource (for example, a prestigious job offer), then the person who loses the competition may not praise the boastful (but accurate) winner as high in competence as our theory would predict.

Self-enhancement as a strategic choice

If social agents must decide between self-enhancement and self-effacement, how might they evaluate their anticipated social impressions? The morality effect suggests

caution, clearly identifying humility as a safe bet. However, being proven right after claiming to be better than others creates an enticing outcome: a reputation of high competence. This decision can be initially evaluated according to the accuracy correlation in the population between self-judgments and performance. If this correlation is high (as in trivia performance estimation; Section I), betting on a positive comparative self-evaluation is a good one. In cases where it is more difficult to predict performance or standing, however, the safer choice may be to claim self-inferiority. However, this strategic framework suffers from a paradox of its own, where valuable and evaluative domains are often the areas in which self-judgment accuracy is the lowest (Vazire, 2010). The simplest conclusion to draw here is that in matters of low consequence (and high accuracy), individuals should feel free to claim superiority, as this will increase perceptions of their competence. If these individuals are proven wrong by objective evidence, then perhaps the sting of committing a self-enhancement error will be assuaged by the relative unimportance of the task performance. In sum, it may not be so easy to strategically manipulate when and when not to self-enhance, after all.

Upon zooming out to explore self-enhancement bias at the population level, accuracy remains a necessary concern. What happens, for example, if a large number of individuals attempt to leverage known accuracy in the population? Knowing that accuracy is high, many individuals will be encouraged to self-enhance. If all (or none) of the population chooses to self-enhance, the accuracy correlation remains unchanged. If some, but not all, choose to self-enhance, however, this correlation weakens.[9] To summarize, choosing to self-enhance because accuracy is known to be high may undermine the very accuracy used to justify the original self-enhancement. This outcome begins to approach a process of magical

---

[9] Tested and confirmed with a computer simulation.

thinking where individuals change their thinking or behavior because of a correlational relationship between this thinking or behavior and some positive outcome (Mijović-Prelec & Prelec, 2010). This type of decision-making is irrational by means of generalizing a correlational effect to a causal outcome in one's own experience. The dangerous logical consequence here is that individuals may begin to conclude that by thinking themselves to be better than others, it follows that they will become better than others. As previously mentioned, this type of thinking can have a detrimental effect on population-level accuracy. Indeed, those who self-enhance because they believe a positive self-view increases the chance of obtaining a positive self are engaging in self-deception (von Hippel & Trivers, 2011). Compared to these individuals, those who self-enhance in order to improve or protect their reputation can be argued to behave rationally. Research on impression management suggests that this type of strategic self-presentation is common (Paulhus, 1984).

Finally, the decision to self-enhance can be cast as a social dilemma (Dawes, 1980; Van Lange, Joireman, Parks, & Van Dijk, 2013). What happens when those who face such a dilemma know their outcome depends on the choice made by others? Individuals should seek to ensure that not everyone chooses to self-enhance, but that each individuals is given an opportunity to uniquely self-enhance so that the payoffs can be maximized for each individual in the group. Conversely, a pure egocentrist would focus simply on his own reputational payoffs while paying little mind to the decisions of others or their long-term consequences.

This example has some similarities with the theoretical argument that giving and receiving esteem within a group can resemble the Prisoner's Dilemma (Krueger, Vohs, & Baumeister, 2008). Esteem is maximized if all choose to cooperate, but individual decision-

70

makers are faced with the temptation to defect, thus free-riding on their stockpiled esteem and the esteem shared by others. This model is slightly different for the decision to self-enhance, however. Here, if all members choose to self-enhance, conflict ensues and costly evidence may be demanded. If both say nothing, there is no conflict. Each individual, however, is faced with the temptation to self-enhance when the others do not, as this creates a large esteem boost for the self relative to the other players. As such, deciding to when and where to self-enhance should take into account what others are likely to do, and whether another is likely to challenge the self-enhancing assertion.

When self-enhancement is cast as defection, new avenues appear for research on its adaptiveness or harmfulness for both the self and for others. Whereas self-esteem is often generated by interpersonal affiliation and social approval (MacDonald, Saltzman, & Leary, 2003), defecting in a comparative self-enhancement scenario by claiming to be better than others may garner self-esteem without validation or approval by others. As past work has shown, however, the immediate benefits of self-enhancement are inevitably overcome by the reputational damage sustained by self-inflation and narcissism (Paulhus, 1998; Robins and Beer, 2001). This appears to be especially true when self-enhancing claims are made explicitly and in direct comparison with others (Hoorens et al., 2012; Van Damme, Hoorens, & Sedikides, 2015). If social agents refuse to consider longer-term reputational consequences in favor of immediate boosts to self-worth, it can be argued that these individuals suffer from irrational, myopic self-presentations (Moore & Kim, 2003).

Section Conclusion

In two studies, social perceivers were shown to be sensitive to the distinction between self-enhancement bias and error, particularly in the domain of competence. In the

moral domain, self-enhancers were disparaged regardless of their performance. Treating

self-enhancement as a decision-strategy cast according to the decision-theoretic framework

allows for new questions to be asked of the adaptiveness or detriments of self-enhancement

bias and error, and helps to provide a window into why individuals may be motivated to

self-enhance in certain contexts while remaining humble in others. Section II serves as a

bridge between Section I (measurement) and Section III (motivation) in self-enhancement

bias and error by identifying performance, accuracy, and self-perceptions as valenced

domains detectable by both decision theory and lay social perceivers. In Section III, I

proceed by asking questions of self-enhancement motivation in terms of both bias and error.

Knowing that self-enhancement error is a particularly negative category to occupy, how do

motivated individuals subjectively mitigate the aversiveness of this risk and negative

outcomes?

.

Section III: Motivated Reasoning in Self-Enhancement Bias and Error

In Section I I argued for the prevalence of accuracy in comparative self-judgment on a performance measure using the decision-theoretic approach to self-enhancement bias and error. In Section II, it became clear that lay social perceivers were sensitive to the distinction between accurate and inaccurate self-enhancement, praising the former and disparaging the latter. Little has been said, however, about the role of motivation in self-enhancement bias and error. Section III asks the question of and how and when individuals may distort perceptions of themselves or their environment in self-serving ways. Thus, the goal of Section III is to explore motivated reasoning that serves to enhance or protect the self-image from within the decision-theoretic framework. A brief introduction to motivated self-enhancement precedes three studies which aim to demonstrate two novel, motivated strategies utilized by individuals to 1.) maximize the perceived likelihood of achieving a favorable outcome or 2.) ameliorate the aversiveness of receiving accurate negative feedback.

Motivated Self-Enhancement

In important or desirable domains, holding an inflated self-perception can be psychologically beneficial. The first evidence for the adaptiveness of so-called positive illusions involved "unrealistically positive self-evaluations, exaggerated perceptions of control or mastery, and unrealistic optimism" (Taylor & Brown, 1988, p. 193). Taylor and Brown argued that inaccurate, positive biases in self-judgment were useful and adaptive because they *feel good*. Perceiving oneself to be better than others should generate feelings of efficacy, competence, self-esteem, and adjustment, all of which contribute to an individual's perception of self-worth (Crocker & Wolfe, 2001). When

73

reasoning within an inherently subjective or opaque criterion domain (i.e., morality, friendship quality, etc…), individuals have little incentive to suppress self-inflating claims and perceptions. In other words, if it feels good and is unlikely to be proven wrong, there may be some intrinsic value in overclaiming positive personality traits and outcomes. Recently, this was demonstrated in the domain of morality, where individuals' moral self-perceptions were found to exceed the normative level of morality in the population as determined by a collection of self- and other-estimates (Tappin & McKay, 2016).

A variety of studies have emerged from the Taylor and Brown hypothesis, arguing that manipulating individuals' motivations has a causal effect on self-enhancing behavior. Brown (2011) demonstrated that the perceived importance of a trait influenced how likely individuals were to claim possession of it. Here, the BTAE was significantly stronger for traits rated as important (honest, kind, responsible) than for traits rated as less important (outgoing, imaginative). This is in line with Pelham's (1991) treatment of trait importance, which he calls an "emotive investment in the self" (p. 520). Brown and Han (2012) manipulated self-worth by having participants either self-evaluate or receive negative feedback from a computer program after completing an intelligence task. When self-worth was threatened, participants were eager to restore positive feelings of self by rating themselves as above average on important traits moreso than those who did not experience threat. A later extension of this study showed that in the same self-threatening paradigm, participants rated a close friend or relationship partner on a series of positive traits after self-evaluating or receiving negative feedback. Those with low self-esteem in the experimental condition self- and partner-enhanced by rating the self and a close other

higher on positive traits than did those in a control group. Brown summarized his and the larger motivational approach to self-enhancement succinctly: "People believe they are better than others largely because it makes them feel good to do so" (Brown, 2012, p. 217).

These accounts aimed to reestablish the link between self-esteem and motivated self-perception, but they also clearly demonstrated that self-enhancement bias can be caused by motivation. Although I have so far treated self-enhancement primarily as a byproduct of cognition and information processing systems, work in this area presents evidence that self-enhancement bias cannot be attributed to strictly cognitive causes alone. It is unclear whether this is similarly true for self-enhancement error. Non-comparative self-enhancement also appears to be sensitive to motivated reasoning. Dunning, Leuenberger, and Sherman (1995) demonstrated that ego-threatened participants were more likely to provide egocentric than global definitions of what it means to succeed. Participants were less likely to recall self-relevant behaviors when they were negative and highly diagnostic of the target trait than when the same behaviors were negative and low in diagnosticity (Green & Sedikides, 2004). For example, participants were more likely to recall a minor behavioral indicator of trustworthiness (e.g., "I would use the toothpaste of a roommate without asking") than a more diagnostic one, (e.g., "I would be unfaithful in an intimate relationship") (Green & Sedikides, 2004, p. 79). Similarly, participants were more likely to disregard failure feedback on their own task performance than on a random other person's (Guenther & Alicke, 2008). Epley and Gilovich (2016) decompose motivated reasoning into two primary information processes: evidence recruiting and evidence evaluation. For the first process, a motivated individual

can selectively attend to self-positive information, or simply choose to ignore negative (or all) information. After being exposed to self-relevant information, however, one can only alter how that information is evaluated.

A separate but related set of motivated self-enhancement processes can be categorized as self-protection. Self-protection processes often occur proactively, or as a response to a previous negative experience (Alicke & Sedikides, 2009). This type of processes aims to mitigate future threats before a self-threat is even encountered. Whereas traditional self-enhancement typically serves to promote the self in the present or in hindsight, self-protective mechanisms attempt to avoid or diminish future experiences of social rejection or poor performance. One explanation for the prevalence of self-protective mechanisms is that negative events often loom larger than positive ones (Baumeister, Bratslavsky, Finkenaur, and Vohs, 2001). If potential negative future events are weighted more heavily than positive ones, cognitive resources can be efficiently devoted to protecting the self rather than to repairing it.

Strategies and Hypotheses

In Section III, I propose that individuals may engage in motivated, self-protective (or –enhancing) reasoning within the decision-theoretic framework. Specifically, I argue that motivated processes can occur even after individuals have completed an objective task and provided self- and other-estimates. Because individuals can no longer manipulate the most directly self-relevant information (their performance or estimates), motivated reasoning must arise through peripheral, post-hoc assessments of the accuracy or prevalence of their comparative self-judgments. Such individuals, already categorized according to the framework, may strategically manipulate how certain or important their

self-judgment is to them, or inflate their estimates of others who performed as they themselves did. In three experiments, this section aims to explore a.) two possible strategies that individuals may employ after being 'locked in' to their performance and comparative estimates and b.) how several common individual-difference measures of self-enhancing behavior associate with or predict these strategies. If motivated, self-protecting or -enhancing mechanisms can be detected after individuals have completed an objective task and generated their own comparative estimates, then the validity of the decision-theoretic approach to self-enhancement bias and error will increase substantially as it becomes possible to capture self-enhancement error *and* post-hoc motivated reasoning expressed by the same individual using a parsimonious and face-valid tool.

Favorable uncertainty. One strategy individuals may employ to protect the self from a negative outcome involves expressing a variable amount of certainty in the accuracy of their comparative self-assessment. By varying how certain an individual is that she truly performed better (or worse) than average, this individual can manipulate her perceived likelihood of achieving a desirable outcome even after providing objective, immutable performance estimates. According to the decision-theoretic framework, there are two possible favorable outcomes: performing above average and making an accurate comparative self-judgment (see Section II for evidence from a reputational perspective). Conversely, performing below average and expressing inaccurate self-perceptions were both associated with lower competence (argued in Section I and demonstrated in Section II), which can reasonably be labeled as unfavorable outcomes. This crossing of performance and self-judgment accuracy is the basis for the motivational strategy I hereafter refer to as Favorable Uncertainty.

The Favorable Uncertainty strategy posits that individuals may express low confidence or certainty in their self-judgments when learning that this judgment was wrong can lead to a favorable outcome. This is a necessary scenario for those who claim to be worse than average on a desirable task. By making this claim, these individuals forego the possible benefit of performing well for the likely benefit of accurately identifying their poor performance, a strategy familiar to research on self-handicapping and strategic self-presentation (Baumeister, Tice, & Hutton, 1989; Tice & Baumeister, 1990). The motivational strategy presented in this section can be measured specifically when considering how *certain* individuals are of a comparative claim. Thus, Favorable Uncertainty predicts that those who claim to be worse than average should be less certain that this claim will turn out to be true than those who claim to be better than average. Conversely, those who claim to be better than average are betting that their claim will be proven right, and are thus motivated to convince themselves of their own accuracy. As discussed in Section II, it would be strange for an individual who expresses a better-than-average belief to be doubtful of its accuracy. It is reasonable, however, for a self-effacer to make a humble claim while believing (or hoping) that that claim does not represent reality.

Misery loves company. A second motivational strategy individuals may use is to manipulate how they perceive their decision category. To mitigate the negativity of discovering that a person committed a judgment error, that individual may adopt an ameliorative perception of that category, namely, by overestimating the number of community or ingroup members that commit a similar error. This strategy, hereafter referred to as a Misery Loves Company (MLC) effect, posits that individuals' egocentric

tendency to project to others will increase with the negativity of the category that they inhabit. Schachter's (1959) famous studies on affiliation demonstrated that threat can cause an affiliation response, an effect later argued to be due specifically to specific psychological mechanisms including sadness and social regret rather than conformity (Cooper & Rege, 2011; Gray, Ishii, & Ambady, 2011). Thus, evidence has shown precedence for a motivated, ameliorative response to negative feedback.

The proposed application of a MLC effect to the decision-theoretic approach to self-enhancement bias and error starts by assuming social projection. When individuals learn which category of decision they made, their estimates of how many others will make a similar decision should be reasonably projective (Epley, Converse, Delbosc, Monteleone, & Cacioppo, 2009; Robbins & Krueger, 2005). The motivated MLC hypothesis extends specifically to those who committed a self-enhancement error. These individuals, who learn that they committed the worst possible decision, should estimate that a greater number of others will also commit False Alarms than a.) the actual number of False Alarms in the population, and b.) any other category member's projective estimates of their own category. Put simply, feeling like others have committed the same error may be one motivational strategy that self-enhancers use to diminish the aversiveness of learning that they both performed poorly and made an inaccurate judgment.

To summarize, the first goal of the studies presented in this section is to detail two possible nonexhaustive motivated strategies that individuals may adopt after having made a comparative self-judgment. A Favorable Uncertainty strategy may be in use when those who claim to be worse than average demonstrate lower certainty in this judgment than

79

those who claim to be better than average. A Misery Loves Company strategy occurs when those who find themselves committing the worst type of decision error – a False Alarm – overestimate the number of individuals who make the same error in order to mitigate the aversiveness of this realization.

Individual Differences: Predicting Self-Enhancement Bias, Error, and Motivation

The second goal of this section is to introduce several individual differences measures, often argued to be associated with self-enhancement tendencies, into the decision-theoretic approach to self-enhancement bias and error. These measures may mediate or moderate the effects of the two strategy types proposed above on self-judgment, self-enhancement, or actual task performance. They may also be able to discriminate between accurate and inaccurate self-enhancers, which would be a novel finding. By including such measures in a predictive model of both judgment category and task performance, we can conclude with greater certainty whether motivational strategies and tendencies can uniquely predict task performance beyond the robust effect of simple self-judgment.

Self-esteem. Perhaps the most common individual difference measure thought to be associated with self-enhancement is self-esteem. This is the gold standard for measuring the 'benefits' of positive illusions (Taylor & Brown, 1988). In Study 2, False Alarm participants were found to have higher self-esteem than Hits. In Study 7, participants complete a validated measure of trait self-esteem (Rosenberg, 1965) to determine whether this classic personality trait is over or underrepresented in the False Alarm category, and how it may be related to individuals' motivated strategic attempts to main a positive self-image.

Task desirability. Alicke (1985) demonstrated that trait desirability had a strong and robust moderating effect on individuals' tendency to self-enhance relative to others. Since then, individuals have consistently been shown to claim that they possess more desirable (and fewer undesirable) traits than their conception of 'the average person' (Alicke & Sedikides, 2011). Study 7 includes a measure of trivia knowledge desirability, expecting that those who value trivia knowledge will perform better, judge themselves to be better, and possibly to engage in more motivated reasoning than those who do not value trivia knowledge.

Desirable responding. Other work has been conducted on self-enhancement as a form of strategic self-presentation. The Behavioral Inventory of Desirable Responding (BIDR) has emerged from this work, comprising two subscales (Paulhus, 1988). The first, self-deceptive enhancement, purports to capture genuine self-enhancement tendencies and includes items that measure the extent to which individuals deceive themselves in favorable ways (example item: "*I am a completely rational person*"). The second subscale, impression management, attempts to capture individuals' propensity to answer questions in a socially desirable way, even if they disagree with the content (example item: "*When I hear people talking privately, I avoid listening*"). Each subscale is argued to correlate with self-enhancement behavior that makes agents feel good and shapes others' opinions of them in a favorable way.

Non-pathological narcissism. Typically, those who are thought to highly value the self-image will go to greater lengths to enhance or protect it, even if doing so requires distorting reality (Alicke & Sedikides, 2011). Unsurprisingly, this measure has previously been shown to correlate with self-enhancement (Paulhus & John, 1998; Paulhus, Harms,

Bruce, & Lysy, 2003). Like self-enhancement, the adaptive nature of narcissism is often debated. Despite the complex and unclear relationship between the two, a measure of narcissism may provide insight into individuals' motivated responding to questions asked about their performance on a simple task.

By determining whether strategic, motivated responding occurs within the context of the decision-theoretic approach to self-enhancement bias and error, we gain a clearer picture of how individuals' motivated preferences and tendencies to protect the self-image occur despite objective evidence of their performance on a task thought to be low in desirability (trivia knowledge). Similarly, by introducing individual difference measures known to be associated with self-enhancement bias, I ask whether these measures can differentiate between bias and self-enhancement error. Study 5 seeks to demonstrate a Favorable Uncertainty pattern. Study 6 aims to replicate this pattern and provide evidence for a Misery Loves Company effect unique to those who commit a self-enhancement error. Finally, Study 7 attempts to overcome some of the limitations of Studies 5 and 6 by obtaining a larger sample, clarifying the dependent variables, and introducing exploratory individual difference measures.

Study 5: Favorable Uncertainty

Method

Participants. Brown University undergraduates (N = 162; 80 female) were recruited by student experimenters as part of an assignment for a laboratory course in social cognition. Participation in the experiment was voluntary and participants were not compensated for their time. Although age information was not collected, most participants were juniors and seniors in college (mean year = 3.42, SD = .86).

Materials. A 30 item trivia task was adapted from Moore and Small (2007). The task contained items varying in difficulty and spanning six domains of trivia knowledge: pop culture, history, science, geography, music, and sports. Each item was presented with a dropdown menu containing the correct answer and three additional foils. The trivia task was presented using Qualtrics online survey platform (Qualtrics, 2014). Item and answer order was randomized for each participant.

Procedure. Student experimenters presented volunteering participants with a laptop computer displaying the survey form. Participants were asked to complete the experiment in a single session and to refrain from using any outside sources or materials to answer the questions. Student experimenters remained in the room with each participant for the duration of the experiment. Participants provided informed consent and then completed the trivia task. After this task was completed, participants were asked to report both how many questions out of 30 they themselves and the average Brown University student answered correctly. These questions were presented in counterbalanced order.

Participants were then shown their answers to the estimation questions described above in the following manner: "*You estimated that you answered [self-estimate] out of 30 questions correctly. You estimated that the average Brown University student answered [other-estimate] out of 30 questions correctly.*" Participants were then prompted to indicate whether these estimates indicated that they thought to have performed better than the average Brown University student, or worse than (or equal to) the average Brown University student. Depending on this categorization, participants then completed a certainty measure by estimating the probability that their categorization

was accurate (0 to 100). For example, if a participant claimed to correctly answer more questions than the average Brown student, they were shown the following prompt: "*You estimated that your performance was better than that of the average person. However, we know that these estimates can always contain some error. Using the slider below, rate (on a scale from 0 to 100) the probability that you actually performed better than the average person.*" Participants also rated how content they were with their life and how content they thought the average Brown student to be with their life or a 7-point scale ranging from 'Very Discontent' to 'Very Content.' Finally, participants were debriefed and asked to exit the survey.

Results

Nine participants failed to accurately interpret the self- and other-estimates they provided. These participants misinterpreted their estimates, claiming to be above (or below) average when the numbers they provided indicated the opposite. After excluding these participants, a final sample size of $N = 153$ remained for analysis.

Descriptive analyses and categorization. Seventy-two participants claimed to have scored above average. Of these 72, 47 (65.28%) were accurate in this claim (H) (25 FA's; 34.72%). Eighty-one participants claimed to have scored below (or equal to) average, and 53 (65.43%) were accurate (CR) (28 M's; 34.57%). See Table 5 for variable descriptives for each of these four target categories. The relative prevalence of the two accurate category types suggested that participants had a good sense of their performance on the task. Indeed, the accuracy correlation over all participants was moderate, $r_{S,T}(151) = .42$, though there was a substantial difference in this correlation between those who claimed to be above average ($r_{S,T}(70) = .58$) and those who claimed to be below average

$(r_{S,T}(79) = .10)$. As expected, participants claimed to be more content with their own life

$(M = 5.68, SD = 1.10)$ than their estimation of the average Brown student's contentedness

with their life $(M = 4.89, SD = 1.09)$, $t(152) = 7.52$, $p < .001$, $d = .72$.

Table 5

Descriptive Statistics for Study 5(Favorable Uncertainty)

n	$T > T_{median}$	$T \leq T_{median}$	
S > O	$n = 47$	$n = 25$	$n = 72$
S ≤ O	$n = 28$	$n = 53$	$n = 81$
	$n = 75$	$n = 78$	

T	$T > T_{median}$	$T \leq T_{median}$	
S > O	23.23 (2.05)	17.08 (4.32)	21.1 (4.21)
S ≤ O	22.14 (1.43)	17.85 (2.18)	19.33 (2.83)
	22.83 (1.91)	17.6 (3.03)	

S	$T > T_{median}$	$T \leq T_{median}$	
S > O	21.57 (3.64)	17.32 (5.11)	20.1 (4.65)
S ≤ O	14.71 (4.55)	13.72 (5.38)	14.06 (5.1)
	19.01 (5.19)	14.87 (5.53)	

O	$T > T_{median}$	$T \leq T_{median}$	
S > O	16.98 (3.8)	13.68 (5.25)	15.83 (4.6)
S ≤ O	16.82 (4.56)	16.06 (4.72)	16.32 (4.65)
	16.92 (4.07)	15.29 (4.99)	

Certainty	$T > T_{median}$	$T \leq T_{median}$	
S > O	70.09 (18.31)	61.6 (15.35)	67.14 (17.71)
S ≤ O	62.18 (18.05)	57.51 (17.75)	59.12 (17.88)
	67.13 (18.5)	58.82 (17.02)	

Note. Cell and marginal means are displayed with standard deviations in parentheses.

Favorable uncertainty. The Favorable Uncertainty hypothesis can be tested

using the novel certainty measure, reported by participants as the probability that their

claim to be above (or below/equal to) average would be validated by their test score. To

ensure that this probability measure behaved lawfully, and as an indicator of whether

participants understood it, certainty was regressed on participants' S – O scores (see

Figure 8). Because self-judgments and performance were positively correlated, extreme

comparative judgments (whether positive or negative) should elicit greater certainty in

the comparative claim's accuracy. This regression approach yielded a significant

quadratic trend in the predicted direction, $F(2,150) = 21.30$, $b_{quadratic} = .295$, $p < .001$,

such that extreme positive and negative S – O scores were associate with greater

certainty[10]. For the focal test of the Favorable Uncertainty effect, average certainty

ratings were compared between above- and below-average claiming participants. Indeed,

those who estimated that they performed worse than (or equal to) average ($N = 81$) were

less certain of this judgment ($M = 59.12$, SD $= 17.88$) than those who estimated that they

performed better than average ($N = 72$) ($M = 67.14$, SD $= 17.71$), $t(151) = 2.78$, $p < .003$,

$d = .45$.

This result was the first piece of evidence for a Favorable Uncertainty effect in the

context of own and others' performance estimation on an objective task. It remains

unclear, however, if self-enhancing or self-effacing certainty estimates differ from the

accuracy rates observed in the sample. Because certainty estimates were measured on a

probability scale (0 to 100), it becomes possible to compare mean scores with the

observed accuracy proportions. This approach is merely illustrative, however, as

---

[10] A linear trend also emerged, $b = .601$, $F(2,151) = 14.32$, $p < .001$, which may help to explain the mean
level differences in the focal hypothesis test.

statistical inference is inappropriate for this type of comparison. Of those participants

who claimed to be better than average ($M_{certainty} = 67.14$), 47 (65.28%) indeed exceeded

the average total score on the trivia task. Of those who claimed to be worse than (or equal

to) average ($M_{certainty} = 59.12$), 53 (65.43%) were correct. These results suggest that self-

effacing individuals were *under*certain in the accuracy of their claim, even though the

observed prediction accuracy was much the same for the two categories (Misses and

Correct Rejections). Conversely, probability estimates for those who claimed to be better

than average nearly matched the accuracy proportion, suggesting that self-enhancers were
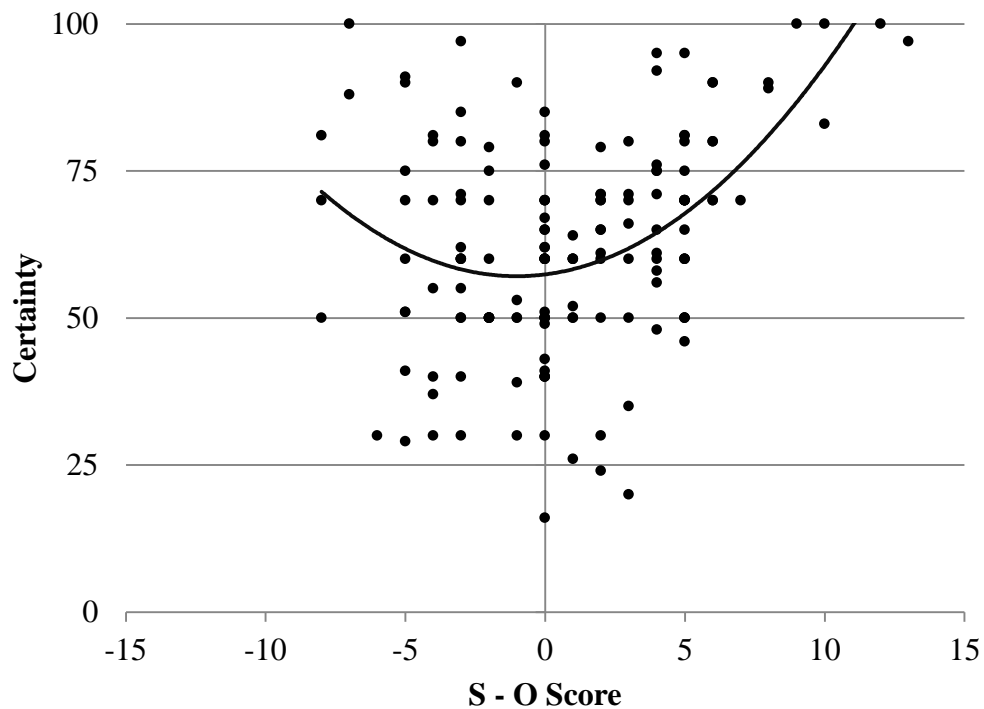
appropriately certain in their comparative claims.

Figure 8. Scatterplot displaying certainty varying according to participants' S (self-judgment) minus O (other-judgment) score.

To further understand how certainty contributes to self-enhancement bias and error, it is possible to test whether certainty differs between accurate and inaccurate category types. When considering only those who claimed to be above average (H, FA), Hits reported greater certainty ($M = 70.09$, $SD = 18.31$) than False Alarms ($M = 61.6$, $SD = 15.35$), $t(70) = 1.98$, $p < .03$). There was no such difference between the two types of self-effacers, Misses ($M = 62.18$, $SD = 18.05$) and Correct Rejections ($M = 57.51$, $SD = 17.75$). A more sophisticated approach asks whether certainty can predict False Alarms from among those who claimed to be above average. To do so, the two category types contained within this group were assigned dummy code variables (Hit = 0, False Alarm = 1). Certainty correlated negatively with this dummy code measure, $r(70) = -.230$, $p < .05$, suggesting that participants' subjective measure of certainty in their above average claim could differentiate between those who were accurate (Hits) and inaccurate (False Alarms).

Study 5 Discussion

To summarize, participants were shown to be uncertain in their comparative self-judgments when claiming to be worse than (or equal to) average, compared with those who claimed to be better than average. This result supported the proposed motivational hypothesis by demonstrating that participants were less certain when uncertainty would suggest the possibility of achieving a favorable outcome: being wrong (and thus, performing well on the task). For those who claimed to be above average, certainty appeared to track accuracy and was able to differentiate between Hit and False Alarm categories of judgment. However, the results of Study 5 remain preliminary. Specifically, it was difficult to interpret the certainty measure as varying from 0 to 100. It remains

unclear what any probability rating of 49 or below should indicate: do participants believe the *opposite* of their claim will occur? Or are participants simply treating this probability measure as a continuous (low to high) scale of certainty? Study 6 attempts to replicate the Favorable Uncertainty effect with a clearer measure of certainty, and introduces the second motivational strategy theorized in this section: Misery Loves Company.

Study 6: Misery Loves Company

Having found evidence for the first proposed motivated self-enhancement strategy, Study 6 had two primary goals. The first was to replicate the Favorable Uncertainty effect demonstrated in Study 5 using a more interpretable measure. Specifically, the 0 to 100 certainty measure was restricted to range from 50 (maximally uncertain) to 100 (maximally certain) to better represent a probability estimate. This restriction would prevent participants from choosing a low probability, thus reversing their original comparative claim. The second goal was to determine whether individuals who committed a self-enhancement error (False Alarm) would overestimate the number of others in the population who commit a similar error. The so-called Misery Loves Company effect is another motivational strategy that individuals may use to protect or enhance the self-image in the presence of negative outcomes.

Method

As in Study 5, Brown University students (N = 129; 71 female; M_{age} = 20.56) were recruited by student experimenters. Two subjects were excluded for failing to correctly interpret their numerical estimates and were excluded from the sample, leaving 127 participants for analysis.

Materials and procedure. Performance was measured using the same 30-item trivia task presented in Study 6 (Moore & Small, 2007) again presented electronically via Qualtrics survey platform. Item and answer order was randomized for each participant.

Participants were asked to complete the task sitting in a quiet room with a laptop computer displaying the survey. Student experimenters remained in the room with each participant for the duration of the experiment. After providing informed consent, participants completed the trivia task and were then asked to provide their self-estimate and estimate for the average Brown University student in counterbalanced order. Similarly to the procedure in Study 5, participants were then fed back their answers to the estimation questions and asked to affirm whether these numbers indicated an above or below/equal to average claim. Depending on whether participants claimed to perform above average or below (or equal) to average, they were then presented with the new certainty measure asking them to estimate "The probability that I actually performed better (worse than or equal to) the average person" on a scale from 50 (maximally uncertain) to 100 (maximally certain). Descriptive text clarified that choosing 50 would equate to a 50% chance, the same probability as either outcome of a coin flip.

After completing the certainty estimate, participants were given accurate feedback about their performance. Specifically, they were told how many questions they answered correctly and whether they performed above or below (equal to) average. The feedback average was computed from the sample of Brown University students who completed the same task in Study 5. All four decision categories were described with participant's decision category bolded for emphasis. Participants received the feedback, *"Based on the answers you provided, you are an example of Category [1, 2, 3, or 4]."* Following this

90

feedback, participants were asked to imagine 100 other randomly sampled Brown

students who also took the trivia quiz and to estimate how many of these 100 students fell

into each of the four possible decision categories. All four categories were listed with an

entry box next to each. As participants entered their estimates, the estimate total updated

in a separate text box. Participants were instructed that their estimates must sum to 100,

and told that this task may take some deliberation to complete properly. To remind

participants which category they belong to, parenthetical text stated *"you are here"* next

to their category. Figure 9 displays the full extent and wording of this feedback and

population estimation measure for a participant who wrongly claimed to be above

average.

Your score on the trivia task was ${gr://SC_cHEKIxjCiirkgaV/Score} out of 30. The average score on this trivia task is 20.2 out of 30. **This indicates that you actually performed worse than the average Brown student.**

In our research, we have identified four categories of people, those who:

1. Claim to be better than average and actually perform better than average.
2. **Claim to be better than average but actually perform worse than average.**
3. Claim to be worse than average but actually perform better than average.
4. Claim to be worse than average and actually perform worse than average.

Based on the answers you provided, **you are an example of Category #2.**

We would like you to imagine 100 randomly sampled Brown students who took this quiz. Please estimate how many of these 100 students fall into each of the four categories. Your answers must sum to 100.

Note that this task may take some deliberation. One possible strategy is to identify which category contains the most (or fewest) students first and then fill in the remaining categories. You may edit the numbers until you are happy with your estimates.

	Number of Brown students (out of 100)
1. Claim to be better than average and actually perform better than average.	
2. Claim to be better than average but actually perform worse than average (you are here).	
3. Claim to be worse than average but actually perform better than average.	
4. Claim to be worse than average and actually perform worse than average.	
Total	

Figure 9. Example instructions given to participants during the population estimate task.

Finally, participants rated their own and the average Brown student's overall life contentedness, completed several demographic measures (trivia desirability, age, gender, ethnicity, country of birth, and whether they used any outside materials to complete the trivia task) before being thanked and debriefed.

Results

Descriptive analyses and categorization. Categorization yielded 29 participants who claimed to be better than average (16 Hits, 13 False Alarms) and 98 participants who claimed to be worse than (or equal to) average (30 Misses, 68 Correct Rejections). This strong positive skew was unexpected and should be noted as a major limitation of Study 6. The now familiar prevalence of accurate category types suggested that this sample demonstrated accuracy in their comparative self-judgments. Indeed, the accuracy correlation over all participants was moderate, $r_{S,T}(125) = .46$. In line with previous studies, participants rated themselves as more content with their lives ($M = 5.78$, SD = 1.18) than their estimate of the average Brown University student ($M = 5.21$, SD = 1.09), $t(126) = 5.27$, $p < .001$, $d = .50$.

Favorable uncertainty. Because of the strong positive skew in better-than-average claims, certainty scores were not suited for a regression approach attempting to replicate the quadratic trend observed in Study 5. It was still appropriate, however, to conduct a mean-level test. The focal test of the Favorable Uncertainty hypothesis was to compare certainty ratings, now ranging from 50 to 100, between those who claimed to be better than average and those who claimed to be worse than (or equal to) average. Participants who claimed to be below (or equal to) average did not differ in their reported certainty ($M = 73.40$, $SD = 13.58$) from those who claimed to be above average ($M = 70.97$, $SD = 12.68$), $t(125) = .86$, $p < .20$. Because the critical difference was not observed, and due to the restrictively small sample of participants who claimed to be above average, tests seeking to predict self-enhancement bias and error were omitted. It was unclear whether the failure to replicate a favorable uncertainty effect was due to the

modified certainty scale, the widely varying sample size in comparative self-judgments, or a true null effect. A second replication attempt is presented in Study 7. For now, we turn to analysis of the population estimation measure.

Misery loves company. Because only 29 participants claimed to be above-average, and only 13 made the target decision error (False Alarm), analytic tests of the Misery Loves Company hypothesis should be considered exploratory and nonconfirmatory. As such, the following results should be replicated before conclusions can be drawn. Despite these limitations, however, initial evidence for a Misery Loves Company effect was encouraging specifically due to large observed effect sizes.

To answer the simplest question posed by the Misery Loves Company hypothesis, averages were computed for the number of hypothetical individuals estimated to make the same type of judgment as participants in each category type (Hits estimating Hits, False Alarms estimating False Alarms, etc…). A one-way between-subjects ANOVA on these judgments revealed differences in own-category estimations between category types, $F(3, 123) = 7.99$, $p < .001$, $\eta^2 = .163$ (see Figure 10). Post-hoc tests using Tukey's HSD procedure confirmed that participants who committed a False Alarm estimated a greater number of False Alarms in the population ($M = 41.54$, $SD = 14.91$) than Hits estimating Hits ($M = 29.81$, $SD = 15.23$), $p < .039$, Misses estimating Misses ($M = 28.07$, $SD = 13.76$) , $p < .006$, and Correct Rejections estimating Correct Rejections ($M = 22.94$, $SD = 11.60$) , $p < .001$. Importantly, and in support of the MLC hypothesis, no other differences between groups were significant. Even for the weakest comparison between False Alarm and Hit participants, the difference in estimated self-enhancement error in the population had large effect size (Cohen's $d = .97$). From this initial assessment it was

clear that participants who wrongly claimed to be better than average uniquely inflated

the number of hypothetical others who would find themselves in the same category.

However, this encouraging initial assessment is insufficient evidence that False Alarm

targets viewed their own category type as uniquely common. A proper analysis must

address how the four category types of participants viewed their own *and* others'

categories (see Figure 11 for a full display of means).

Figure 10. Mean estimates of number of individuals who would make the same category of decision, grouped by participant decision category. Error bars represent one standard error of the mean.
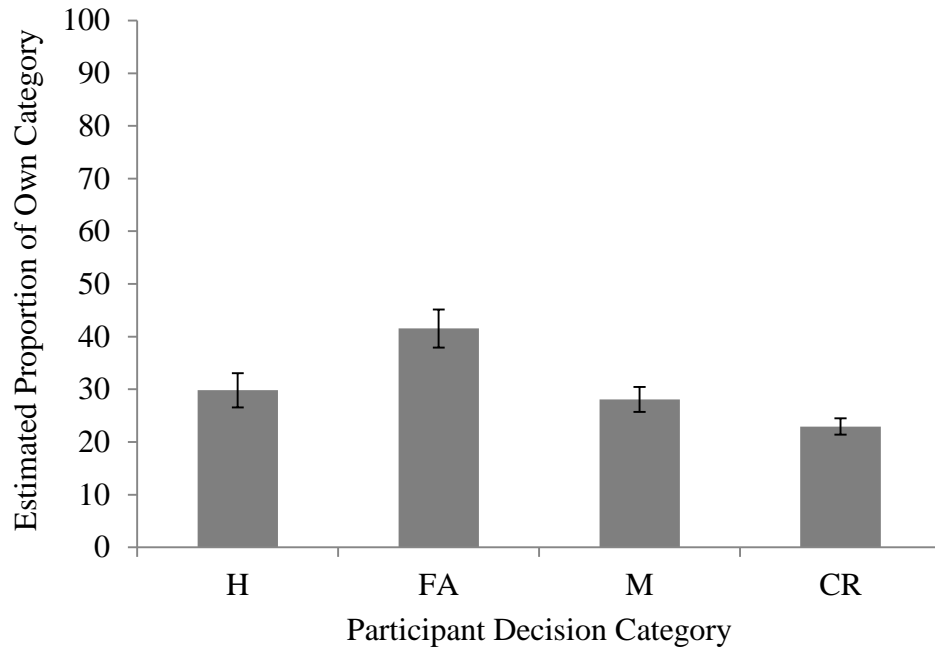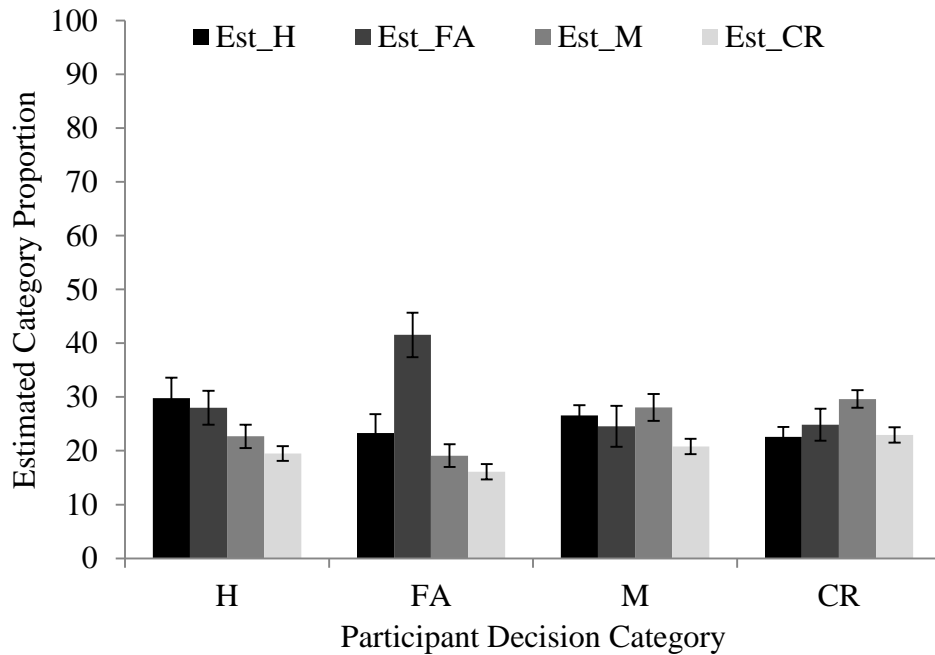

Figure 11. Mean estimates of number of individuals who would make each of the four categories of decision, grouped by participant decision category. Answers must sum to 100 within each participant type. Error bars represent one standard error of the mean.

The focal prediction made by the Misery Loves Company hypothesis is that

inaccurate self-enhancers (FA participants) will inflate the number of estimated others in

their own category compared specifically to those who correctly claim to be above

average (H). If committing a self-enhancement error motivates individuals to

overestimate those committing a similar error, then False Alarm participants should

estimate that more above-average claims will turn out to be false than Hits do. This can

be formalized as a two-way interaction effect whereby False Alarm participants estimate

a larger difference between False Alarms and Hits than do Hit participants. This results in

a 2 (Accuracy: accurate (H) / inaccurate (FA)) * 2 (Estimate Category: own category /

other category) repeated measures ANOVA with Accuracy as a between-subjects factor

(See Figure 12, top panel). Though peripheral to the primary hypothesis, the main effect

of Estimate Category was significant, $F(1, 27) = 5.11$, $p < .032$, $\eta^2 = .15$, suggesting that

participants tended to estimate greater numbers of their own category types than the

opposite category type, consistent with social projection. The inconsequential main effect

of Accuracy was not significant, $F(1, 27) = 1.57$, $p < .221$. Critically, the Accuracy *

Estimate Category interaction approached significance, $F(1, 27) = 3.43$, $p < .075$,

yielding the expected pattern. Simple effects testing confirmed that False Alarm

participants estimated a greater number of False Alarms ($M = 41.54$) than Hits ($M =

23.31$), $t(12) = 3.42$, $p < .005$, $d = .97$. Importantly, Hit participants did not differ in their

number of estimated False Alarms ($M = 28.00$) and Hits ($M = 29.81$), $t(15) = .270$, $p <

.790$. This pattern provided initial evidence for a Misery Loves Company effect,

suggesting that False Alarm participants were unique, relative to Hits, in their

overestimation of how many hypothetical others would make the same type of decision on a similar task.

The same question can be asked of those who claimed to be below (or equal to) average. If the Misery Loves Company effect applied to those who made *any* decision error, then a similar pattern should appear, whereby Misses estimate more hypothetical others in their own category than correct rejections. If, however, this effect is unique to self-enhancement error, this pattern should not emerge. A similar ANOVA (Accuracy * Estimate Category) produced no main effects (ps > .242). A significant Accuracy * Estimate Category interaction emerged, $F(1, 96) = 12.13$, $p < .001$, $\eta^2 = .11$, though the resulting pattern did *not* demonstrate a Misery Loves Company effect for self-effacers (see Figure 12, bottom panel). Here, estimates of Misses were greater for both accurate (CR) participants, $t(67) = 3.04$, $p < .003$, $d = .58$, and inaccurate (M) participants, $t(29) = 2.14$, $p < .041$, $d = .63$. In other words, inflating estimates of the number of others in the same category was not characteristic of all inaccurate individuals, but appeared to be unique to inaccurate self-enhancers.
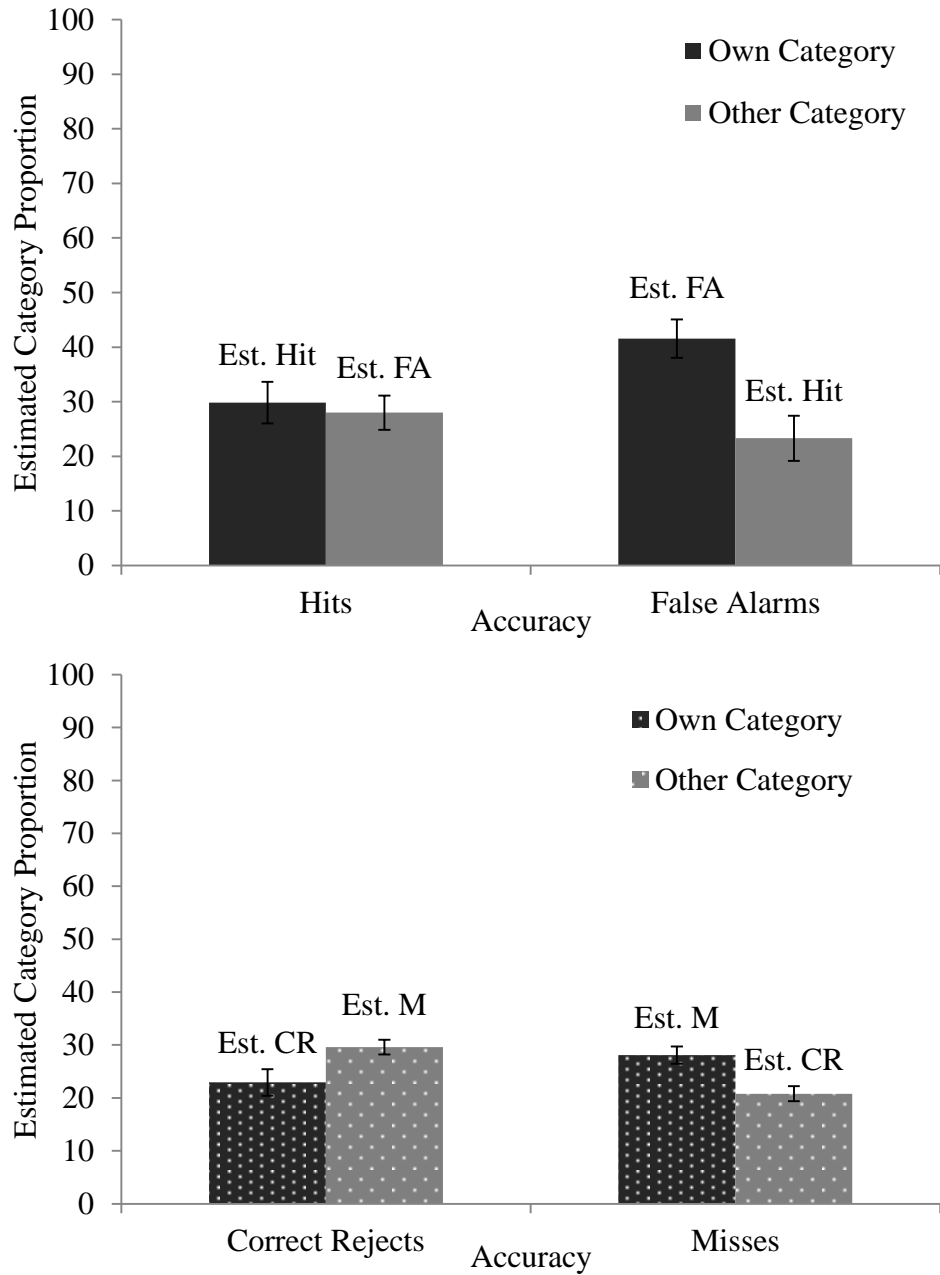
Figure 12. Estimated category membership for own and other category displayed separately for participants who claimed to be better than average (top) and participants who claimed to be worse than (or equal to) average (bottom). Error bars represent one standard error of the mean.

To summarize, the Misery Loves Company effect appeared to be unique to inaccurate participants who claimed to be better than average (False Alarms). FA participants inflated the number of others they estimated to be in the same category, though Hit participants did not. Conversely, for those who claimed to be worse than (or equal to) average (M, CR), participants estimated a greater number of inaccurate individuals (M) in the population than accurate ones (CR), regardless of which category they themselves belonged to. A true focal test of the difference between these interaction patterns requires a three-way ANOVA model entering claim (above or below average) as a between-subjects measure. Because of the skewed nature of this sample, however, several key assumptions of this ANOVA model are severely violated. Study 7 aims to address these assumption violations so that the Misery Loves Company Effect can be replicated and better detailed.

Study 6 provided initial, though limited, evidence in support of the Misery Loves Company effect. Those who learned that they made a self-enhancement error by wrongly claiming to be above average overestimated the number of others who would make the same error in a similar situation. This effect appeared to be unique to False Alarm targets, suggesting a motivated response to receiving unfavorable feedback after self-enhancing. In contrast to the primary result of Study 5, participants did not differ in the certainty they expressed about the accuracy of their comparative self-judgments. This failure to replicate the Favorable Uncertainty effect may have been due to issues with the new scale measure, a skewed sample, or a true null effect. Study 7 aims to replicate both of these

effects and introduces several individual differences measures commonly associated with self-enhancement bias.

Study 7: Motivated Strategies and Individual Differences

Studies 5 and 6 provided an interesting first glance at the two proposed motivational strategies that individuals may employ to protect or enhance the self. However, these studies were limited by a small, skewed sample of the target population and an unclear dependent measure of the Favorable Uncertainty effect. Study 7 attempted to address these limitations by adopting a clearer measure of certainty, including an orthogonal measure, importance (Pelham, 1991), and replacing Study 5 and 6's skewed performance task with a new, pretested trivia quiz designed specifically to produce a similar number of better- and worse-than-average claims. Furthermore, changing the participant pool to online subjects allowed us to collect a larger sample and to include several individual differences measures commonly associated with self-enhancement bias.

Method

Participants ($N = 300$) were recruited on Amazon Mechanical Turk (MTurk; Amazon, 2014). All participants were screened using TurkGate to ensure that they had not previously participated in our studies on self-enhancement (Turkgate, 2013). Participants received $0.75 as compensation.

Sample size was determined to obtain a sufficient number of participants who claimed to be either better or worse than average, and to maximize the likelihood of

observing decision errors (False Alarms and Misses). Similarly, the MLC effect observed in Study 6 was large but arose from only a sample of 13 participants. According to G*Power, detecting a large effect ($d = .9$) with 90% power would require a sample size of $N = 44$ (Faul et al., 2007). Given the nature of population accuracy in similar trivia tasks, we estimated a sample size of 300 participants would yield approximately 50 False Alarm errors.

Materials & Procedure

Trivia task. A 20 item trivia task was adapted from Moore & Small (2007) and presented using Qualtrics online survey platform (Qualtrics, 2014). The task contained medium difficulty items across six domains of trivia knowledge: pop culture, history, science, geography, music, and sports. This task was pretested on 50 Mechanical Turk users over four iterations in order to create a measure with reasonably centered performance and self/other estimates. The final iteration resulted in mean scores, self-estimates, and other-estimates of ~10 out of 20 for each. Each item was presented using a box of text accompanied by a dropdown menu containing the correct answer and four additional foils. All items and all possible answers were presented in random order for each participant. As a screening tool, one question asked participants to select the word "Goodbye" from a list of five synonyms (participants were notified that this question would not affect their score).

Self-esteem. A 10 item scale of trait self-esteem was adapted from Rosenberg (1965). Participants are asked to rate their level of agreement with a series of statements on a scale varying from 1 (Strongly Disagree) to 4 (Strongly Agree), with no neutral

midpoint. The sum of these responses comprise the self-esteem measure. Items were presented in a randomized order for each participant.

Narcissism. A brief measure of non-pathological narcissism was adapted from Ames, Rose, and Anderson (2005). This measure contained a subset of 16 items take from the larger, 40 item Narcissistic Personality Inventory (NPI-40). The NPI-16 was validated as a useful measure of narcissism in place of the more cumbersome NPI-40. Each item asks participants to choose either of two sentences, where one expresses a narcissistic sentiment (e.g., "I am going to be a great person") and the other expresses a similar but non-narcissistic sentiment ("I hope I am going to be successful"). A participant's summed total of narcissistic responses represents their magnitude of narcissism on a 0-16 continuous scale. Items and the two statements contained within each were presented in randomized order for each participant.

Desirable responding. A similarly brief version of the Behavioral Inventory of Desirable Responding was adapted from Hart, Ritchie, Hepper, and Gebauer (2015). The BIDR-16 was validated as a useful measure of desirable responding in cases where administering the complete BIDR-40 is impractical. The scale comprises two eight-item subscales: Self-Deceptive Enhancement (SDE) and Impression Management (IM). Participants are asked to rate the extent to which they agree or disagree with a series of statements on a scale ranging from 0 (Strongly Disagree) to 7 (Strongly Agree), with a neutral midpoint.

Other items and demographics. To measure perceived desirability of trivia knowledge, participants were asked to rate their level of agreement with the statement, "Being good at trivia is socially desirable" on a scale ranging from 1 (Strongly Disagree)

to 7 (Strongly Agree), with a neutral midpoint. Participants also answered the questions, "How content are you with your life, overall?" and "How content do you think the average MTurk user is with their life, overall?" on similar seven-point scales ranging from "Very Discontent" to "Very Content."

On the final page of the survey, participants reported whether they used any outside materials to help complete the trivia task after being reassured that the answer to this question would not affect their approval rating or HIT completion. Participants then provided optional basic demographic information, including gender, ethnicity, country of birth, and where they tend to stand on social and economic issues (from 1 (extremely liberal) to 7 (extremely conservative))

Procedure. Survey materials were presented online (Qualtrics, 2014) to be accessed by residents of the United States. All participants provided informed consent and were told that they would be asked to complete a brief trivia task and complete a few survey measures about themselves. After completing the trivia task, participants were asked to estimate how many questions (out of 20) they answered correctly and to estimate how many questions the average MTurk user answered correctly. These questions were presented in counterbalanced order.

As in Studies 5 and 6, participants were then shown their answers to the estimation questions described above. Using these numbers, participants were asked to identify whether they claimed to perform worse than or equal to the average MTurk user, or better than the average MTurk user. Depending on this self-categorization, participants

were then asked to provide certainty and importance ratings that their self-categorization would turn out to be true. For example, participants were shown the following prompt (parentheses indicate a worse-than-average claim): "*You estimated that your performance was better than (worse than or equal to) the performance of the average person. However, we know that these estimates may contain some error. How certain are you (important is it to you) that your claim will turn out to be true?*" Ratings were made on a seven-point scale ranging from 1 (Not at all Certain/Important) to 7 (Extremely Certain/Important).

Following the certainty measure, participants received accurate feedback on their performance and completed two measures of population estimates: a new exploratory measure and the same measure presented in Study 7. Note that the exploratory measure did not cohere and was not used for any analyses. For this measure, participants were told how many questions they answered correctly and the average score on the task obtained from an earlier pretest on a similar population. This allowed us to clarify to participants which decision type they exhibited. For example, a participant who committed a False Alarm was presented the following summary: "*Your score on the trivia task was [score] out of 20. The average score on this trivia task is 10 out of 20. This indicates that you actually performed worse than (or equal to) the average Mechanical Turk user.*" Participants were also told that out of 100 people who take this test, 50 scored below (or equal to) average. Finally, half of participants were asked to estimate the number of people (out of 50) who made the same claim as them (match condition), while the other half were asked to estimate the number of people who made the opposite claim as them (mismatch condition). Because participants were restricted to estimate category types out

of 50, this allowed for a recovery of above-average performers' estimates of Hits and Misses in the population, and below-average performers' estimates of False Alarms and Correct Rejections.

 The second population estimate task was the same as the measure presented in Study 6. Participants received accurate feedback about their own performance and category standing and then provided estimates for each of the four category types out of a sample of 100 Mechanical Turk users (see Figure 9 for the full item text; note that "Brown students" was replaced with "Mturk users," and the average used for feedback was replaced with the average observed in pretesting).

 After completing these two measures of interest, participants completed measures of trivia desirability, the Rosenberg Self-Esteem scale, the BIDR-16, the NPI-16, self and others' overall contentedness with life, and demographics. Finally, participants were debriefed, given the option to view the answers to the trivia task, and given a completion code to be entered in Mechanical Turk for payment.

Results

Participant characteristics and demographics. $N = 284$ participants remained for analysis after exclusions (eight were excluded for failing the trivia quiz attention check, 11 were excluded for failing to correctly interpret their self- and other-estimate, and four were excluded for admitting to using outside help on the trivia task).

Task performance, estimation, and categorization. Means and intercorrelations for the three target variables (S, O, T), new variables Certainty and Importance, and population estimates of each of the four category types are displayed in Table 6. Over all participants, self-judgment accuracy was moderate to high, $r(282) = .59$, $p < .001$.

Participants also exhibited social projection as represented by the correlation between self-judgment and judgment of the average person, $r(282) = .60$, $p < .001$. O and T were weakly correlated, $r(282) = .33$, $p < .001$.

Table 6

Descriptive Statistics and Intercorrelations over all Participants

$N = 284$	Mean (SD)	O	T	Certainty	Importance	Est. H	Est. FA	Est. M	Est. CR
S	10.27 (4.11)	**0.60**	**0.59**	0.11	**0.12**	**0.19**	0.02	-0.11	-0.09
O	10.56 (3.13)	-	**0.33**	0.00	-0.02	-0.02	-0.10	0.11	0.02
T	11.01 (3.12)		-	0.01	-0.04	**0.20**	-0.05	-0.05	-0.08
Certainty	3.9 (1.47)			-	**0.34**	**0.15**	-0.09	-0.04	0.01
Importance	2.72 (1.58)				-	**0.18**	0.02	-0.06	**-0.14**
Est. H	23.08 (10.54)					-	**-0.29**	**-0.40**	**-0.23**
Est. FA	29.33 (12.51)						-	**-0.40**	**-0.47**
Est. M	24.15 (11.39)							-	**-0.21**
Est. CR	23.44 (10.63)								-

Note. Bold correlation values represent statistical significance at $p < .05$

Participants were categorized according to the decision-theoretic approach. Of 284 total participants, 119 (41.90%) claimed to be better than average, while 165 (58.10%) scored above average. This yielded 90 Hits (31.69%), 29 False Alarms (10.21%), 75 Misses (26.41%), and 90 Correct Rejections (31.69%), suggesting a high degree of accuracy in participants' comparative judgments.

Certainty and importance. Pelham (1991) argued that certainty and importance represent two orthogonal parts of the self-concept, with certainty corresponding to the valence and magnitude of a belief ('confidence') and importance corresponding to that belief's integration in the self-image ('consequence'). Indeed, over all participants, certainty and importance ratings were only weakly correlated, $r(282) = .336$. This

correlation did not differ substantially when computed separately for those who claimed to be better, or worse, than average.

As in Study 5, I expected certainty and importance scores to be related to the extremity of better (or worse) than average claims. Regressing certainty on S – O scores yielded a significant quadratic trend, $F(2,281) = 7.08$, $b_{quadratic} = .001$, $p < .013$, suggesting that extreme claimants were more certain in the accuracy of these claims than those whose self-estimates were similar to their estimates of the average person (see Figure 13, top panel). A similar and unexpected quadratic trend emerged for importance ratings, $F(2,281) = 6.54$, $b_{quadratic} = .013$, $p < .013$ (see Figure 13, bottom panel)[11].

---

[11] Both regression models also had significant linear trends, $b_{Certainty} = .082$, $p < .023$; $b_{Importance} = .097$, $p < .004$. Including these trends in the prediction model did not diminish the quadratic effects.
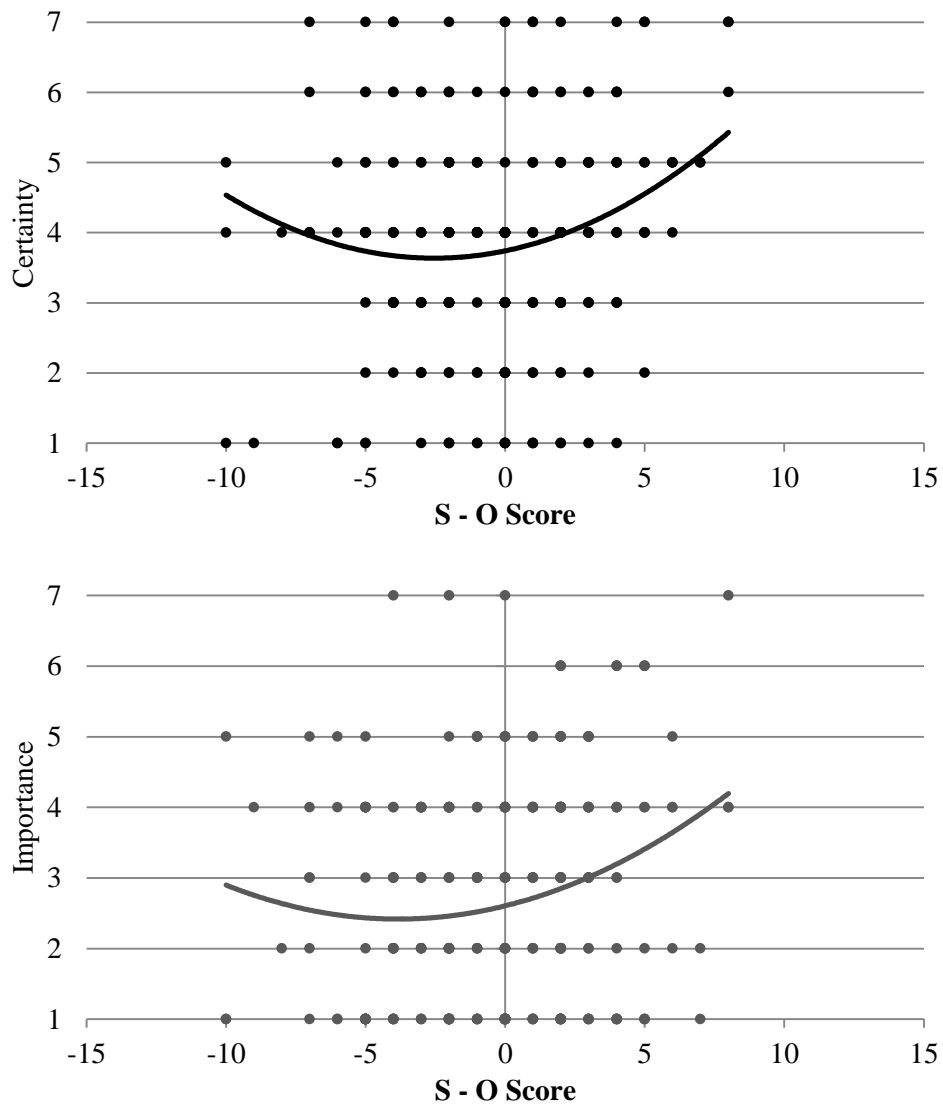
Figure 13. Scatterplots displaying certainty (top) and importance (bottom) varying according to participants' S (self-judgment) minus O (other-judgment) score.

To test for differences in the certainty and importance of self-enhancers' (S > O) and self-effacers' (S ≤ O) comparative claims, each response (certainty; importance) was subjected to an independent groups *t*-test comparing those who claimed to be better than average with those who claimed to be worse than (or equal to) average (see Figure 14 for these results)[12]. Both results were significant in the predicted direction, such that participants who claimed to perform worse than (or equal to) average were less certain, $t(282) = 2.21$, $p < .014$, $d = .27$, and rated it less important, $t(282) = 3.75$, $p < .001$, $d = .45$, that their claim would turn out to be true, even after adopting a stricter alpha for two planned comparisons ($\alpha = .05/2 = .025$). These results successfully replicated and supported the Favorable Uncertainty hypothesis, demonstrating that those who claimed to perform worse than average were less certain in this judgment that those who claimed the opposite. This result was extended by including a novel measure of importance, providing supporting evidence that participants viewed performing better than average as more important than worse than (or equal to) average, although participants judged the overall outcome of their claim as relatively unimportant ($M = 2.78$ on a seven-point scale). This can serve as an operationalization of how favorable participants viewed performing better or worse than average: overall unimportant, but more important to perform better than to perform worse than average.

---

[12] Note that although an ANOVA model may seem appropriate given the 2x2 nature of the independent measures, there is no theoretical reason to expect (or test for) a main effect of judgment domain (certainty *and* importance), or an interaction effect between comparative claim and judgment domain.
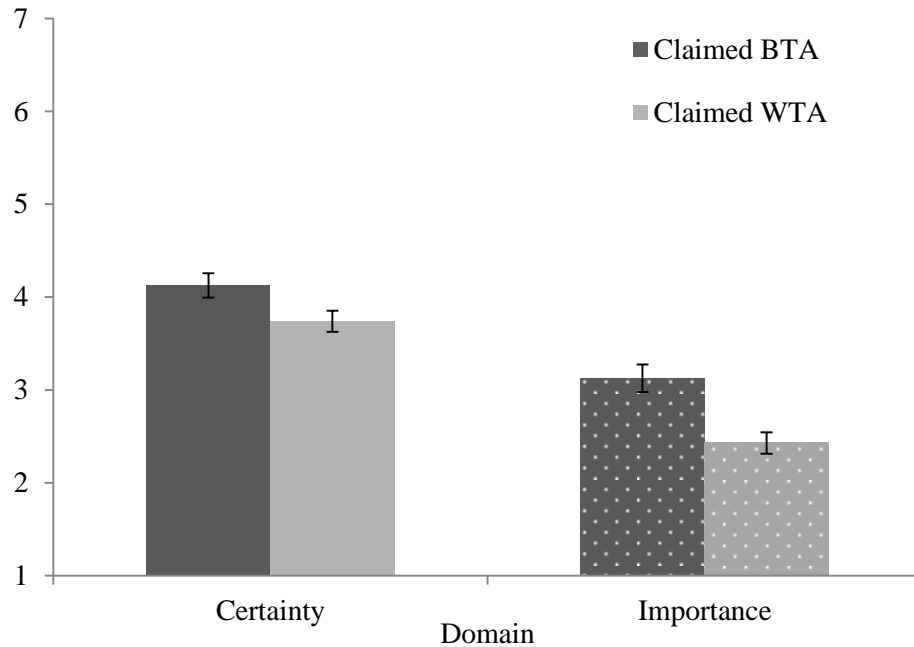
Figure 14. Mean certainty and importance ratings grouped by claims to be better or worse than (equal to) average. Error bars represent one standard error of the mean.

To examine the unique contributions of both importance and certainty to comparative, self-enhancing (or effacing) claims, a regression approach was adopted. We attempted to predict comparative judgments (S – O) by simultaneously entering certainty and importance ratings into a multiple regression model separately over all participants. Importance was a significant predictor of S – O scores, $t(281) = 2.24$, $p < .026$, but certainty was not, $t(281) = .088$, $p < .156$. This indicates that as directional claims of being better than average increased, participants viewed it as more important that this claim would turn out to be true. This is in line with the favorable uncertainty hypothesis: participants appeared to view accurately claiming to be better than average as more important than accurately claiming to be worse than average. For certainty, however, another analytic step is necessary. Because certainty should increase with the *magnitude* of S – O scores, rather than their signed positivity, a second multiple regression was run

entering certainty and importance as predictors of the absolute value of S – O scores.
Here, certainty was expected to predict extreme differences between S and O, as the
quadratic trend suggested. The prediction for importance was less clear. Indeed, certainty
was a significant predictor of the magnitude of S – O scores, $t(281) = 2.38$, $p < .018$,
while Importance was not, $t(281) = .822$, $p < .41$. When analyzing only who claimed to
be above average, Certainty predicted S – O scores, $t(116) = 3.43$, $p < .001$, but
importance did not, $t(116) = .345$, $p < .731$. For those who claimed to be worse than (or
equal to) average, neither Certainty nor Importance could predict S – O scores, p's $> .46$.

Certainty and importance were also used to distinguish between category types.
Because there is overall accuracy in the population, accurate self-perceivers (Hits and
Correct Rejections) likely had greater certainty in their claims than inaccurate self-
perceivers (False Alarms and Misses). Once again, the prediction for importance is
unclear. To answer this question, Hits and False Alarms were assigned dummy codes of 0
and 1, respectively. This coded measure was then regressed on certainty and importance
ratings simultaneously using a logistic regression model. As predicted, certainty was a
significant predictor of Hits and False Alarms, $\beta = -.537$, Wald $= 8.36$, $p < .004$, while
importance was not, $\beta = .210$, Wald $= 1.87$, $p < .171$, suggesting that Hits were more
certain that their above average claim would turn out to be true than False Alarms. The
same approach was used to distinguish between Correct Rejections and Misses with
similar (but weaker) results. Here, certainty was a marginally significant predictor of
Correct Rejections and Misses, $\beta = -.216$, Wald $= 3.38$, $p < .066$, while importance
contributed little, $\beta = .070$, Wald $= .379$, $p < .538$. Based on this analysis, accurate self-
perceivers tended to be more certain that their self-other judgment was correct than

112

inaccurate individuals, although these groups did not differ in how important they rated their claim's accuracy.

In line with the Favorable Uncertainty hypothesis, those who claimed to be better than average were both more certain, and viewed it as more important, that this claim would turn out to be true compared with those who claimed to be below (or equal to) average. Additional regression analyses clarified that participants viewed it as increasingly important that their comparative self-judgment was accurate as the positivity of their better-than-average claim increased. Interestingly, certainty was associated with the magnitude of S – O scores for participants who claimed to perform better than average, but not for those who claimed the opposite. This may be one potential window into the mean-level difference driving the favorable uncertainty effect. Finally, certainty, but not importance, could be used to predict accurate (H, CR) from inaccurate (FA, M) individuals.

Population estimates. The new exploratory measure of population estimation performed poorly and was excluded from analyses. Specifically, estimates provided and derived from these items correlated weakly (~0 - ~.3) with one another and with the more complete measure utilized in Study 6. As such, present analyses were conducted on the long-form population estimate measure where participants estimated how many out of 100 MTurk users fall into each of the four decision categories.

To answer the simplest question posed by the Misery Loves Company hypothesis, averages were once again computed for the number of hypothetical individuals estimated to make the same type of judgment as participants in each category type (Hits estimating Hits, False Alarms estimating False Alarms, etc…). A one-way between-subjects

ANOVA on these estimates replicated the differences observed in Study 6, $F(3, 280) =$ 4.62, $p < .004$, $\eta^2 = .047$ (see Figure 15). Post-hoc tests using Tukey's HSD procedure again confirmed that participants who committed a False Alarm estimated a greater number of False Alarms in the population ($M = 35.69$, $SD = 13.35$) than Hits estimating Hits ($M = 28.30$, $SD = 10.96$), $p < .018$, Misses estimating Misses ($M = 27.36$, $SD = 12.25$) , $p < .007$, and Correct Rejections estimating Correct Rejections ($M = 26.62$, $SD = 11.44$) , $p < .002$. No other differences between groups were significant.

Figure 15. Mean estimates of number of individuals who would make the same category of decision, grouped by participant decision category. Error bars represent one standard error of the mean.

It was clear that participants who wrongly claimed to be better than average uniquely inflated the number similar errors in the population. To further explore the Misery Loves Company effect, analyses must take into account how the four category types of participants viewed their own *and* others' categories (see Figure 16).
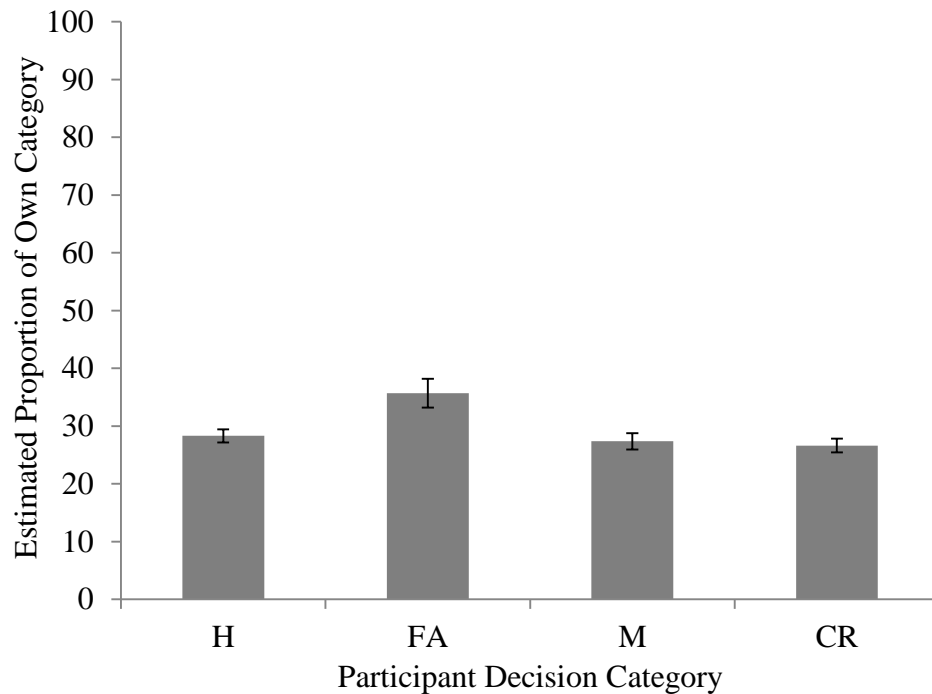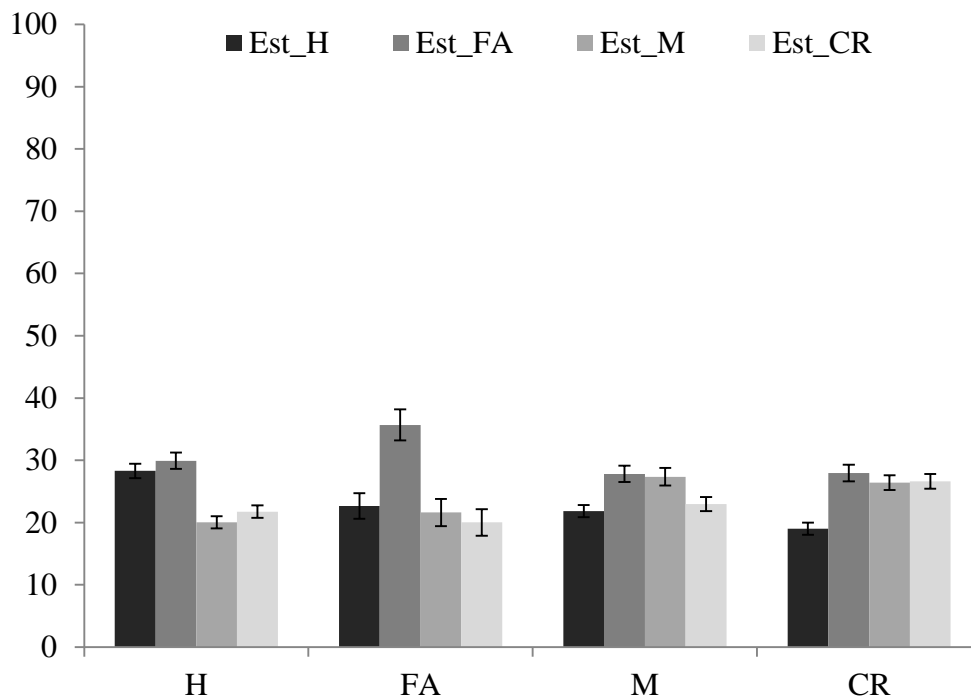
Figure 16. Mean estimates of number of individuals who would make each of the four categories of decision, grouped by participant decision category. Answers must sum to 100 within each participant type. Error bars represent one standard error of the mean.

The omnibus prediction made by the Misery Loves Company hypothesis is that inaccurate self-enhancers (FA participants) will inflate the number of estimated others in their own category compared specifically to groups of other participants including a.) those who correctly claimed to be above average (H), and b.) those who claimed to be worse than (or equal to) average (M, CR). This can be formalized as a two-way interaction effect whereby False Alarm participants estimate a larger difference between False Alarms and Hits than do Hit participants. Recall that this pattern emerged in Study 6 with a trending significance test. However, this two-way interaction should *not* appear for those who claim to be below average: Miss participants should not estimate a greater difference between Misses and Correct Rejections than Correct Rejection participants do.

115

To summarize, a full test of the Misery Loves Company hypothesis predicts a three-way interaction consisting of a significant two-way interaction between judgment accuracy and estimate category for those who claimed to be better than average, but not for those who claimed to be below average. To conduct this analysis, participants' estimates were subjected to a 2 (Claim: (S > O) / (S ≤ O)) * 2 (Accuracy: accurate / inaccurate) * 2 (Estimate: own category type / opposite category type) repeated-measures ANOVA with Claim and Accuracy as between-subjects factors (See Figure 17). The two focal tests comprise 1.) the Claim*Accuracy*Estimate interaction effect, to determine whether a different pattern emerged for those who claimed to be better (worse) than average, and 2.) the Accuracy*Estimate interaction effects separately for these two groups. Indeed, the three-way interaction effect was significant, $F(1, 280) = 4.34$, $p < .038$, $\eta^2 = .015$, suggesting that the focal two-way interaction pattern differed between BTA and WTA claiming participants.

To drill down into this three-way interaction effect, tests of the two-way interaction were conducted separately for BTA-claiming and WTA-claiming participants. For the former, this approach tested the difference between estimates of Hits and False Alarms for those whose claim was accurate (H) or inaccurate (FA). This resulted in a 2 (Accuracy: accurate (H) / inaccurate (FA)) * 2 (Estimate Category: own category / other category) repeated measures ANOVA with Accuracy as a between-subjects factor (See Figure 17, top panel). A main effect of Estimate Category, $F(1, 117) = 11.70$, $p < .001$, $\eta^2 = .091$, revealed that participants estimated a greater numbers of others in their own category ($M = 32.81$) than the opposite category ($M = 25.48$). There was no main effect of Accuracy, $F(1, 117) = 11.70$, $p < .001$, $\eta^2 = .091$, although such an effect was unlikely

because the total estimates of Hits and False Alarms estimated should not differ between category types by nature of the design. The critical test of the Misery Loves Company hypothesis was the Accuracy * Estimate Category interaction. This effect was significant, $F(1, 117) = 7.10$, $p < .009$, $\eta^2 = .057$. Tests of simple effects confirmed the predicted pattern, such that participants who accurately claimed to be above average (Hits) did not differ in their estimates of Hits and False Alarms, $t(89) = .76$, $p < .45$, while participants who inaccurately claimed to be above average (False Alarms), estimated that more False Alarms would occur than Hits, $t(28) = 3.63$, $p < .001$, $d = 1.06$. To confirm the nature of the pattern difference in the three-way interaction, a similar two-way ANOVA was run on participants who claimed to be worse than (or equal to) average. As expected, the Accuracy*Estimate Category interaction effect was not significant, $F(1, 163) = 2.18$, $p < .142$, $\eta^2 = .013$, suggesting that those who made a decision error after claiming to be worse than average (Misses) did *not* produce a Misery Loves Company effect (see Figure 17, bottom panel). Indeed, this effect appears to be unique to self-enhancement error.
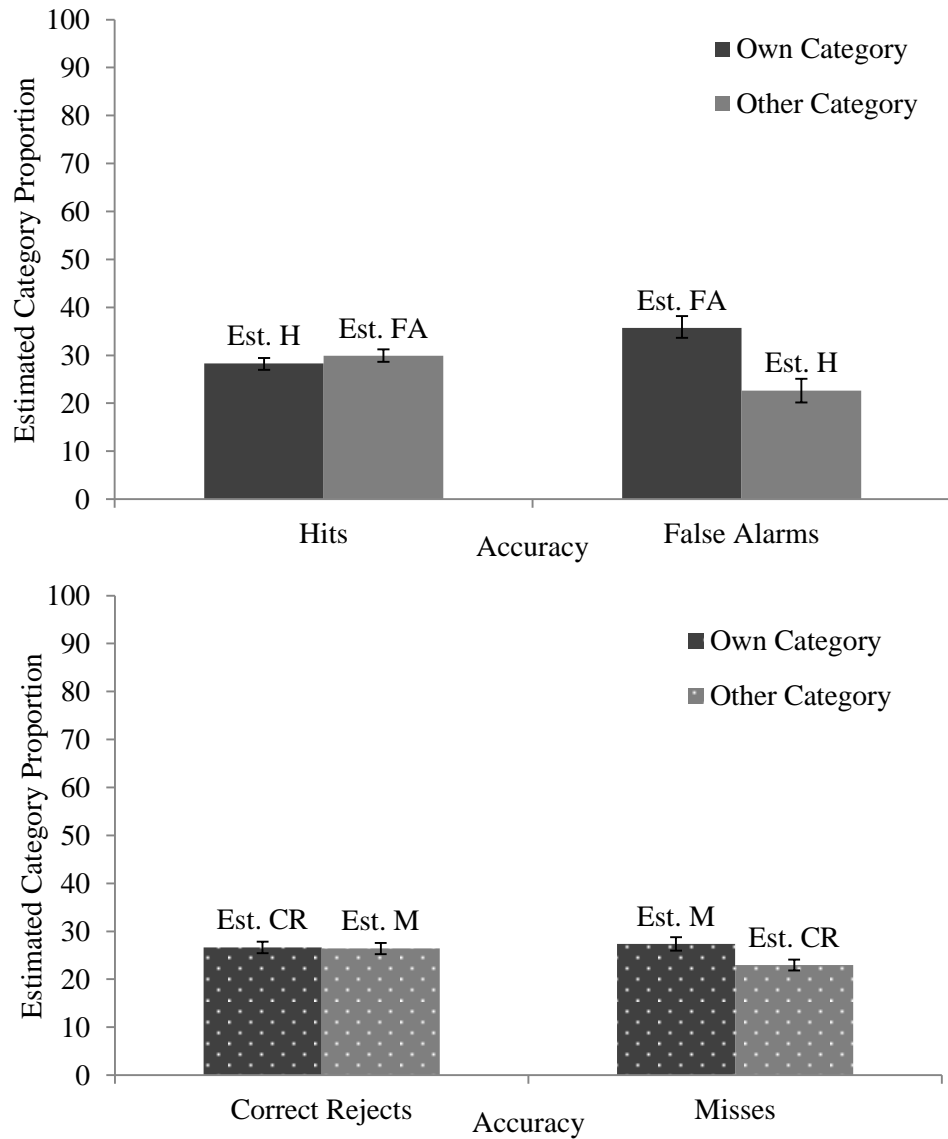
Figure 17. Estimated category membership for own and other category displayed separately for participants who claimed to be better than average (top) and participants who claimed to be worse than (or equal to) average (bottom). Error bars represent one standard error of the mean.

To summarize, the Misery Loves Company effect was unique to those participants who learned that they committed a False Alarm. These individuals overestimated the number of other Mechanical Turk workers who would commit a similar error after

completing a similar task. This effect did not emerge for any other decision category including unfavorable categories revealing either a poor performance (CR) or an inaccurate self-judgment (M). It appears that the unique combination of claiming and failing to be better than average resulted in an overestimation of others who might find themselves in a similarly unfavorable situation.

Individual Differences and Predicting Self-Enhancement Error. The next goal of Study 7 was to introduce several measures thought to be associated with self-enhancing thoughts and behaviors. To date, these measures have yet to be used in attempt to detect self-enhancement *error,* or as a window into the personality of False Alarm participants. Reliability measures and intercorrelations are reported below, followed by regression analyses attempting to predict self-enhancement bias and error.

Table 8 displays reliability coefficients (Cronbach's α) and intercorrelations among scale measures and relevant single-item variables. Items tended to cohere within scales (α between .80 and .95), and scale scores tended to correlate in predictable directions likely due to a combination of shared construct validity and method variance. A few notable relationships deserved mention. Self-esteem was highly correlated ($r =$ .70) with the Self-Deceptive Enhancement subscale of the BIDR, but only modestly correlated with the Impression Management subscale ($r = .23$). This was well in line with work in this area (Hart et al., 2015) and supports the claim that SDE and IM are unique aspects of desirable responding. A similar result emerged for narcissism scores, $r(\text{NPI,SDE}) = .40$; $r(\text{NPI,IM}) = .05$, once again corroborating Hart et al.. Conservativism (political orientation) correlated positively with self-esteem ($r = .15$) and Self-Deceptive Enhancement ($r = .17$). Interestingly, the better-than-average effect for life contentedness

correlated positively with all measures but trivia desirability. Participants' ratings of trivia desirability were not correlated with any other scale measures.

All of the scale measures described have previously been theorized or shown to be associated with self-enhancing thoughts and behaviors, which are often conceptualized as the better-than-average effect. As such, these scale measures were entered into a logistic regression model to determine which measures, if any, could correctly classify those who claimed to perform better than average and those who claimed to perform worse than (or equal to) average. The best performing model contained only the scale measures of self-esteem (SE), $\beta = .041$, Wald $= 3.85$, $p < .050$, and total desirable responding (BIDR), $\beta = -.021$, Wald $= 4.71$, $p < .030$, although classification accuracy remained relatively low (58.5%) even with these significant predictors. No other scale measures were significant predictors when entered into this model independently or together with SE and BIDR scores. There were no differences in inference or directionality when using a multiple regression model to predict the continuous measure of the better-than-average effect, S – O. This allows for the tentative conclusion to be drawn that self-esteem and desirable responding were independently predictive of participants' self-enhancing (or –effacing) claims.

After exploring the scales measures' ability to predict self-enhancement *bias*, we turned to predictions of self-enhancement *error*, namely, committing a False Alarm. Two possible prediction techniques were explored. The first attempted to predict False Alarm participants from all other category types (i.e., False Alarms were assigned a dummy code of 1 and all other category types received a dummy code of 0). Quite surprisingly, no scale measures correlated with this type of coding. Here, there was no model available

that could predict False Alarms from the other three category types using the proposed scale measures.

A second technique restricts analysis to those participants who claimed to be better than average. Here, the goal was to predict False Alarms (dummy coded as 1) from Hits (dummy coded as 0). Interestingly, only Trivia Desirability even weakly correlated with this measure, $r = -.20$, $p < .032$, suggesting that Hits viewed trivia knowledge as more desirable than False Alarms. This measure could significantly predict Hits from False Alarms in a logistic regression model, $\beta = -1.13$, Wald $= 28.13$, $p < .032$, although this result should be considered tentative given the large number of comparisons and exploratory nature of analyses using these scale measures. Contrary to expectation, the individual differences measures presented in Study 7 contributed little to explaining self-enhancement bias and error. Furthermore, none of these measures correlates with the other sets of dependent measures (certainty, importance, and the four category estimates). More work is necessary to determine how individual differences can contribute to research on self-enhancement bias, error, and motivation.

Table 8

Scale Reliability and Intercorrelations

Variable	α	BIDR-Total	BIDR-SDE	BIDR-IM	NPI	Political Orientation	Desirability	Relative Contentedness
Self-Esteem	0.95	0.55	0.70	0.23	0.33	0.15	-0.06	0.52
BIDR-Total	0.86	-	0.87	0.85	0.27	0.08	-0.06	0.35
BIDR-SDE	0.84		-	0.47	0.40	0.17	-0.03	0.44
BIDR-IM	0.79			-	0.05	-0.04	-0.08	0.16
NPI	0.85				-	0.08	-0.01	0.26
Political Orient.	0.80					-	-0.03	0.15
Desirability	n/a						-	-0.06
Relative Contentedness	n/a							-

Note. Desirability and contentedness measures have no alpha because they consist of a single item. BIDR = Behavioral Inventory of Desirable Responding; SDE = Self-Deceptive Enhancement; IM = Impression Management; NPI = Narcissistic Personality Inventory; Relative Contentedness = self-contentedness minus other-contentedness.

General Discussion

Studies 5, 6, and 7 provided novel evidence and successful replications for two motivational strategies that individuals may use to self-enhance within the decision-theoretic framework. College students and online participants expressed a lack of certainty in their comparative self-judgments when this uncertainty could lead to a positive outcome (a Favorable Uncertainty effect). Relative to participants in any other judgment category, those who committed a self-enhancement error by erroneously claiming to be better than average overestimated the number of others in the population who would commit a similar error (a Misery Loves Company effect). However, traditional individual difference measures often associated with motivated self-enhancing thoughts and behaviors failed to predict self-enhancement bias and error or to account for the motivational strategies described.

Favorable uncertainty. The Favorable Uncertainty effect demonstrated in Study 5 and Study 7 supported the hypothesis that individuals may express uncertainty in their claims when the wrongness of these claims would lead to a favorable outcome. The effect emerged in Study 5 and was replicated in Study 7 using a simpler scale measure, but failed to replicate in Study 6 where certainty was measured as a probability estimate ranging from 50 to 100. Because the sample obtained in Study 6 was skewed and contained a small number of self-enhancers, it is difficult to conclude whether this failure to replicate was due to the scale measure, sampling error, or a true null effect. Still, the quadratic and linear patterns observed over all participants in Studies 5 and 7 suggested that the strength of certainty increased with the magnitude of an above average claim, but did not increase (or increased less so) as participants claimed to have performed increasingly *worse* than average. This pattern can explain the nature of the effect observed at the mean level: extreme better-than-average self-judgments result in increased certainty of a high performance, while extreme worse-than-average self-judgments elicit little more certainty than claims to be average (or just below). In other words, certainty appeared to track self-enhancement bias but not self-effacement bias.

Contrary to Pelham (1991), participants' perceived importance of their comparative self-judgments behaved similarly to certainty, even though certainty and importance judgments were only weakly correlated. Those who claimed to be worse than average viewed the accuracy of their comparative claim as less important than those who claimed to be better than average. This result suggests that people view at least one of the following as important: a.) performing well; b.) being accurate in one's claim. Similarly, it validates an important assumption made early in this section, namely, that being better

than average is perceived as important. It is worth noting, however, that perceived importance was quite low across all participants ($M = 2.78$ out of 7), suggesting that people did not particularly care whether their judgments were accurate or not. One clear future direction for research on the Favorable Uncertainty effect is to pursue performance on a task where accuracy is more important to participants. This can be achieved by administering a task where individuals are intrinsically motivated to perform well, for example an intelligence test or a contest among peers, or by creating extrinsic motivation by offering incentives for strong performances and accurate self-other judgments. If individuals are put in a situation where performing poorly is highly aversive for one of these reasons, then the magnitude of the Favorable Uncertainty effect should increase substantially.

At present, interpretation of the Favorable Uncertainty effect is limited. Individual difference measures of self-enhancement and self-worth failed to correlate with perceived certainty and importance, which may call into question the validity of a motivational explanation. Similarly, if it the case that high performers are more accurate than low performers (Kruger & Dunning, 1999), then the Favorable Uncertainty effect may be due simply to an artifact of accuracy, where above-average performers who demonstrate greater accuracy in their self-judgments rationally assign greater certainty to their claims than low performers. As such, better understanding of the relationship between certainty, performance, and self-judgment is necessary.

Misery loves company. Compared to the Favorable Uncertainty effect, the observed overestimation of self-enhancement errors specifically by those who committed one provides robust evidence for a motivational, self-enhancing strategy. Receiving

accurate but aversive feedback was associated with inflated population estimates of others who commit the same error. This effect was unique to those who wrongly claimed to be better than average and could be not explained by social projection alone.

Still, more exploration is necessary to conclude that this effect was caused by self-enhancement error as conceptualized by the decision-theoretic framework. Because all participants received feedback, it is unclear whether False Alarm participants inflated estimates of False Alarms in the population due to the information they received in the experiment or due instead to the unique psychological characteristics of errant self-enhancers. To answer this question, an experiment is necessary that withholds feedback from participants before obtaining population estimates. If it is the case that no-feedback False Alarm targets estimate a similar number of self-enhancement errors as False Alarm targets who learned that their self-enhancing claim was wrong, then it can no longer be concluded that receiving category feedback caused the inflated estimates. This would weaken the present evidence for motivated reasoning in self-enhancement.

A similar limitation comes from the accurate nature of the feedback provided. Because all Hit and False Alarm self-enhancers were given accurate feedback, it is unclear that receiving aversive feedback specifically caused False Alarm participants to inflate their estimates of similar errors. Here it is unclear whether the effect is due to the person (self-enhancement error), or the situation (negative performance feedback). To answer this question, another experiment is necessary where feedback accuracy is manipulated such that participants who claim to be above average are randomly assigned to receive favorable (H) or aversive (FA) feedback on their performance. If the Misery Loves Company effect is due simply to the personality of False Alarm participants, then

125

feedback should have no effect on and False Alarms' own-category estimates. If, however, the effect is due to the negative feedback received, then those who claim to be better than average but learn they performed poorly should inflate their estimates of self-enhancement error regardless of how they actually performed.

Finally, it is worth noting that exploring psychological processes and outcomes within participants who commit self-enhancements errors is constrained by the task they complete. Specifically, of the 564 participants reported in Studies 5, 6, and 7 of this section, only 67 (11.88%) met the criteria for the phenomenon of interest (claiming to be better than average but performing worse than or equal to average). On one hand, this number is encouraging because it only provides further evidence for accuracy in the population and the ubiquitous overdiagnosis of self-enhancement error in the population (see Section I). On the other hand, it calls into question the prevalence of the target population. If this number can represent even a conservative estimate of self-enhancement errors over a large number of tasks and domains, the question remains whether studying the sub-population of individuals who regularly commit such errors is valid and important. Furthermore, if researchers agree that self-enhancement error is an important and prevalent behavior, the question shifts to a practical one of how to sample a population of self-enhancers without expending large amount of resources resulting in largely peripheral data with few observations of the phenomenon of interest. Expanding the nature of the performance tasks used to detect self-enhancement bias and error into important, desirable, and ecologically valid domains including intelligence, ethics/morality, and strategic interactions may be successful in encouraging and detecting greater rates of self-enhancement errors. Furthermore, measuring self-enhancement bias

and error repeatedly within an individual over a series of tasks will substantially increase the presence and detection rates of self-enhancement errors. This procedure is detailed in the general discussion section of this dissertation.

Section III Conclusion

After completing an unimportant task and providing simple numerical estimates of their own and others' performance, individuals appeared to engage in motivated, strategic reasoning in a direction consistent with a positive or protected self-image. Expressing uncertainty when being wrong leads to a favorable outcome, and overestimating the number of similarly inaccurate others in the population, are at least two vehicles individuals may use to protect or enhance the self, even after the task was completed and comparative performance estimates were solidified. Studies 5, 6, and 7 demonstrated that although accuracy in self- and other-judgment remained robust, individuals still managed to manipulate perceptions of their environment to better serve themselves.

Integration & Discussion

Accuracy and error are fundamental to the study of judgment and decision-making. However, as Krueger and Funder (2014) argued, current research tends to favor or privilege reports of the latter. To fairly study problems in social psychology and cognition, this privileging of bias and error must be challenged (Jussim, 2012). Historically, it has been common to accept cognitive errors and biases as describing 'how the mind works,' though this can result in a research bias of its own (a bias for research to seek out biases). Early research on social projection propagated the irrationality of the so-called false consensus effect (Ross, Greene, & House, 1977) until Hoch (1987) and

Dawes (1989) demonstrated that projection is a rational process relying on high levels of interpersonal similarity. Harris and Hahn (2011) similarly showed that a widely-accepted judgment bias, the illusion of invulnerability in forecasting negative experiences, is also overdiagnosed by measures biased to detect biases. They demonstrated empirically and analytically that individuals who have not experienced rare catastrophes are behaving rationally when they claim a below-average probability of experiencing one in the future. Gino, Sharek, and Moore (2011) found that the popular illusion-of-control bias only occurs in the complete absence of actual control. If individuals have some control, this effect diminishes as control increases (a lawful regression effect). My goal in this dissertation was to situate self-enhancement bias among these types of accuracy- and rationality-driven conclusions. By implementing a measure sensitive and specific to both bias and error, future research can better address how social cognition works, rather than continuing to simply condemn its failures.

In the time it has taken to write this dissertation, high impact papers have continued to be published in influential journals exploring individual differences in self-enhancement bias (Leising, 2016; *Assessment*), educational outcomes associated with social-comparison and social-reality approaches to self-enhancement (Chung, Schriber and Robins, 2016; *Personality and Social Psychology Bulletin*), and so-called self-enhancement effects independent of mere self-positivity and performance (Humberg et al., 2016; *Journal of Personality and Social Psychology*). All three of these papers adhere to problematic difference-score and residual approaches that fail to distinguish between bias and error. It is as clear as ever that self-enhancement appears in daily life and continues to be studied in restrictive, myopic ways. Across seven studies borrowing from

three unique theoretical perspectives, this dissertation has attempted to ameliorate these myopic approaches by building accuracy and error into the measurement tools available to scholars of self-enhancement bias.

Researchers can separate self-enhancement bias and error. Studies 1 and 2, a computer simulation, and a meta-analysis of 12 independent samples demonstrated that accuracy abounds in self-enhancing (and -effacing) claims. Social perceivers are sensitive to this distinction as well. In Studies 3 and 4, observers condemned self-enhancement bias in the moral domain, and uniquely disparaged self-enhancement error in the competence domain. Finally, the distinction between bias and error appears to motivate agents to see both high performance and accuracy in self-judgment. Studies 5, 6, and 7 provided initial and replicable evidence that individuals were a.) selectively? certain in the accuracy of their claims depending on whether they claimed to be better or worse than average, and b.) motivated to overestimate the numbers of others committing a similar judgment after learning that they committed a self-enhancement error. The three sections presented in this dissertation indicate that accuracy in self-judgment, and the distinction between bias and error, are critical to continued research in self-enhancement. For each of these sections, I proceed by discussing their limitations and future directions for continued research.

Measurement

Studies 1 and 2 proposed a conceptual and formal framework for measuring self-enhancement in a way that allows for accuracy to be detected in comparative self-judgments. This framework is useful because it separately captures self-enhancement bias and error using parsimonious, face-valid categories that allow for individuals to be

described simply but accurately into those who performed above (Hits, Misses) or below (False Alarms, Correct Rejections) average and those who accurately identified their relative performance (Hits, Correct Rejections). However, such a simple approach is not without limitations. By reducing individuals to diagnostic categories, researchers ignore the variance present in their continuous judgments and performance. Although it is possible to leverage continuous measures *after* having categorized individuals, doing so results in an inevitable range restriction whereby categorized individuals no longer exist at the lowest or highest levels of each scale measure. Thus, researchers must rely on variance within each category type when attempting to predict or explain outcomes associated within particular categories. For explorations of correlational and regression analyses within category types, see Heck and Krueger (2015).

A second limitation of this approach is built into the structure of decision theory. Here, and throughout this dissertation, self-enhancers are treated as those who made a a decision that is inaccurate compared to an objective criterion. This decision may give researchers a window into the type of person who makes it, but categorization still only measures an extra-personal observation based on an outcome and not a psychological process. Simply put, the decision-theoretic approach cannot predict or determine who will make a self-enhancement error in the future. Indeed, research has struggled for decades with how to conceptualize self-enhancement as a trait or personality type (John & Robbins, 1994; Krueger & Wright, 2011; Paulhus, 1998).

The goal of the present research was not to predict self-enhancement errors that individuals may make in the future. However, this is a promising research direction. Borrowing from Paulhus' overclaiming technique (1998), the decision theoretic approach

can be applied repeatedly within an individual over a series of tasks. Paulhus attempted to

diagnose bias and error by applying the principles of signal detection theory to

individuals' familiarity responses to real and fake stimuli. If an individual claimed to

recognize a realistic but fabricated term (e.g., 'Plates of Parallax;' 'ultra-lipids') from

among a series of actual terms, they were labeled as having made an error. Furthermore,

each individual was assigned an idiosyncratic error rate, which Paulhus labeled as a

measure of self-enhancement independent of an individuals' bias, or their likelihood of

claiming familiarity of any item.

In order to adapt the decision-theoretic approach to a within-person measure,

researchers must administer a series of objective tasks and self- and other-estimates must

be collected after each task. This series of tasks could be administered according to two

approaches. A within-task approach follows from Paulhus (1998) where a single

participant, who completes a single task comprised of multiple items, provides a self-

estimate and an estimate of the average person's performance *after completing each item*.

For example, a participant may be asked after each trivia question, "do you think you

answered this item correctly?" and "do you think the average test-taker answered this

item correctly?" This type of categorical data obtained over items but within participants

would yield both a parameter for bias (likelihood of claiming to answer correctly while

estimating that the average person does not) and an independent parameter for error

(likelihood of claiming to answer correctly when the average person does not *and* getting

the answer wrong). An alternative approach can remove social comparison completely by

relying instead on only an individual's prediction of answering the question correctly

(here, the average person's performance is irrelevant). Indeed, a within-person approach

131

yields a complete categorization scheme within an individual over repeated brief performances and comparative self-judgments. By doing so, an individual who committed multiple False Alarm errors can be predicted to make more False Alarm errors in the future. As in Paulhus (1998), these within-person estimates are better suited to correlate with other individual differences measures associated with self-enhancement (see Study 7 for the difficulties encountered with correlating multiple scale measures with simple categorical responses). A between-task approach can attempt to categorize an individual over a series of task contexts and domains (e.g. Leising, 2016, where self-enhancement data were collected from respondents who completed 17 unique tasks in domains of creativity, performance, and intelligence, among others). Such an approach relies on similar analytic techniques to determine an individual's unique likelihood of committing a self-enhancement error in the future.

Still, however, the decision-theoretic approach should not be condemned for its simplicity. Indeed, the parsimony and face-validity of the technique is a major strength of this research program. Future research using this technique can turn to exploring the emergent qualities of category types without having to incorporate more complex continuous measures. This approach allows for the radical question to be asked: what may emerge from Hits or False Alarm individuals that cannot be explained by their self- and other-judgment alone? For example, Krueger & Heck (in preparation) explored how categorization labels could predict self-esteem even when controlling for the primary input variables S, O, and T. Though this regression model did not achieve significance, self-esteem scale scores could explain a small amount of variance in categorization labels when entered alongside the variables necessary to categorize individuals in the first place.

In other words, decision-theoretic categorization introduced variance that was unaccounted for by the input variables. Similarly, as demonstrated in Section II, social perceivers were especially responsive to the distinctions between category types yet would likely pay little attention to differences between targets who claimed to be "somewhat" above average and targets who claimed to be "slightly" above average. In other words, the category types explored in Section I may explain outcome measures or questions of psychological process beyond their continuous input variables.

Social Perception

In Section II, social perceivers were shown to be able to discriminate between biased and errant comparative judgments from within a reputational framework. This was an encouraging result because it demonstrated face validity in the decision-theoretic approach to self-enhancement bias and error and generated novel strategic conclusions for self-presentation and impression management in orthogonal domains of social perception. Although Studies 3 and 4 succeeded in adapting decision categories to interpretable and meaningful target descriptions, they suffered from several methodological limitations. Specifically, these studies lacked ecological validity due to the 'single exposure' nature of each target type. Given this constraint, a concern with these studies' design is that participants were merely judging decision types and not the individuals making those decisions. To overcome this limitation, future work in this area should explore participant reactions to individuals whose self-judgments either vary or remain consistent over time. Similarly, to better understand the strategic nature of self-enhancing (or -effacing) claims, I expect that social perceivers will engage in different

types of information search when attempting to determine whether a target's claim was genuine or false.

With regard to the first future direction, participants can be asked to view targets whose comparative self-judgments are either consistent or inconsistent. Recall the result that Hits, those targets who correctly claimed to have performed above average, were perceived as particularly competent and as somewhat moral. Would these targets continue to be lauded if they consistently claimed self-superiority in the same domain? Over varying domains? Paulhus' (1998) and Hoorens' (2015) work on perceived narcissism and the hubris hypothesis, respectively, suggest that social perceivers may find consistent self-enhancing claims to be grating or aversively narcissistic. Tesser's (1988) self-evaluation maintenance model similarly predicts that perceivers in a competitive mindset may severely disparage those who perform well and know it. Conversely, social perceivers may give a benefit of the doubt to False Alarm targets who either perform well or admit their inferiority in other domains. This approach suggests that perceivers may have a threshold or tolerance for self-enhancing behavior, and that this threshold may vary with observers' characteristics and the relevance of the task at hand. It is a reasonable speculation that those who consistently demonstrate accuracy in their self-judgment may be able to 'get away with' an occasional self-enhancement error, while those who consistently but accurately brag to others about their above average performances may be perceived as infallible, or, 'too perfect.' It would be an intriguing and novel result indeed if occasionally making a decision error could improve a target's reputation.

The second future direction is concerned with the strategic nature of making self-enhancing claims. Error management theory and research on deception suggest that individuals should be motivated and expend more resources to detect a costly deception than to detect trivial one (Haselton et al., 2009, von Hippel & Trivers, 2011). The same prediction should hold when considering observers' information search processes regarding others' self-enhancing claims. Studies 3 and 4 demonstrated that claiming to be better than average was riskier than self-effacement in the competence domain: being proven wrong (thus committing a False Alarm) was the worst possible outcome. When the decision to claim self-superiority is risky or strategic, for example in a competitive environment, then observers should invest more resources into uncovering evidence that supports or refutes the target's claim. This proposal suggests that social perceivers should not differ in the extent to which they engage in information seeking when determining only how a target performed, but that observers should be more likely to seek performance information when a self-enhancing claim is made than in the presence of a self-effacing claim. Specifically, participants should react more quickly to performance information, selectively seek it over irrelevant information, and be willing to pay more to access this information when presented with a self-enhancing target than when presented with a self-effacing target. Similarly, observers should experience more positive affect when successfully detecting or revealing a False Alarm target than a Hit, Miss, or Correct Rejection. This affective explanation may amount to feelings of schadenfreude (Feather, 1994).

Motivation

Studies 5, 6, and 7 provided initial evidence for how motivated reasoning can arise in participants who demonstrate self-enhancement bias and error. Section III summarized several of the limitations of these three studies, perhaps the greatest of which is concerned with task versus trait motivation. To date, it cannot be concluded whether participants were motivated by the task itself and the feedback they received or by the experience of merely committing self-enhancement bias or error. In order to differentiate between these types of motivation, additional manipulations are necessary to disentangle the effects of committing a False Alarm and learning that one was committed. These studies are currently in preparation.

Another future direction for motivation research within the decision-theoretic framework borrows from the extension proposed in Section II, namely, information search. When participants are asked whether they would like to know how they performed on the task, a categorical framework generates unique predictions in line with the favorable uncertainty hypothesis. First, individuals should be motivated to either a.) confirm a self-enhancing claim or b.) disconfirm a self-effacing claim. This suggests that Hits and Misses should be more willing to solicit or obtain performance information about themselves than False Alarms or Correct Rejections, both of whom do not benefit from learning their score after having made a claim. A strong prediction can be made if those who committed a False Alarm *know* they committed a False Alarm: these individuals should be uniquely motivated to avoid information about their performance. A substantial amount of metacognitive effort is necessary to both 1.) claim to be above average despite knowing this is likely to be false and 2.) selectively avoid information that may prove their deceptive self-judgment wrong. This prediction contradicts Dunning

and Kruger's (1999) theory of meta-ignorance, which argues that low performers are generally both unskilled and unaware of their lack of skill. Thus it may be the case that False Alarm targets can be further subdivided into deceptive False Alarms—those who actively attempt to deceive themselves or others—and meta-ignorant False Alarms— those who are genuinely unaware of where they stand.

Finally, Studies 5, 6, and 7 are agnostic with regard to which types of motivation matter more to individuals. In all cases, the motivation to perform well was treated as similar in magnitude to the motivation to be accurate. However, there are clear examples of situations where one motivation may be more important or meaningful than the other. On a high-stakes placement exam like the MCAT or GRE, for example, individuals should be strongly motivated by achieving a high performance but would likely not care whether or not they accurately predicted their performance immediately following the test. These individuals can be predicted to privately predict their own superiority without concern for the accuracy of this claim. Conversely, claim accuracy becomes more important when coordinating with others. When planning a wedding, for example, having a large number of friends and family attend is a desirable outcome, but accurately predicting whether you have a small or large number of attendees is more important for coordination and preparation than the number of attendees observed at the time. Performance or reality motivation, separate from accuracy motivation, may be able to be experimentally manipulated with the goal of optimizing decision-making by leveraging motivational tradeoffs.

Conclusion

Self-enhancement is a familiar and intuitive phenomenon that presents the unique challenge of attempting to understand a mundane experience by setting aside one's own experiences with it. It is surely the case that each approach to conceptualizing and measuring the phenomenon appears to its originator to be superior to the alternatives; the research presented in this dissertation is no exception. It is clear, however, that only by allowing for accuracy to emerge can self-enhancement error be properly detected.

References

Abele, A. E., Cuddy, A. J., Judd, C. M., & Yzerbyt, V. Y. (2008). Fundamental dimensions
of social judgment. Special Issue of the *European Journal of Social Psychology*. doi:
10.1002/ejsp.574

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

Alicke, M. D., & Sedikides, C. (Eds.). (2011). *Handbook of self-enhancement and self-
protection*. New York, NY: Guilford Press.

Allison, S. T., Messick, D. M., & Goethals, G. R. (1989). On being better but not smarter
than others: The Muhammad Ali effect. *Social Cognition*, *7*, 275-295. doi:
10.1521/soco.1989.7.3.275

Ames, D. R., Rose, P., & Anderson, C. P. (2006). The NPI-16 as a short measure of
narcissism. *Journal of Research in Personality*, *40*, 440-450. doi:
http://dx.doi.org/10.1016/j.jrp.2005.03.002

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

Anderson, C., Brion, S., Moore, D. A., & Kennedy, J. A. (2012). A status-enhancement
account of overconfidence. *Journal of Personality and Social Psychology*, *103*, 718-
735. doi: http://dx.doi.org/10.1037/a0029395

Baron, J. (2012). Where do non-utilitarian moral rules come from? In J. I. Krueger (Ed.)

Social judgment and decision-making (pp. 261-278). New York, NY: Psychology
Press.

Baumeister,R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger
than good. *Review of General Psychology, 5*, 323-370. doi: 10.1037//1089-
2680.5.4.323

Baumeister, R. F., Tice, D. M., & Hutton, D. G. (1989). Self-presentational motivations
and personality differences in self-esteem. *Journal of personality*, *57*, 547-579.
doi: 10.1111/j.1467-6494.1989.tb02384.x

Beggan, J. K., Vencill, J. A., & Garos, S. (2013). The good-in-bed effect: College
students' tendency to see themselves as better than others as a sex partner. *The
Journal of Psychology*, *147*, 415-434. doi:
http://dx.doi.org/10.1080/00223980.2012.707992

Brambilla, M., & Leach, C. (2014). On the importance of being moral: The distinctive
role of morality in social judgment. *Social Cognition, 32*, 397-408. doi:
10.1521/soco.2014.32.4.397

Brown, J. D. (1986). Evaluations of self and others: Self-enhancement biases in social
judgments. *Social Cognition*, *4*, 353–376. doi:10.1521/soco.1986.4.4.353

Brown, J. D. (2010). Across the (not so) great divide: Cultural similarities in self-
evaluative processes. *Social and Personality Psychology Compass*, *4*, 318-330.
doi: 10.1111/j.1751-9004.2010.00267.x

Brown, J. D. (2012). Understanding the better than average effect: Motives (still) matter.
Personality and Social Psychology Bulletin, *38*, 209–219.
doi:10.1177/0146167211432763

Brown, J. D., & Han, A. (2012). My better half partner enhancement as self-enhancement. *Social Psychological and Personality Science*, *3*, 479–486. doi:10.1177/1948550611427607

Carli, L. L., LaFleur, S. J., & Loeber, C. C. (1995). Nonverbal behavior, gender, and influence. *Journal of Personality and Social Psychology, 68*, 1030-1041, doi: http://dx.doi.org/10.1037/0022-3514.68.6.1030

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

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology*, 68, 1152-1162. doi: 10.1037/0022-3514.68.6.1152

Cooper, D. J., & Rege, M. (2011). Misery loves company: Social regret and social interaction effects in choices under risk and uncertainty. *Games and Economic Behavior*, *73*, 91-110. doi: http://dx.doi.org/10.1016/j.geb.2010.12.012

Crocker, J., & Wolfe, C. T. (2001). Contingencies of self-worth. *Psychological Review*, *108*, 593-623. doi: http://dx.doi.org/10.1037/0033-295X.108.3.593

Dawes, R. M. (1988). *Rational choice in an uncertain world*. San Diego, CA: Harcourt, Brace, Jovanovich.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, *31*, 169-193. doi: 10.1146/annurev.ps.31.020180.001125

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668-1674. doi: 10.1126/science.2648573

Dunning, D. (2012). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York, NY: Psychology Press.

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*, 58–68. doi:10.1037/0022-3514.69.1.58

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., ... & Brown, E. R. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68-82. doi: http://dx.doi.org/10.1016/j.jesp.2015.10.012 0022-1031/

Einhorn, H. J. (1986). Accepting error to make less error. Journal of Personality Assessment, 50, 387-395. DOI: 10.1207/s15327752jpa5003_8

Epley, N., Converse, B. A., Delbosc, A., Monteleone, G. A., & Cacioppo, J. T. (2009). Believers' estimates of God's beliefs are more egocentric than estimates of other people's beliefs. *Proceedings of the National Academy of Sciences*, *106*(51), 21533–21538. doi:10.1073/pnas.0908374106

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self-or social prediction? *Journal of Personality and Social Psychology*, *79*, 861-875. doi:10.1037/0022-3514.79.6.861

Epley, N., & Dunning, D. (2006). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin*, *32*, 641-655. doi: 10.1177/0146167205284007

Eriksson, K., & Funcke, A. (2014). Humble Self-Enhancement Religiosity and the
Better-Than-Average Effect. *Social Psychological and Personality Science*, *5*, 76-
83. doi: 10.1177/1948550613484179

Exline, J. J., & Geyer, A. L. (2004). Perceptions of humility: A preliminary study. *Self
and Identity*, *3*, 95-114. doi: 10.1080/13576500342000077

Epley, N., & Gilovich, T. (2016). The Mechanics of Motivated Reasoning. *The Journal
of Economic Perspectives*, *30*, 133-140. doi: https://doi.org/10.1257/jep.30.3.133

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible
statistical power analysis program for the social, behavioral, and biomedical
sciences. *Behavior Research Methods*, *39*, 175–191. doi: 10.3758/BF03193146

Feather, N.T. (1994). Attitudes toward high achievers and reactions to their fall. In M.P.
Zanna (Ed.) Advances in experimental social psychology: Vol. 26 (pp. 1–73). San
Diego, CA: Academic Press.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*, 117–
140. doi:10.1177/001872675400700202

Fiedler, K., & Krueger, J. I. (2012). More than an artifact: Regression as a theoretical
construct. In J. I. Krueger (Ed.), *Social judgment and decision-making* (pp. 171-
189). New York, NY: Psychology Press.

Gaertner, L., Sedikides, C., & Chang, K. (2008). On pancultural self-enhancement: Well-
adjusted Taiwanese self-enhance on personally valued traits. *Journal of Cross-
Cultural Psychology*, *39*, 463-477. doi: 10.1177/0022022108318431

Galesic, M., Olsson, H., & Rieskamp, J. (2012). Social sampling explains apparent biases

in judgments of social environments. *Psychological Science*, 23, 1515-1523.

doi:10.1177/0956797612445313

Gino, F., Sharek, Z., & Moore, D. A. (2011). Keeping the illusion of control under

control: Ceilings, floors, and imperfect calibration. *Organizational Behavior and

Human Decision Processes, 114*, 104-114. doi:10.1016/j.obhdp.2010.10.002

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The

strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral

Decision Making*, *26*, 213-224. doi: 10.1002/bdm.1753

Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person

perception and evaluation. *Journal of Personality and Social Psychology, 106*,

148-168. doi: 10.1037/a0034726

Gray, H. M., Ishii, K., & Ambady, N. (2011). Misery loves company: When sadness

increases the desire for social connectedness. *Personality and Social Psychology

Bulletin*, *37,* 1438-1448, doi: 10.1177/0146167211420167

Green, J. D., & Sedikides, C. (2004). Retrieval selectivity in the processing of self-

referent information: Testing the boundaries of self-protection. *Self and Identity*,

3, 69–80. doi:10.1080/13576500342000059

Guenther, C. L., & Alicke, M. D. (2008). Self-enhancement and belief perseverance.

Journal of Experimental Social Psychology, *44*, 706–712.

doi:10.1016/j.jesp.2007.04.010

Guenther, C. L., & Timberlake, E. A. (2012). The motivated Self: Self-affirmation and

the Better-Than-Average Effect. *Personality and Social Psychology Bulletin*,

144

0146167212457074. doi:10.1177/0146167212457074

Harris, A. J. L., & Hahn, U. (2011). Unrealistic optimism about future life events: A cautionary note. *Psychological Review, 118*, 135-154. http://dx.doi.org/10.1037/a0020997

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open*, *5*, 2158244015621113.

Hartley, A. G., Furr, R. M., Helzer, E. G., Jayawickreme, E., Velasquez, K. R., & Fleeson, W. (2016). Morality's Centrality to Liking, Respecting, and Understanding Others. *Social Psychological and Personality Science*, online first. doi: 10.1177/1948550616655359.

Haselton, M. G., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition, 27*, 733-763. doi: 10.1521/soco.2009.27.5.733

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. *Journal of Personality and Social Psychology*, *60*, 895-910. doi: 10.1037/0022-3514.60.6.895

Heck, P.R., & Krueger, J.I. (2015). Self-enhancement diminished. *Journal of Experimental Psychology: General, 144,* 1003-1020. doi: http://dx.doi.org/10.1037/xge0000105

Heck, P. R., Krueger, J. I., & Sachs, J. F. (unpublished). *Decision-theoretic self-enhancement after false performance feedback*. Brown University, Providence, RI.

145

Heine, S. J., & Hamamura, T. (2007). In search of east Asian self-enhancement.

    Personality and Social Psychology Review, *11*, 4–27.

    doi:10.1177/1088868306294587

Hoch, S. J. (1987). Perceived consensus and predictive accuracy: The pros and cons of

    projection. *Journal of Personality and Social Psychology*, *53*, 221-234. doi:

    10.1037/0022-3514.53.2.221

Hoorens, V., Pandelaere, M., Oldersma, F., & Sedikides, C. (2012). The Hubris

    hypothesis: You can self-enhance, but you'd better not show it. *Journal of*

    Personality, 80, 1237-1274. OI: 10.1111/j.1467-6494.2011.00759.x

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual

    differences in self-enhancement and the role of narcissism. *Journal of Personality*

    and Social Psychology, 66, 206-219. doi: 10.1037/0022-3514.66.1.206

Jussim, L., Stevens, S. T., & Salib, E. R. (2012). The strengths of social judgment: A

    review based on the goodness of fit index. In J. I. Krueger (Ed.), *Social judgment*

    and decision making (pp. 97-114). New York, NY: Psychology Press.

Kennedy, J. A., Anderson, C., & Moore, D. A. (2013). When overconfidence is revealed

    to others: Testing the status-enhancement theory of overconfidence.

    Organizational Behavior and Human Decision Processes, *122*, 266-279.

    doi:10.1016/j.obhdp.2013.08.005

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

    NY: Guilford Press.

Klar, Y., & Giladi, E. E. (1999). Are most people happier than their peers, or are they just happy? *Personality and Social Psychology Bulletin*, *25*, 586–595. doi:10.1177/0146167299025005004

Klein, N., & Epley, N. (2016). Maybe holier, but definitely less evil, than you: Bounded self-righteousness in social judgment. *Journal of Personality and Social Psychology*, *110*, 660-674. doi: http://dx.doi.org/10.1037/pspa0000050

Krueger, J. I. (1998). Enhancement bias in the description of self and others. *Personality and Social Psychology Bulletin*, *24,* 505-516.

Krueger, J. I. (2007). From social projection to social behaviour. *European Review of Social Psychology, 18,* 1-35. doi:10.1080/10463280701284645

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: Outcome bias and evidential reasoning. *American Journal of Psychology, 120*, 593-618.

Krueger, J. I., Freestone, D., & MacInnis, M. L. (2013). Comparisons in research and reasoning: Toward an integrative theory of social induction. *New Ideas in Psychology*, *31*, 73-86. doi: 10.1016/j.newideapsych.2012.11.002

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*, 313–327. doi: 10.1017/S0140525X04000081

Krueger, J.I., Freestone, D., & Heck, P.R. (unpulished). The inductive reasoning model: Rationale and simulation.

Krueger, J. I., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality & Social Psychology*, *82*, 180–188. doi:10.1037//0022-3514.82.2.180

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, NY: Guilford.

Krueger, J. I., Vohs, K. D., & Baumeister, R. F. (2008). Is the allure of self-esteem a mirage after all? *American Psychologist, 63, 64-65.* doi: http://dx.doi.org/10.1037/0003-066X.63.1.64

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology*, *77*, 221-232. doi: 10.1037//0022-3514.77.2.221

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121-1134. doi: 10.1037/0022-3514.77.6.1121

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

Kurt, A., & Paulhus, D. L. (2008). Moderators of the adaptiveness of self-enhancement: Operationalization, motivational domain, adjustment facet, and evaluator. *Journal of Research in Personality*, *42*, 839-853. doi: 10.1016/j.jrp.2007.11.005

Kwan, V. S. Y., John, O. P., Kenny, D. A., Bond, M. H., & Robins, R. W. (2004). Reconceptualizing individual differences in self-enhancement bias: An interpersonal approach. *Psychological Review*, *111*, 94–110. doi:10.1037/0033-295X.111.1.94

Lafrenière, M.-A., Sedikides, C., Van Tongeren, D. R., & Davis, J. (2015). On the perceived intentionality of self-enhancement. *The Journal of Social Psychology, 156*, 28-42. doi: 10.1080/00224545.2015.1041447

Lamba, S., & Nityananda, V. (2014). Self-deceived individuals are better at deceiving others. *PloS one*, *9*, e104562. doi: 10.1371/journal.pone.0104562

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*, 76-94. doi: http://dx.doi.org/10.1016/j.obhdp.2006.10.002

Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 32, pp. 1–62). San Diego: Academic Press.

Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2015). Quantifying the association of self-enhancement bias with self-ratings of personality and life satisfaction. *Assessment*, 23, 588-602. doi: 10.1177/1073191115590852

Lynn, S. K., & Feldman Barrett, L. (2014). "Utilizing" signal detection theory.

      Psychological Science. Advance online publication. doi:

      10.1177/0956797614541991

MacDonald, G., Saltzman, J. L., & Leary, M. R. (2003). Social approval and trait self-

      esteem. *Journal of Research in Personality*, *37*, 23-40. doi: 10.1016/S0092-

      6566(02)00531-7

MATLAB 8.0 and Statistics Toolbox Release 2012b [software]. (2012). Natick,

      Massachusetts: The Mathworks, Inc.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York, NY: Wiley.

Mijović-Prelec, D., & Prelec, D. (2010). Self-deception as self-signaling: a model and

      experimental evidence. *Philosophical Transactions of the Royal Society B:*

      Biological Sciences, *365*(1538), 227-240. doi: 10.1098/rstb.2009.0218

Moore, D. A., & Cain, D. M. (2007). Overconfidence and underconfidence: When and

      why people underestimate (and overestimate) the competition. *Organizational*

      Behavior and Human Decision Processes, *103*, 197-213.

      doi:10.1016/j.obhdp.2006.09.002

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological*

      Review, *115*, 102-117. doi: 10.1037/0033-295X.115.2.502

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison

      effect. *Journal of Personality and Social Psychology*, *85,* 1121-1135. doi:

      http://dx.doi.org/10.1037/0022-3514.85.6.1121

Moore, D. A., Kurtzberg, T. R., Fox, C. R., & Bazerman, M. H. (1999). Positive illusions

      and forecasting errors in mutual fund investment decisions. *Organizational*

Behavior and Human Decision Processes, *79*, 95-114. doi:
10.1006/obhd.1999.2835

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, *9*, 972–989. doi:10.1037/0022-3514.92.6.972

Pascal, B. (1669/1962). *Pensées*. London, England: Harvill Press.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of personality and social psychology*, *46*, 598. doi: http://dx.doi.org/10.1037/0022-3514.46.3.598

Paulhus, D. L. (1998). Interpersonal and intrapsychic adaptiveness of trait self-enhancement: A mixed blessing? *Journal of Personality and Social Psychology, 74*, 1197-1208. doi.org/10.1037/0022-3514.74.5.1197

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*, 890–904. doi:10.1037/0022-3514.84.4.890

Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*, 1025-1060. doi: 10.1111/1467-6494.00041

Pelham, B. W. (1991). On confidence and consequence: the certainty and importance of self-knowledge. *Journal of Personality and Social Psychology*, *60*, 518-530. doi: http://dx.doi.org/10.1037/0022-3514.60.4.518

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self

    versus others. *Personality and Social Psychology Bulletin*, *28*, 369-381. doi:

    10.1177/0146167202286008

Qualtrics Research Suite [Survey software]. (2013). Provo, UT: Qualtrics.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in

    interpersonal perception. *Psychological Review*, *86*, 61-79. doi:

    http://dx.doi.org/10.1037/0033-295X.86.1.61

Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A

    review and meta-analysis. *Personality and Social Psychology Review*, *9*, 32–47.

    doi:10.1207/s15327957pspr0901_3

Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits

    and long-term costs. *Journal of Personality and Social Psychology, 80*, 340-352.

    doi.org/10.1037/0022-3514.80.2.340

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton

    University Press.

Ross, L., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric

    bias in social perception and attribution processes. *Journal of Experimental Social*

    Psychology, 13, 279-301. doi:10.1016/0022-1031(77)90049-X

Rozin, P., Millman, L., & Nemeroff, C. (1986). Operation of the laws of sympathetic

    magic in disgust and other domains. *Journal of Personality and Social*

    Psychology, *50*, 703-712. doi: http://dx.doi.org/10.1037/0022-3514.50.4.703

Schachter, S. (1959). *The Psychology of Affiliation*. Stanford, CA: Stanford University

    Press.

Schroeder,-Abé, M., Rentzsch, K., Asendorpf, J. B., & Penke, L. (2015). Good enough

    for an affair: Self-enhancement of attractiveness, interest in potential mates and

    popularity as a mate. *European Journal of Personality*, *30*, 12-18. doi:

    10.1002/per.2029

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives*

    on Psychological Science, 3, 102-116. doi: 10.1111/j.1745-6916.2008.00068.x

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal*

    of Personality and Social Psychology, *84*, 60-79. doi:

    http://dx.doi.org/10.1037/0022-3514.84.1.60

Sedikides, C., Meek, R., Alicke, M. D., & Taylor, S. (2014). Behind bars but above the

    bar: Prisoners consider themselves more prosocial than non-prisoners. *British*

    Journal of Social Psychology, *53*, 396-403. doi: 10.1111/bjso.12060

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in

    impression formation: A review of explanations. *Psychological Bulletin*, *105*,

    131-142. doi: http://dx.doi.org/10.1037/0033-2909.105.1.131

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological*

    Bulletin, 87, 245-251. doi.org/10.1037/0033-2909.87.2.245

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition, 131*, 159–

    171. doi: 10.1016/j.cognition.2013.12.005

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta*

    Psychologica, 47, 143-148. doi: 10.1016/0001-6918(81)90005-6

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*, 1–26. doi: 10.1111/1529-1006.001

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.

Tappin, B. M., & McKay, R. T. (2016). The Illusion of Moral Superiority. *Social Psychological and Personality Science*, advance online publication. doi: 10.1177/1948550616673878

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, *103*, 193–210. doi:10.1037/0033-2909.103.2.193

Taylor, S. E., Lerner, J. S., Sherman, D. K., Sage, R. M., & McDowell, N. K. (2003). Portrait of the self-enhancer: Well adjusted and well liked or maladjusted and friendless? *Journal of Personality and Social Psychology*, *84*, 165–176. doi:10.1037/0022-3514.84.1.165

Tenney, E. R., & Spellman, B. A. (2011). Complex social consequences of self-knowledge. *Social Psychological and Personality Science*, *2*, 343-350. doi:10.1177/1948550610390965

Tenney, E. R., Vazire, S., & Mehl, M. R. (2013). This examined life: The upside of self-knowledge for interpersonal relationships. *PloS One*, *8,* e69605. doi: 10.1371/journal.pone.0069605

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. *Advances in Experimental Social Psychology*, *21*, 181-228.

Tice, D. M., & Baumeister, R. F. (1990). Self-esteem, self-handicapping, and self-

   presentation: The strategy of inadequate practice. *Journal of Personality*, *58*, 443-

   464. doi: 10.1111/j.1467-6494.1990.tb00237.x

Van Damme, C., Hoorens, V., & Sedikides, C. (2016). Why self-enhancement provokes

   dislike: The hubris hypothesis and the aversiveness of explicit self-superiority

   claims. *Self and Identity*, *15*, 173–190. doi: 10.1080/15298868.2015.1095232

Varnum, M. E. (2015). Higher in status, (even) better-than-average. *Frontiers in*

   Psychology, *6*: 496. doi: 10.3389/fpsyg.2015.00496

Vazire, S. (2010). Who knows what about a person? The self–other knowledge

   asymmetry (SOKA) model. *Journal of personality and social psychology*, *98*,

   281-300. doi: http://dx.doi.org/10.1037/a0017908.

Van Lange, P. A. M., Joireman, J., Parks, C., Van Dijk, E. (2013). The psychology of

   social dilemmas: a review. *Organizational Behavior and Human Decision*

   Processes, 120, 125-140. doi:10.1016/j.obhdp.2012.11.003

Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-

   deception. *Behavioral and Brain Sciences*, *34*, 1-16. doi:

   http://dx.doi.org/10.1017/S0140525X10001354

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly*

   Journal of Experimental Psychology, 12, 129-140. doi**:**

   10.1080/17470216008416717

Wojciszke, B. (2005). Morality and competence in person-and self-perception. *European*

   Review of Social Psychology, *16*, 155-188. doi: 10.1080/10463280500229619

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A

metasynthesis. *Perspectives on Psychological Science*, *9*, 111-125. doi:

10.1177/1745691613518075

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A