MICHIGAN
ROSS SCHOOL OF BUSINESS

Working Paper

Social Comparison and Confidence: When Thinking You're Better than Average Predicts Overconfidence

Katherine A. Burson
Stephen M. Ross School of Business
at the University of Michigan

Richard P. Larrick
Duke University

Jack B. Soll
Duke University

Ross School of Business Working Paper Series
Working Paper No. 1016
July 2005

Social Comparison and Confidence:

When Thinking You're Better than Average Predicts Overconfidence

Richard P. Larrick

Duke University

Katherine A. Burson

University of Michigan

Jack B. Soll

Duke University

Manuscript under review. Please do not cite without permission.

RUNNING HEAD: Social Comparison and Overconfidence

Abstract

A common social comparison bias—the better-than-average-effect—is frequently described as psychologically equivalent to the individual judgment bias known as overconfidence. However, research has found “hard-easy” effects for each bias that yield a seemingly paradoxical reversal: Hard tasks tend to produce overconfidence but worse-than-average perceptions, whereas easy tasks tend to produce underconfidence and better-than-average effects. We argue that the two biases are in fact positively related because they share a common psychological basis in subjective feelings of competence, but that the “hard-easy” reversal is both empirically possible and logically necessary under specifiable conditions. Two studies are presented to support these arguments. We find little support for personality differences in these biases, and conclude that domain-specific feelings of competence account best for their relationship to each other.

Social Comparison and Confidence:

When Thinking You're Better than Average Predicts Overconfidence

How do people evaluate their own abilities? This was one of the basic questions underlying Festinger's original formulation of social comparison theory. Festinger (1954) proposed that people have a fundamental desire to evaluate their abilities, but often cannot test them against an objective standard. Therefore the abilities of others become the subjective reality that people use to reduce this uncertainty. Festinger largely portrayed this as a "cold" process (Goethals, Messick, & Allison, 1991), although with the recognition that there is a "unidirectional drive upward" in evaluations: People prefer to be better than others on a given ability, not worse.

A "hotter" version of social comparison theory emerged in the 1980s and 1990s that emphasized the importance of "downward comparisons" (Hakmiller, 1966; Wills, 1981) as a source of self-enhancement and positive affect (Alicke, 1985; Goethals, Messick, & Allison, 1991; Taylor, 1989; Taylor, Wayment, & Collins, 1993). Theories of downward comparison proposed that people seek and recall social comparison information favorable to themselves in order to hold the view that they are superior to others. Perhaps the most famous example of downward comparison is the "better than average" (BTA) effect (Goethals et al., 1991; Taylor & Brown, 1988), demonstrated in an early study which found that 90% of drivers believed that they were above average in driving ability (Svenson, 1981). Hundreds of studies have since replicated this pattern across a wide range of ability domains (Sedikides, Gaertner, & Toguchi, 2003; Windschitl, Kruger, & Simms, 2003; Chambers & Windschitl, 2004).

In the early 1990s, Goethals et al. (1991) observed that an important question not directly raised by Festinger (1954) was, "How *well* do people evaluate their own abilities?" They proposed that many social comparison evaluations are prone to systematic biases. For example, the BTA effect is typically interpreted as a bias because of the statistical unlikelihood that a majority of people would be above average. More careful studies have elicited a percentile estimate on an ability domain within a well-defined population. These studies have found that more than fifty percent of a population believes it is above the 50th percentile within that population, which is a statistical impossibility (e.g., Klar & Giladi, 1997).

The question "how well do people evaluate their abilities?" has also received attention from researchers in cognitive psychology in work on overconfidence. Decades of research have compared measures of subjective confidence with objective performance on a variety of tasks (e.g., Klayman, Soll, Gonzalez-Vallejo, & Barlas, 1999; Lichtenstein & Fischhoff, 1977; Yates, 1990). In one common paradigm, participants are given general knowledge questions with two possible answers. They are then asked to choose the answer they think is correct and to estimate the probability that they are right. Over many judgments, the average probability given can be compared to the actual proportion of choices that are correct. If people are insightful about their ability on these knowledge questions, we would expect that their expressed confidence would match the rate at which they answered questions correctly. A gap between average confidence and proportion correct indicates a lack of insight about ability. And, indeed, such a gap often occurs. Many of the original studies found that people were overconfident (OC): Average confidence exceeded average proportion correct.

Thus, the question, "how well do people evaluate their abilities?" has received similar answers in these two literatures: People systematically overestimate their abilities. And many researchers have noted this similarity. The better-than-average effect and overconfidence are frequently described as related—even identical—phenomena (e.g., Alba & Hutchinson, 2000; Daniel, Hirshleifer, & Subrahmanyam, 1998; Hoelzl & Rustichini, 2005; Juslin, Winman, & Olsson, 2000; Moore, Kurtzberg, Fox, & Bazerman, 1999). For example, one popular book on behavioral economics uses one phenomenon to illustrate the other: "[O]verconfidence often appears in the form of unrealistically high appraisals of one's own qualities versus those of others. The classic example of this tendency is a 1981 survey of automobile drivers in Sweden, in which 90% of them described themselves as above average drivers." (Belsky & Gilovich, 1999, p. 153-154). Intuitively, the connection between BTA and OC is appealing, and nonacademics also readily endorse the relationship between them (Yates, Lee, & Shinotsuka, 1996).

But is there, in fact, a direct relationship between the two biases? If one knew, for example, that Ann thought she was in the 80th percentile of performance on a geography quiz and Bill thought he was in the 50th percentile, would one be able to predict that Ann is more overconfident than Bill if she was asked to give a confidence level for the individual answers? Similarly, if one learned that sports quizzes elicit higher percentile estimates on average than do geography quizzes, would one be able to predict that sports quizzes elicit more overconfidence than do geography quizzes? Surprisingly, these direct questions about the relationship between BTA and OC have not been tested empirically.

The apparent similarity of BTA and OC has been cast in doubt in recent years when "hard-easy" manipulations in each literature were discovered to have opposite effects on the two biases. In the overconfidence literature, people have been found to be overconfident on "hard" questions but underconfident on "easy" questions (Brenner, 2003; Lichtenstein & Fischhoff, 1977), where hard and easy are defined in terms of actual performance. For example, if general knowledge questions are sorted based on the proportion of respondents who answered them correctly, then those questions that are frequently answered incorrectly will show overconfidence and those that are frequently answered correctly will show underconfidence. In contrast, researchers in the BTA tradition have found that "easy" tasks produce the BTA effect, and that "hard" tasks actually produce a worse than average (WTA) effect (Burson, Larrick, & Klayman, in press; Kruger, 1999; Moore & Kim, 2003). Thus, hard tasks appear to produce greater overconfidence but weaker BTA effects, whereas easy tasks produce less overconfidence but stronger BTA effects. If BTA and OC are related (even identical) phenomena, why does varying task difficulty have *opposite* effects on each bias? Is it a real reversal that is replicable within the same study, or is it an illusion created by looking across studies and methods? And, if it is real, why does it occur?

The studies presented in this paper explore the relationship between BTA effects and OC to identify their similarities and differences. We propose that BTA and OC are in fact fundamentally related, and therefore the common academic practice of linking them together is justified. The key factor uniting them is that a subjective sense of competence in a domain leads various subjective measures of ability in that domain to be highly correlated with each other (but only poorly correlated with objective measures of ability). We also propose that hard-easy manipulations do in fact have opposite effects on the two biases, and that this reversal does not represent a paradox—in fact, it is necessary under specifiable circumstances. We will show that changes in task difficulty can affect actual performance more than confidence. In addition, changes in task difficulty can inflate perceived percentile. When both occur, there must be a negative relationship between BTA and OC. In the next section, we provide a systematic analysis of the relationship between BTA and OC using a standard BTA measure (perceived percentile) and a standard overconfidence measure (the difference between average confidence and average proportion correct, which is sometimes called calibration-in-the-large (Yates, 1990)).

The relationship between BTA and OC. One of the fundamental results in both the BTA and OC literatures is that subjective perceptions are poorly correlated with objective measures (see Alba & Hutchinson, 2000, and Ehrlinger & Dunning, 2003, for recent reviews). An early and classic demonstration of this pattern was found by Oskamp (1965) who showed that forecast accuracy was poorly correlated with confidence. The weak relationship has now been found in many studies. We discuss the significance of this weak relationship first in the context of overconfidence. The poor correlation between objective and subjective measures leads to predictable patterns of bias depending on how data are conditioned (Erev, Wallsten, & Budescu, 1994; Soll, 1996). Figure 1 shows a stylized pattern of results from the literature on overconfidence. It depicts a weak linear relationship between confidence and proportion correct across a set of judgments, and plots the same relationship in two ways (Erev et al., 1994). The top panel shows the classic pattern of overconfidence: When average proportion correct is plotted on levels of confidence, there is a region of underconfidence on the lefthand side of the diagram, in which proportion correct falls above the identity line, and a region of overconfidence on the righthand side, in which proportion correct falls below the identity line. A typical result is that when people say they are 90% confident about their performance, they are correct only 70% of the time.

The bottom panel plots average confidence on proportion correct, once again showing the weak relationship between the two variables. The lefthand side of the diagram depicts the "hard" region in which proportion correct is low and confidence exceeds the identity line. Thus, "hard" questions tend to produce overconfidence. The righthand side of the diagram depicts the "easy" region in which proportion correct is high and average confidence falls short of it. "Easy" questions tend to produce underconfidence. In recent years, many scholars have attributed the hard-easy effect observed when items are sorted based on proportion correct to mean reversion (Erev et al., 1994; Dawes & Mulford, 1996; Juslin, Winman, & Olsson, 2000; Klayman et al., 1999).[1] While mean reversion can and does contribute to the hard-easy effect in overconfidence, it is important to note that psychological factors may contribute to these effects as well (Ayton & McClelland, 1997; Griffin & Varey, 1996; Klayman et al., 1999).

More recently, the poor correlation between subjective and objective measures has been demonstrated in the BTA literature (Ackerman, Beier, & Bowen, 2002; Ames & Kammrath, 2004; Burson et al., in press; Ehrlinger & Dunning, 2003; Krueger & Mueller, 2002; Kruger & Dunning, 1999). In these studies, actual percentile in a domain is measured by giving participants a test and then assigning them a percentile rank based on their proportion correct. This percentile rank is then compared to the participant's percentile estimate within the population performing that task. Figure 2 depicts the resulting relationship when actual percentile is plotted on perceived percentile (top panel) and vice versa (bottom panel). It is worth noting a difference between percentile calibration and confidence calibration. Actual percentile in an ability domain must necessarily average to 50, as shown in the top panel, whereas proportion correct has no corresponding constraint in the overconfidence literature, and the line in the top panel of Figure 1 could have any elevation. (Because actual percentile is a monotonically-increasing function of proportion correct, the line in the top panel of Figure 2 would have a similar slope if proportion correct were the dependent variable.)

Although actual percentile in an ability domain must average to 50, perceived percentile can average to well above or well below 50. Kruger (1999) has shown that perceived percentile varies directly with perceptions of absolute performance in a domain. Tasks on which a population's absolute performance is high tend to produce BTA effects, whereas tasks on which absolute performance is low tend to produce WTA effects (a pattern also demonstrated by Burson et al. (in press), Camerer and Lovallo (1999), and by Moore and Kim (2003)). The bottom panel of Figure 2 depicts the hard-easy effect for perceived percentile. Important to our argument is that several studies (Kruger, 1999; Moore & Kim, 2003) have manipulated perceived difficulty by varying actual difficulty, thereby changing the average proportion correct across conditions and mirroring how the "hard-easy" difference is operationalized in the overconfidence literature.[2]

Figures 1 and 2 summarize the main findings from the separate literatures on overconfidence and better-than-average effects. We now consider how the two measures will be related to each other. In the following analysis we will consider degrees of BTA and OC, where perceived percentile can range from worse-than-average to better-than-average effects, and calibration can range from underconfidence to overconfidence. Thus, if BTA and OC are positively correlated it simply means that as one increases in magnitude the other increases in magnitude, regardless of absolute magnitude.

Although objective and subjective measures are poorly correlated, we expect that related subjective measures will tend to be highly correlated. When individuals estimate their confidence in a performance and their percentile on a performance, they will tend to draw on similar evidence in assessing both: Memory of the recent performance, views of the self in that domain, and general feelings about the self. We therefore predict that confidence will be strongly related to perceived percentile, as depicted in the top panel of Figure 3. However, because proportion correct will be weakly related to perceived percentile, overconfidence will increase with perceived percentile. The difference between confidence and proportion correct is calculated in the bottom panel of Figure 3, and shows that as perceived percentile increases a region of diminishing underconfidence gives way to a region of increasing overconfidence.

Our basic prediction is that perceived percentile is positively related to greater degrees of overconfidence. Thus, we would expect that the answer to the earlier question, "If one knew that Ann thought she was in the 80th percentile of performance on a geography quiz and Bill thought he was in the 40th percentile, would one be able to predict that Ann is more overconfident than Bill if she was asked to give a confidence level for the individual answers?" is yes. We believe that demonstrating this empirical relationship would provide initial justification for the common practice of linking these constructs. However, we also want to explore the basis of this relationship. Thus, in the studies that follow we examine whether the link arises because of general individual differences based in personality that influence all perceptions of competence, such as differences in self-esteem and narcissism, or whether the link is due to domain-specific self-views that affect only domain-related perceptions of competence (Ehrlinger & Dunning, 2003).

If there is a positive relationship between perceived percentile and degree of overconfidence, and the analysis summarized in Figure 3 predicts that there will be one, why then does the positive relationship seem to reverse when task difficulty changes? We close this section by dissecting the seeming paradox of this reversal. Figure 4 is a modified version of Figure 3. At no loss of generality, Figure 3 was drawn to depict a "hard" task in which proportion correct was low, on average, in a population. These two lines are repeated in the top panel of Figure 4. However, the top panel of Figure 4 adds a new line that plots the average proportion correct for an easy task (which by definition in this case has a higher average proportion correct). The effect of making a task easier on overconfidence is shown in the bottom panel of Figure 4 as a shift of the entire line to the bottom right: The consequence is that the same level of perceived percentile will translate into less overconfidence.

However, manipulating proportion correct on a task will affect not just average overconfidence, but also average perceived percentile (Kruger, 1999). Figure 5 completes the analysis by depicting both effects simultaneously. The top panel shows the effect of a hard-easy manipulation on perceived percentile (shown as two different positions on the x-axis) and on overconfidence (shown as mapping the two different positions into different linear functions with the same slope but different constants). Making the task easier increases perceived percentile while simultaneously reducing overconfidence. This pattern is depicted in the bottom panel of Figure 5 as a pair of points linked by a downward sloping solid line.

There are two key features that lead the previously hypothesized positive relationship between BTA and OC (see Figure 5) to reverse in this situation. First, the difficulty manipulation affects perceived percentile: A task that leads to a high proportion correct leads to, on average, higher estimates of percentile. Second, the difficulty manipulation leads to a bigger change in proportion correct than in mean confidence. Thus, in moving from a hard task to an easy one, confidence increases but accuracy increases even more. Consequently, the OC associated with the hard task is attenuated with the easy task. The final result is that, compared to the hard task, the easy task yields higher BTA but lower OC.

We note that this pattern is not inevitable. Some difficulty manipulations could change mean confidence more than proportion correct. In this case, the hard-easy manipulation would not reverse the relationship between BTA and OC. In fact, it would yield a more positive relationship between BTA and OC. We consider the boundary conditions for the reversal at greater length in the Discussion and in a model in Appendix B. We simply note that our studies were designed to facilitate the reversal and thereby provide useful confirmation that the reversal is logically and empirically possible. However, we caution that this reversal need not hold across all “difficulty” manipulations.

We close this analysis with a brief consideration of an alternative operationalization of BTA. In the previous arguments, we have asked how perceived percentile is related to overconfidence. Perceived percentile is simply a subjective measure that, at the individual level, does not constitute a bias (one might actually be above average in an ability!). A bias can only be attributed at a population level when mean perceived percentile differs significantly from 50. In contrast, overconfidence is an individual level bias. A natural question is whether both biases could be calculated at the individual level and their relationship explored? For example, a measure of “overplacement” could be constructed by subtracting actual percentile from perceived percentile for each individual. This measure could then be correlated with overconfidence. However, the interpretation of the resulting relationship is problematic. Specifically, if the objective measures used in calculating overconfidence (i.e., proportion correct) and overplacement (i.e., actual percentile) are derived from the same performance, they will be monotonically-increasing functions of each other, and therefore highly correlated for purely mathematical reasons. Because the same term appears on both sides of the correlation between overconfidence and overplacement, they will be positively correlated for an artifactual reason (a more extended analysis is offered as part of the model presented in Appendix B).

The artifactual relationship between overplacement and overconfidence leads us to downplay exploring this relationship. In recent years, however, researchers have proposed ways around this artifact within the OC (Juslin et al., 2000; Klayman et al., 1999) and BTA (Krueger & Mueller, 2002) literatures, which is to measure both subjective and objective variables on different performances. Such “split-sample” methods remove the biasing effects of a shared term. At several points in the paper we test the overplacement-overconfidence relationship when we can use a split-sample method.

We note that the relationship between perceived percentile and overconfidence is not subject to the artifactual concern that plagues the overplacement-overconfidence relationship—that is, the same measure does not appear twice in the variables that are being correlated. It is an empirical question how strongly correlated perceived percentile, confidence, and proportion correct are with each other. Although we hypothesize that the two subjective measures will be more highly correlated than either subjective measure is with proportion correct, this hypothesis is capable of empirical falsification. These statistical considerations are one reason we focus our analysis on the relationship between perceived percentile and overconfidence. A second reason is that most social psychological studies of BTA effects do not assess objective ability at an individual level and do not measure overplacement (a tendency that has changed in recent years—see Footnote 2). Thus, our basic unit of analysis parallels that used in most social psychological studies of BTA effects, allowing for a more direct comparison between literatures.

The following studies were designed to test the hypothesized relationships between BTA and OC. Both studies used the same basic methodology. Participants were given a series of 10 questions within a specific domain and asked to give their best estimate for each question. For example, one domain was the year in which Nobel Prizes in literature were awarded to different authors. Answers were correct if they deviated from the truth by less than a fixed value designated by the experimenters (e.g., within 5 years of the truth). A set of 10 difficult and 10 easy questions was created within each domain simply by varying the stringency of the criterion for being correct (e.g., being within 5 years of the truth versus 30 years of the truth). After making an estimate on a question, participants estimated the probability that their answer was within the criterion. Their average confidence with a set of 10 items was then compared to the proportion correct in order to measure under-/overconfidence. Finally, participants estimated their percentile of performance for each set of 10 items. These constituted our two main measures in the following analyses.

Method

Participants

Study 1. Forty University of Chicago students were recruited with posted advertisements and were paid nine dollars for this 45-minute experiment.

Study 2. Thirty-five University of Michigan students were recruited from their introductory marketing class and received course credit for this 45-minute experiment.

Materials

In both studies, participants saw questions from five different domains. Each domain consisted of two subsets of 10 questions. These questions were drawn at random from a larger list of items that were representatively sampled for a domain. Each 10-question subset was presented in either a difficult or an easy version. However, the order of the 100 estimates was the same across participants, consisting of 10 questions from each of the five domains, followed by another 10 questions from each of the five domains. The order of difficulty was counterbalanced such that half the participants received the first five subsets of questions in the difficult version and the second five in the easier version. For the other half, the first five subsets were in the easier version and the second five in the difficult version.

Study 1 domains. Study 1 used five domains: college acceptance rates, dates of Nobel prizes, length of time pop songs had been on the charts, financial worth of richest people, and games won by hockey teams. The questions in these domains were selected randomly from the available information sources. Sample questions are provided in the appendix.

Study 2 domains. Study 2 employed a different set of five domains: University of Michigan student demographics, distances between campus landmarks, University of Michigan football scores, characteristics of marketing students, and local pizza delivery costs. Sample questions are provided in the appendix.

Procedure

In both studies, participants were told that they would be making a series of estimates about a range of topics. They were given a booklet containing 10 subsets of 10 estimates. The introduction of the booklet provided an example of the overall procedure using questions from an unrelated domain. In the next part of the booklet, participants read 10 pages, each devoted to a different subset of questions. For each subset of 10 questions, participants read an explanation of the required estimates, the criterion for being correct, and information about the mean of the sample and the range in which 90% of the sample fell. They then made an estimate for each question and provided confidence in the accuracy of that estimate. Finally, at the end of each subset, participants also indicated their predicted percentile standing.

After completing this section, participants in both studies answered questions about their mood and personality on the Self Esteem Scale (Rosenberg, 1965), Need for Closure (Webster &

Kruglanski, 1994), and Need for Cognition scales (Cacioppo, Petty, & Kao, 1984). Participants in Study 1 also completed the Positive and Negative Affect Schedule measure (Watson, Clark, & Tellegen, 1988)) and Need for Uniqueness (Snyder & Fromkin, 1977) scales. Participants in Study 2 also completed the Defensive Pessimism Scale (Norem & Cantor, 1986), Hypersensitive Narcissism Scale (Hendin & Cheek, 1997), and Narcissistic Personality Inventory (Raskin & Terry, 1988).

Results

Our first analysis explores the relationship between perceived percentile, average confidence, proportion correct, and overconfidence. Our unit of analysis is each set of 10 questions for each person, yielding 400 data points for Study 1 and 350 data points for Study 2. Figures 6 (Study 1) and 7 (Study 2) show the scatter plot and regression equations when average confidence, proportion correct, and their difference is plotted against perceived percentile. As expected, average confidence was strongly related to perceived percentile but proportion correct was only weakly related. Consequently, perceived percentile predicted degree of overconfidence.

Tables 1 and 2 extend this analysis by including additional variables. Equation 2 in each table adds a dummy variable for the difficulty manipulation (where 1 = hard criterion). As expected, more difficult domains significantly increase overconfidence. Including the difficulty manipulation increases the coefficient for perceived percentile, indicating that the difficulty manipulation, when uncontrolled, acts a suppressor variable on the relationship. Finally, additional dummy variables were added for domain and participant (omitting one domain and one participant in each analysis). These “fixed effects” tests control for variation attributable to domains and individuals. It may be seen that the basic relationship between perceived percentile and overconfidence remains unaffected. In both studies, a one point increase in perceived percentile translates into a .4 increase in overconfidence after controlling for domain differences and individual differences.

A second way to examine the relationship between perceived percentile and overconfidence is to examine means at the level of domain. The top panel of Figure 8 plots average overconfidence and average perceived percentile for the 10 domains in Studies 1 and 2. The relationship between the variables across domains is positive and strong (R-squared = .710, $F(1, 9) = 23.04$, $p = .001$). These analyses indicate that a one point increase in average perceived percentile in a domain translates into a 1.33 increase in overconfidence. Because this analysis is conducted at the level of a domain, average perceived percentiles that deviate from 50 can regarded as an under- or overplacement bias. This level of analysis indicates a strong positive relationship between overplacement and overconfidence.

To examine the hard-easy reversal, we analyzed perceived percentile and overconfidence separately in a repeated-measures ANOVA, with difficulty and domain as within-participant variables. The means for this analysis are presented in the first two columns of Tables 3 (Study 1) and 4 (Study 2). In Study 1, there was a significant main effect of domain ($F(4, 152) = 9.25$, $p < .001$) and of difficulty ($F(1, 38) = 19.31$, $p < .001$) on perceived percentile; a marginal main effect of domain ($F(4, 152) = 2.33$, $p = .059$) and a main effect of difficulty ($F(1, 38) = 666.12$, $p < .001$ on overconfidence; and an interaction of domain and difficulty for on overconfidence ($F(4, 152) = 28.42$, $p < .001$). In Study 2, there was a significant main effect of domain ($F(4, 128) = 3.17$, $p = .016$) and of difficulty ($F(1, 32) = 13.26$, $p = .001$) on perceived percentile; a significant main effect of domain ($F(4, 132) = 11.29$, $p < .001$) and of difficulty ($F(1, 33) = 69.38$, $p = .001$ on overconfidence; and no interactions ($F < 1$). As the overall means in each study show, the difficulty manipulation significantly increased the degree of overconfidence while significantly decreasing perceived percentile.

Tables 3 and 4 also provide the means for confidence and proportion correct for each domain, which were analyzed in repeated-measures ANOVA, with difficulty and domain as within-participant variables. In both studies, there were significant main effects of difficulty and domain on proportion correct and on confidence (ps < .001), as well as a significant difficulty by domain interaction (ps < .05), indicating that the difficulty effect was stronger in some domains. As expected, the easy conditions produced more confidence and higher proportion correct than did hard conditions. The means in Tables 3 and 4 make clear the important underlying cause of the hard-easy reversal: The difficulty manipulation produced a difference in confidence across conditions, but an even larger difference on proportion correct. As we note in the Discussion—and analyze at greater length in Appendix—it is this pattern that drives the hard-easy reversal across BTA and OC.

This difficulty-induced reversal is depicted graphically in the bottom panel of Figure 8. The top panel shows the generally strong, positive relationship between perceived percentile and overconfidence. The bottom panel plots the means for overconfidence and perceived percentile within the hard and easy versions of the 10 domains used in Studies 1 and 2. The means for the hard version of a domain are linked to the means for the easy version of the same domain with a solid line. This graph reveals that the strong, positive relationship in the top panel masks a second relationship: Within a domain, there is a pronounced inverse relationship between perceived percentile and overconfidence, consistent with the pattern anticipated in Figure 5. This pattern held for 10 out of 10 domains depicted in the bottom panel of Figure 8 (also in the means in Tables 3 and 4).

Overall, we find that perceived percentile is strongly related to overconfidence. The initial analyses (Figures 6 and 7) demonstrate that it is the strong relationship between perceived percentile and confidence that drives this relationship. As expected, however, the positive relationship can be reversed when difficulty is manipulated within a domain (bottom Figure 8).

We conclude our analysis with an exploration of the factors that lead perceived percentile and overconfidence to be highly correlated. We explore three levels of explanation: General personality differences, domain-specific self-views, and item-specific influences. An individual-level analysis is presented in Table 5 that allows us to assess the influence of personality differences. In this analysis, average perceived percentile and average overconfidence are first calculated for an individual across the 10 subdomains before performing any correlations. For example, a person might have average perceived percentiles across the 10 subdomains of 25, 35, 40, 70, 50, 35, 70, 40, 55, 65, yielding an overall individual-level average of 48.5. The first line shows the high correlation between average perceived percentile and average overconfidence when calculated at the individual level. However, the next section of the table shows that these individual level measures have only weak relationships with the personality measures. In Study 1, there was some evidence that Need for Closure correlated with perceived percentile and overconfidence. However, this pattern did not replicate in Study 2. None of the other relationships was even suggestive. Because these personality measures were not correlated reliably with either perceived percentile or overconfidence, they do not represent a plausible explanation for the observed relationship between perceived percentile and overconfidence.

These weak personality relationships are consistent with other research in this area. Jonsson and Alwood (2003) found no relationship between realistic confidence judgments and Need for Cognition. Ehrlinger and Dunning (2003) found no correlation between performance estimates and Self Esteem or PANAS. Interestingly, Ames and Damrath (2004) have found a correlation between narcissism and overconfidence in a social judgment task, but this was with a scale of their own design. We find no similar pattern when using more common narcissism scales on tasks that have less social content. Our conclusion is that standard personality measures do little to predict level of perceived percentile or degree of overconfidence.

Table 6 continues an individual level analysis by examining overconfidence, overplacement, and their components. As in Table 5, each measure is calculated for an individual by averaging together his or her responses for the 10 subdomains. Practically, this level of analysis is akin to creating a quasi-personality scale by aggregating measures across several domains of knowledge to get a reliable measure of these tendencies. The results show that many of the basic relationships hold across individuals: Notably, average perceived percentile and average confidence are strongly correlated in both studies, and both are more strongly related than either subjective measure is to an objective measure. At this level of analysis, perceived percentile predicts overconfidence well in Study 1, but only weakly in Study 2. (We do not discuss all the relationships in Table 6 but provide them because researchers in different literatures may find different combinations of interest.)

We note that there is a very strong relationship between overplacement and overconfidence at this level of analysis. In the introduction, we suggested that the relationship between under-/overplacement and under-/overconfidence is interesting but problematic. It is interesting because it compares one measure of individual-level bias with another. However, for the reasons described in the introduction (and elaborated further in footnotes 1 and 2), the relationship is problematic if the same performance is used to calculate both actual percentile and proportion correct. The issue is that the same variable (slightly transformed) therefore appears in the estimates of both under-/overplacement and under-/overconfidence, thereby inducing a positive correlation between the measures. Not surprisingly, Table 6 shows that actual percentile and proportion correct are extremely highly correlated, which confirms that interpreting the high overplacement-overconfidence relationship is problematic. Thus, we conducted a split-samples analysis predicting overconfidence on easy subdomains from overplacement on hard subdomains (.339 in Study 1, .495 in Study 2), and vice versa (.475 in Study 1, .194 in Study 2).These individual level analyses suggest that, even after removing the artifactual bias in the overplacement-overconfidence relationship, there remains an average correlation of approximately .35 between the two measures. Overplacement is positively related to overconfidence.

Table 7 presents a new type of analysis that examines the influence of domain-specific self-views (Ehrlinger & Dunning, 2003; Markus, 1977). The basic assumption underlying this test is that people hold views of themselves tied to specific ability domains—"I know a lot about literature," "I'm not a hockey fan," and so on. These domain-specific self-views lead perceived percentile and overconfidence to move together within a domain more strongly than between domains. The domain-specificity proposal implies that perceived percentile measured in one domain (e.g., literature) will correlate more strongly with overconfidence in the same domain than in another domain (e.g., hockey). However, given our measures, one reason there may be a high correlation between perceived percentile and overconfidence within a subdomain is that both judgments are based on reactions to the same items—"I'm below average and not very confident about these particular literature questions, but I would be more knowledgeable about other literature questions." Our data allow us to separate the influence of domain-specific self-views from reactions to specific items by looking at the correlation within a domain but across difficulty and thereby question subsets. If these within-domain-and-across-difficulty correlations are higher than across-domain correlations, it indicates the influence of a domain-specific self-view that exerts an influence across different instantiations of the same domain.

For this analysis, we calculated four sets of correlations that compare the correlation between perceived percentile and overconfidence a) within a subdomain (i.e., within a domain and within difficulty using the same set of items), b) within a domain but across difficulty, c) across domains but within difficulty, and d) across domains and across difficulty. The average correlation is shown in Table 6, by study. Not surprisingly, the average correlation between perceived percentile and overconfidence within a subdomain is fairly high (this correlation analysis merely restates the regression results in Tables 1 and 2). The correlations calculated within domain but across difficulty are not significantly smaller than the within subdomain correlations (*ns*), but are significantly larger than the across-domain correlations (*p*s $< .01$).[3] We believe this shows that domain-specific self-views do exert an influence across separate measures drawn from the same domain. Thus, one can predict overconfidence from perceived percentile better when the measures come from the same domain than when they come from different domains. Finally, the fact that across-domain correlations are positive—and in Study 1 averaged above .20—does suggest a general individual difference that underlies perceived percentile and overconfidence.

Our final analysis is of under-/overplacement, which is calculated here as perceived percentile minus actual percentile within a subdomain. The analysis in Table 8 presents the same grouping of correlations as in Table 7, except it replaces perceived percentile with overplacement. Table 8 shows that within-subdomain correlations are quite high, and higher than the comparable correlations in Table 7. This pattern is expected because the correlations are inflated by the inclusion of a common variable. The second row of Table 8 is revealing because it provides within-domain correlations that have no common variable problem. The within-domain-across-difficulty correlations are, not surprisingly, lower than the within-subdomain correlations (*p*s < .01). They are also higher than the across-domain correlations, marginally in Study 1 (p = .15) and significantly in Study 2 (p = .02 in Study 2), indicating that domain-specificity provides a modest enhancement of the overplacement-overconfidence relationship when analyzed at the domain level.

Discussion

These analyses show that better-than-average effects and overconfidence are fundamentally related to each other, yet their relationship can be reversed. Overall, there is a positive relationship between overconfidence and better-than-average effects. This relationship holds across individuals and across domains. Thus, the answer to the question, “If one knew that Ann thought she was in the 80th percentile of performance on a geography quiz and Bill thought he was in the 50th percentile, would one be able to predict that Ann is more overconfident than Bill if she was asked to give a confidence level for the individual answers?” is yes. Similarly, knowing that one domain produces a higher average perceived percentile than another allows one to predict it will produce higher average overconfidence. The positive relationship justifies the common academic practice of treating each tendency as related to the other.

The positive relationship between BTA and OC arises because subjective assessments of confidence and percentile estimates are highly correlated with each, but each is poorly correlated with actual performance. Several split-sample tests indicate that overplacement—measured as the difference between actual and perceived percentile at the individual level—predicts overconfidence We find some evidence that these relationships are stronger within domains, indicating that domain-specific self-views help drive the relationship. By comparison, we find little evidence that personality measures help explain the relationship between percentile estimates and overconfidence.

A seemingly paradoxical reversal of hard-easy effects for the better-than-average effect and overconfidence has been observed across different studies. Hard tasks appear to produce worse-than-average effects but overconfidence; easy tasks appear to produce better-than-average effects but underconfidence. Our results show that this apparent reversal is a real one—it is empirically possible. Moreover, it is not a paradox. It is possible to specify the conditions that drive it. In our studies, a difficulty manipulation within a domain produced larger changes in average proportion correct than in average confidence. The same manipulation tended to change perceived percentiles systematically: Hard tasks yielded lower perceived percentiles on average than did easy tasks (Kruger, 1999). Thus, increased task difficulty within a domain tended to decrease the BTA effect while increasing OC. This reversal, however, will not always occur. Some difficulty manipulations will in fact strengthen the positive relationship between BTA and OC. The next section provides an intuitive sketch of the conditions that moderate whether the BTA-OC relationship is positive or negative.

Conditions for the "hard-easy" reversal. The relationship between BTA and OC can be clarified by comparing three different analyses. One analysis compares participants responding to the same topic and difficulty level. For convenience, participants can be thought of as being two types based on a median split of percentile estimates. Because subjective measures are highly correlated, those who estimate high percentiles will tend to be the most confident, and also the most overconfident (because accuracy is imperfectly correlated with the subjective measures). This type of analysis is depicted in Figures 3 and 4. A second analysis compares participants responding to different topics and the same difficulty level. Across topics, subjective measures are highly correlated with each other and weakly correlated with accuracy. Thus the topics with the greatest percentile estimates (and hence also the greatest BTA) will tend to also exhibit high confidence, accuracy will be regressive, and consequently the high BTA topics will tend to be the same ones as those that exhibit the most OC (e.g., see the top panel of Figure 8). In both of these analyses, BTA and OC are positively correlated.

Finally, consider a third analysis in which different levels of difficulty are compared for the same topic domain. The easier version consistently has both higher percentile estimates and higher accuracy than does the harder version. However, the easy version may or may not produce greater overconfidence. This depends on whether mean confidence changes more or less than mean accuracy across the difficulty manipulation. In our studies, mean confidence was consistently higher in the easy version than in the hard version, but the change in mean confidence was less than the change in mean accuracy. Consequently, the easy version yielded greater BTA but less OC than did the hard version of a domain. Thus, in these circumstances, BTA and OC are negatively correlated. However, difficulty manipulations that affect mean confidence more than mean accuracy would not produce a negative correlation. These conditions are considered at greater length in Appendix B. The next section proposes difficulty manipulations that would yield a positive relationship between BTA and OC.

Undoing the "hard-easy" reversal. Recent research in the BTA literature has manipulated task difficulty in a way that manipulates both perceptions of performance and actual task performance (Burson et al., in press; Kruger, 1999; Moore & Kim, 2003; Windschitl et al, 2003). This was of interest in the BTA literature because mean changes in actual task performance at the population level is irrelevant to judgments of perceived percentile—mean percentile in a population is always 50 regardless of the level of absolute performance in the population. For these types of manipulations, average performance (e.g., proportion correct) tends to change more than average perceptions (e.g., confidence), yielding the reversal documented here. However, other difficulty manipulations could change perceptions more than performance. If tasks were designed to influence perceptions of performance more than actual performance, the hard-easy reversal would not occur. We offer one example and sketch two other possibilities.

In his classic study on the relationship between confidence and accuracy, Oskamp (1965) manipulated the amount of information people had available on which to base forecasts. He found that confidence increased with the number of cues available. However, accuracy of the forecasts did not (that is, the cues were not particularly diagnostic). We believe that Oskamp's paradigm could be used to manipulate feelings of difficulty without actually changing performance. Providing increasing amounts of information would tend to inflate both confidence and perceived percentile without actually improving performance (especially if the information is not diagnostic). This difficulty manipulation, unlike the current studies, would reinforce the underlying positive relationship between BTA and OC.

Several other paradigms suggest that perceptions of ability can be influenced without influencing actual performance. For example, Fox and colleagues have shown that the sequence of tasks can affect perceptions of ability (Fox & Tversky, 1995; Fox & Weber, 2002). When participants confront a comparatively difficult task before a target task, they increase their sense of competence on the target task; a comparatively easy task has the opposite effect. Of course, actual performance did not change on the target task. Schwarz and his colleagues (e.g., Schwarz, Bless, Strack, Klumpp, Rittenauer-Schatka, & Simons, 1991) have shown that asking for 10 reasons why something is true versus 2 reasons made the retrieval of reasons either difficult or easy, respectively. A similar manipulation applied to judgments about the self could affect perceived competence without changing performance on a related task. Both paradigms suggest ways in which future studies might manipulate perceptions of difficulty independently of actual difficulty and thus eliminate hard-easy reversals.

Conclusion

Social comparison theory was one of the first psychological theories to consider how people evaluate their own abilities, which ultimately led to the question of how well they did it (Goethals et al., 1991). Pervasive better-than-average effects and judgmental overconfidence suggest that people are biased in these ability assessments. The current results confirm that these two judgments of ability are closely related. Individuals who believe they are better than average are also more likely to be overconfident. And domains that produce better-than-average effects also produce greater overconfidence. Across many ways of analyzing the relationship between percentile judgments and overconfidence, the following relationship hold within a knowledge domain: The higher one's assessment of ability relative to others, the more likely one is to be overconfident when making judgments related to that domain. This robust relationship justifies the common practice of treating better-than-average effects and overconfidence as closely related phenomena.

These two assessments of ability, however, need not always show a positive relationship. Task difficulty can lead the two assessments to have a negative relationship: Higher assessments of ability relative to others will be accompanied by less overconfidence. This effect of task difficulty is both empirically demonstrable and logically explained. However, it will occur under only limited but specifiable circumstances. We hope that research on social comparison and on decision making will benefit from understanding when better-than-average effects and overconfidence will occur together and when they will diverge.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33*, 587-605.

Alba, J. W., & Hutchinson, J. W. (2000). Knowledge calibration: What consumers know and what they think they know. *Journal of Consumer Research, 27*, 123-156.

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

Ames, D., & Kammrath, L. (2004). Mind-reading and metacognition: Narcissism, not actual competence, predicts self-estimated ability. *Journal of Nonverbal Behavior, 28*, 187-209.

Ayton, P., & McClelland, A. G. R. (1997). How real is overconfidence. *Journal of Behavioral Decision Making, 10*, 153-285.

Belsky, G., & Gilovich, T. (1999). *Why smart people make big money mistakes and how to correct them: Lessons from the new science of behavioral economics*. New York, NY: Fireside.

Brenner, L. (2003). A random support model of the calibration of subjective probabilities. *Organizational Behavior and Human Decision Processes, 90*, 87-110.

Burson, K. A., Larrick, R. P., & Klayman, J. (in press). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*.

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

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306-307.

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

Daniel, K., Hirshleifer, D., & Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance, 53*, 1839-85.

Dawes, R. M. & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment, or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes*, *65*, 201-211.

Ehrlinger, J., & Dunning, David (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5-17.

Erev, I., Wallsten, T. S. & Budescu, D. V. (1994). Simultaneous over- and under-confidence: The role of error in judgment processes. *Psychological Review, 101*, 519-528.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception & Performance, 3*, 552-564.

Fox, C. R., & Tversky, A. (1995). Ambiguity aversion and comparative ignorance. *The Quarterly Journal of Economics, 110*, 585-603.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506-528.

Goethals, G. R., Messick, D. M., & Allison, S. T. (1991). The uniqueness bias: Studies of constructive social comparison. In J. Suls and T, A. Wills (Eds.), *Social comparison: Contemporary theory and research* (pp. 149-176). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

Griffin, D. W., & Varey, C. A. (1996). Towards a consensus on overconfidence. *Organizational Behavior and Human Decision Processes* , *65*, 227-231.

Hakmiller, K. L. (1966). Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology, 2*(Supplement 1), 32-39.

Hendin, H.M., & Cheek, J.M. (1997). Assessing hypersensitive narcissism: A re-examination of Murray's narcissism scale. *Journal of Research in Personality, 31*, 588-599.

Hoelzl, E., & Rustichini, A. (2005). Overconfident: Do you put your money on it? *Economic Journal, 115*, 305-318.

Jonsson, A. C., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences, 34*, 559-574.

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 1038-1052.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107*, 384-396.

Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology, 73*(5), 885-901.

Klayman, J., Soll, J. B., Gonzalez-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*, 216-247.

Krueger, J., & 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 and Social Psychology, 82*, 180-188.

Kruger, J. (1999). Lake Wobegon be gone! The "Below-Average Effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*, 221-232

Kruger, J., & 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*, 1121-1134.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? The calibration of probability judgments. *Organizational Behavior and Human Performance*, *20*, 159-183.

Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality & Social Psychology, 35*, 63-78.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85*, 1121-1135.

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.

Norem, J. K., & Cantor, N. (1986). Anticipatory and post hoc cushioning strategies: Optimism and defensive pessimism in "risky" situations. *Cognitive Therapy and Research, 10*, 347-362.

Oskamp, S. (1965). Overconfidence in case-study judgments. *The Journal of Consulting Psychology, 29*, 261–265.

Pfeifer, P. E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes, 58*, 203-213.

Raskin, R., & Terry, J. (1988). A principal-components analysis of the Narcissitic Personlaity Inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology, 54*, 890-902.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology*, 61, 195–202.

Sedikides, C. Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84*, 60-79.

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

Soll, J. B. (1996). Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes, 65*, 117-137.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica, 47*, 143-148.

Taylor, S. E. (1989). *Positive illusions: Creative self-deception and the healthy mind*. USA: Basic Books.

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

Taylor, S. E., Wayment, H.A., & Collins, M. A. (1993). Positive illusions and affect regulation. In D. M. Wegner, J. W. Pennebaker et al. (Ed.), *Handbook of mental control* (pp. 325-343). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of the brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 44*, 1063-1070.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in causal undertainty. *Journal of Personality and Social Psychology, 67*, 1049-1062.

Wills, T. A. (1981). Downward comparison principles in social psychology. *Psychological Bulletin, 90*, 245-271.

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

Yates, J. F. (1990). *Judgment and Decision Making*. Englewood Cliffs, NJ: Prentice Hall.

Yates, J. F., Lee, J., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes, 65*, 138-147.

Author Notes

Richard P. Larrick and Jack B. Soll, Fuqua School of Business, Duke University; Katherine A. Burson, Ross School of Business, University of Michigan.

Initial results were reported at the Behavioral Decision Research in Management Conference held at Duke University in May, 2004. Study 1 uses data from Study 2 of Burson et al. (in press), but provides new and more extensive analyses; means for perceived percentile for some domains are reported in Burson et al. (in press).

Questions regarding this research may be directed to Rick Larrick (larrick@duke.edu), Katherine Burson (kburson@umich.edu), or Jack Soll (jsoll@duke.edu).

Footnotes

[1] The hard-easy effect in overconfidence is inevitable for some methods of sorting questions into hard and easy categories (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin et al., 2000; Klayman et al., 1999). Part of the problem is that the independent variable (accuracy) and dependent variable (overconfidence = confidence – accuracy) are bound to be correlated because the same measure of accuracy shows up in both halves of the equation. Error in the accuracy measure guarantees the effect (see Gigerenzer et al., 1991; Juslin et al., 2000; Klayman et al., 1999). Some versions of the hard-easy effect hold up to statistical control. For example, individuals who are less accurate on one set of questions (i.e., the questions are *hard* for them) are more overconfident on a different set of questions on the same topic (Klayman et al., 1999).

[2] It is worth noting that the "hard-easy" effect depicted in the bottom panel of Figure 2 differs from that traditionally discussed in the overconfidence literature because average proportion correct is manipulated for a whole population in an ability domain. Thus, the hard-easy effect on percentiles is manifested as an increase in perceived percentile (i.e., the line in the bottom panel of Figure 2 shifts upward). Kruger and Dunning (1999) conducted an analysis of percentile estimates that more directly parallels the original hard-easy analysis in the overconfidence literature (see Footnote 1) when they sorted participants based on their actual percentile and examined their percentile estimates. They found that the worst performers (as measured by actual percentile) overestimated their percentile, whereas the best performers underestimated theirs. In a direct parallel to the overconfidence literature, this pattern has been reinterpreted as a necessary effect of regression (Ackerman et al., 2002; Burson et al., in press; Krueger & Mueller, 2002).

[3] Mean differences in the correlations were tested after performing an *r*-to-*z* transformation on the individual correlations, taking an average, and then performing a set of planned, non-orthogonal contrasts ($df = 96$) on the means of the transformed correlations. The contrasts compared Row 1 vs. Row 2, and Row 2 vs. Rows 3 and 4.

Appendix A: Stimuli

University of Chicago Quizzes (harder criteria in brackets)

College Acceptance

In this section, you will estimate the acceptance rate of colleges. The acceptance rate is the rate of acceptance to applications. Therefore, if a college lets in only half of the people who apply, it would have a 50% acceptance rate. You should try to be accurate within 20% [5%] of the truth. These 10 colleges were selected randomly from the top 51 colleges of 2001 published by US News and World Report. Within the 20 colleges in this packet, the average acceptance rate is 44% and 90% of the colleges fall between 19% and 77%.

	What is the acceptance rate (%) of this college?		How confident are you that this estimate is within 20% [5%] of being correct?
1	Rice University		%
2	University of Michigan--Ann Arbor		%
3	New York University		%
4	Rensselaer Polytechnic Inst.		%
5	University of Washington		%
6	Boston College		%
7	Vanerbilt University		%
8	University of Rochester		%
9	University of California--Irvine		%
10	Duke University		%
11	Cornell University		%
12	Wake Forest University		%
13	Pennsylvania State University (Penn State)		%
14	University of Wisconsin, Madison		%
15	Georgetown University		%
16	University of California, Berkeley		%
17	Massachusetts Institute of Technology (MIT)		%
18	Johns Hopkins University		%
19	Dartmouth College		%
20	California Institute of Technology (Cal Tech)		%

Nobel Prizes

In this section, you will estimate the year in which particular people received the Nobel Prize in Literature. You should try to be accurate within 30 [5] years of the truth. These 10 Nobel Laureates were selected randomly from the 100 Nobel Laureates in Literature. Within the 20 Laureates in this packet, the average year of the Nobel Prize is 1949 and 90% of the Laureates fall between 1921 and 1985.

	What year did this person receive the Nobel Prize?		How confident are you that this estimate is within 30 [5] years of being correct?
1	Claude Simon		%
2	Wladyslaw Stanislaw Remont		%
3	Thomas Stearns Eliot		%
4	Patrick White		%
5	Pablo Neruda		%
6	Romain Rolland		%
7	Johannes Vilhelm Jensen		%
8	Anatole Fronce		%
9	Sinclair Lewis		%
10	William Faulkner		%
11	Ivo Andric		%
12	Frans Eemil Sillanpaa		%
13	Elias Canetti		%
14	Albert Camus		%
15	William Butler Yeats		%
16	Juan Ramon Jimenez		%
17	Pearl Buck (pen name Pearl Walsh)		%
18	Hermann Hesse		%
19	Singrid Undset		%
20	Kenzaburo Oe		%

Pop Songs

In this section, you will estimate the number of weeks that a particular single has been on the charts. You should try to be accurate within 7 weeks of the truth. These 10 hit singles were selected at random from the top 20 singles listed on the Billboard chart for March 31, 2001. Within the 20 singles in this packet, the average number of weeks is 14 and 90% of the singles fall between 3 and 24 weeks.

	How many weeks has this single been on the charts?		How confident are you that this estimate is within 7 weeks of being correct?
1	"Jaded", Aerosmith		%
2	"Butterfly", Crazy Town		%
3	"Survivor", Destiny's Child		%
4	"Don't Tell Me", Madonna		%
5	"Hanging By a Moment", Lifehouse		%
6	"Promise", Jagged Edge		%
7	"Put It On Me", Ja Rule Featuring Lil'Mo & Vita		%
8	"It's Over Now", 112		%
9	"Get Over Yourself", Eden's Crush		%
10	"I Hope You Dance", Lee Ann Womack		%
11	"All For You", Janet		%
12	"Angel", Shaggy Featuring Rayvon		%
13	"Stutter", Joe Featuring Mystikal		%
14	"South Side", Moby Featuring Gwen Stefani		%
15	"Love Don't Cost a Thing", Jennifer Lopez		%
16	"Again", Lenny Kravitz		%
17	"Nobody Wants To Be Lonely" Rick Martin With Christina Aguilera		%
18	"Thank You", Dido		%
19	"Crazy", K-Ci & JoJo		%
20	"If You're Gone", matchbox twenty		%

Financial Worth

In this section, you will estimate the worth of the U.S.'s richest people. You should try to be accurate within $10 [$1] billion of the truth. These 10 people were selected at random from the top 50 richest people in the U.S. as listed by Forbes for the year 2000. Within the 20 people in this packet, the average worth is $16 billion and 90% of the people fall between $5 and $58 billion.

	What is the financial worth of this person—in billions?		How confident are you that this estimate is within 10 [1] billion of being correct?
1	John Werner Kluge (Metromedia)		%
2	Lawrence Joseph Ellison (Oracle Corp.)		%
3	Henry Samueli (Broadcom)		%
4	Jeffrey P. Bezos (Amazon.com)		%
5	James Goodnight (Software)		%
6	S. Robson Walton (Wal-Mart stores)		%
7	Paul Gardner Allen (Microsoft)		%
8	Edward Crosby Johnson III (Fidelity)		%
9	Daniel Smith (Fiber optics)		%
10	Gururaj E. Deshpande (Fiber optics)		%
11	Charles Ergen (Satellite Television)		%
12	William H. Gates III (Microsoft Corp.)		%
13	David Filo (Yahoo!)		%
14	Craig O. McCaw (McCaw Cellular)		%
15	Thomas J. Pritzker (Inheritance)		%
16	Steven Ballmer (Microsoft)		%
17	Theodore W. Waitt (Gateway 2000)		%
18	Sumner M. Redstone (Viacom)		%
19	Henry T. Nicholas (Broadcom)		%
20	Sanjiv Sidhu (Software)		%

Hockey

In this section, you will estimate the number of wins that National Hockey League hockey teams have as of April 6, 2001. You should try to be accurate within 20 [5] wins of the truth. These 10 teams were selected at random from the 30 teams in the league. Within the 20 teams in this packet, the average number of wins is 34 and 90% of the teams fall between 24 and 46.

	How many wins does this hockey team have so far?		How confident are you that this estimate is within 20 [5] wins of being correct?
1	New Jersey Devils		%
2	San Jose Sharks		%
3	Washington Capitals		%
4	Mighty Ducks of Anaheim		%
5	Carolina Hurricanes		%
6	Detroit Red Wings		%
7	Pittsburgh Penguins		%
8	Chicago Blackhawks		%
9	Boston Bruins		%
10	Edmonton Oilers		%
11	St. Louis Blues		%
12	Montreal Canadiens		%
13	Minnesota Wild		%
14	Calgary Flames		%
15	New York Rangers		%
16	Florida Panthers		%
17	Atlanta Thrashers		%
18	Vancouver Canucks		%
19	Columbus Blue Jackets		%
20	Buffalo Sabres		%

University of Michigan Knowledge Quiz (harder criteria in brackets)

University of Michigan Student Demographics

In this section, you will estimate the percent of University of Michigan students who are in a particular category as of Fall 2003. You should try to be accurate within ± 15 [5] percentage points of the truth. Within the 20 categories in this packet, the average percent of students is 46% and 90% of the percents fall between 17 and 84.

	What **percent** of UM students are in each category?		How confident are you that this estimate is within ± **15 [5] percentage points** of being correct?
1	Asian American freshman		%
2	freshman applicants accepted		%
3	freshman with a 4.0 high school GPA		%
4	freshman who drop-out		%
5	undergraduates who enroll in graduate school		%
6	out-of-state students		%
7	graduates students		%
8	freshmen who live in residence hall		%
9	undergraduates who enter the Greek System		%
10	undergraduates enrolled in LS&A		%
11	accepted freshman applicants who enroll		%
12	Caucasian freshman		%
13	graduates who are employed within 9 months		%
14	minority undergraduates		%
15	freshmen who receive AP/IB credits		%
16	students with a 3.9 high school GPA or higher		%
17	of transfer applicants, those accepted		%
18	undergraduates enrolled in the School of Engineering		%
19	freshmen applicants who score in the middle 50^{th} percentile range on the SATs		%
20	students who graduate within five years		%

Distances Between University of Michigan Points of Interest

In this section, you will estimate the distances (in tenths of miles) between University of Michigan points of interest. You should try to be accurate within ± .5 (five-tenths) [.2] of a mile of the truth. Within the 20 distances in this packet, the average distance between points is 1 mile and 90% of the distances fall between .5 and 2.6 miles.

	What is the **distance** (in tenths of miles) between each of these two points?		How confident are you that this estimate is within ± **.5 [.2] of a mile** of being correct?
1	Angell Hall to Stockwell Hall		%
2	Graduate Library to West Quad		%
3	Baits to Markley Hall		%
4	Stockwell Hall to Michigan Union		%
5	Yost Arena to Michigan Theatre		%
6	Ulrich's Bookstore to Media Union		%
7	Michigan Book & Supply to Business School		%
8	Media Union to Stadium		%
9	East Hall to Espresso Royale		%
10	Michigan Theatre to Chemistry Building		%
11	Angell Hall to Bursley		%
12	Angell Hall to Michigan League		%
13	Bursley Hall to Michigan League		%
14	Graduate Library to University Hospital		%
15	Lane Hall to University Hospital		%
16	Michigan Union to Undergraduate Library		%
17	East Quad to Michigan Book & Supply		%
18	Michigan Theatre to Undergrad Library		%
19	Bschool to Yost Arena		%
20	East Hall to Michigan Theatre		%

University of Michigan Football Games

In this section, you will estimate the number of points scored by either Michigan or their opponent during the 2003-2004 football season. You should try to be accurate within ± 10 [3] points of the truth. Within the 20 games included in this packet, the average points of teams is 28 and 90% of the scores fall between 10 and 45.

	How many **points** did the team score in this game?		How confident are you that this estimate is within ± **10 [3] points** of being correct?
1	Michigan's points against Central Michigan		%
2	Michigan's points against Notre Dame		%
3	Oregon's points against Michigan		%
4	Michigan's points against Indiana		%
5	Michigan's points against Minnesota		%
6	Illinois's points against Michigan		%
7	Michigan's points against Purdue		%
8	Michigan State's points against Michigan		%
9	Ohio State's points against Michigan		%
10	Michigan's points against Southern California		%
11	Central Michigan's points against Michigan		%
12	Michigan's points against Houston		%
13	Indiana's points against Michigan		%
14	Michigan's points against Iowa		%
15	Minnesota's points against Michigan		%
16	Michigan's points against Illinois		%
17	Purdue's points against Michigan		%
18	Northwestern's points against Michigan		%
19	Michigan's points against Ohio State		%
20	Southern California's points against Michigan		%

M300 Students

In this section, you will estimate the percent of students in M300 this semester who are in a particular category. You should try to be accurate within ± 15 [5] percentage points of the truth. Within the 20 categories in this packet, the average percent of students is 25% and 90% of the percents fall between 8% and 54%.

	What **percent** of M300 students are in each category?		How confident are you that this estimate is within **± 15 [5] percentage points** of being correct?
1	one of their favorite tv shows is reality tv?		%
2	one of their favorite tv shows is The OC?		%
3	come from outside the US?		%
4	major in Communications?		%
5	major in Industrial & Operations Engineering?		%
6	are in LS&A?		%
7	are in Undergraduate Engineering?		%
8	are seniors?		%
9	have missed no M300 classes this semester?		%
10	have missed more than half of the classes this semester?		%
11	one of their favorite tv shows is The Family Guy?		%
12	one of their favorite tv shows is Sex and the City?		%
13	one of their favorite tv shows pertained to sports?		%
14	one of their favorite tv shows is Friends?		%
15	come from someplace in Michigan?		%
16	major in Economics?		%
17	major in Psychology?		%
18	major in General Studies?		%
19	missed the first day of M300?		%
20	missed more than three days of M300 this semester?		%

Ann Arbor Delivery Pizza

In this section, you will estimate the price (not including delivery or tax) of delivery pizza from various restaurants in Ann Arbor. You should try to be accurate within ±$3.00 [$1.00] of the truth. Within the 20 delivery pizzas in this packet, the average price of the delivery pizza is $11.53 and 90% of the prices fall between $6.99 and $14.80

	What is the **price** of this Ann Arbor delivery pizza?		How confident are you that this estimate is within ±**$3.00** [$1.00] of being correct?
1	Cottage Inn small cheese		
2	Mr. Pizza small mushroom		
3	Pizza House large cheese deep dish		
4	Bell's Pizza medium sausage		
5	Pizza House small Hawaiian		
6	Little Caesars small vegetarian		
7	New York Pizza Depot small 4 meat combo		
8	Little Caesars large vegetarian		
9	Dominos medium peperoni		
10	New York Pizza Depot medium 4 meat combo		
11	Mr. Pizza medium sausage		
12	Dominos large pepperoni		
13	New York Pizza Depot large 4 meat combo		
14	Pizza Hut medium pepperoni and olive		
15	Bell's Pizza small sausage		
16	Cottage Inn large cheese		
17	Pizza House medium pepperoni deep dish		
18	Little Caesars medium vegetarian		
19	Pizza Hut medium pepperoni		
20	Pizza House large Hawaiian		

Appendix B

This appendix presents the statistical conditions that underlie the relationship between perceived percentile, overplacement, and overconfidence. We begin with the relationship between perceived percentile and overconfidence. Let p´ and p represent perceived percentile and actual percentile, and x´ and x represent mean confidence and proportion correct. Presently, we are interested in the relationship between p´ and OC = x´ - x.

The direction of this relationship is determined by the sign of the covariance. In the following derivation, *S* refers to the standard deviation of the subscripted variable and *r* to the correlation of the subscripted variables.

$$\begin{aligned}\operatorname{cov}\left(p', x'-x\right) &= \operatorname{cov}\left(p', x'\right) - \operatorname{cov}\left(p', x\right)\\ &= S_{p'}S_{x'}r_{p'x'} - S_{p'}S_{x}r_{p'x}\\ &= S_{x'}r_{p'x'} - S_{x}r_{p'x}\end{aligned}$$

The covariance is positive if

$$\frac{r_{p'x'}}{r_{p'x}} > \frac{S_x}{S_{x'}} \quad ,$$

and negative if

$$\frac{r_{p'x'}}{r_{p'x}} < \frac{S_x}{S_{x'}} .$$

In other words, the direction of the relationship can be determined by comparing a ratio of correlations to a ratio of standard deviations.

Thinking through the empirical results in this paper provides some insight into this result. When observations correspond to participants responding to a single topic and difficulty level, $r_{p'x}$ is very low (poor correlation between perceived percentile and accuracy), the ratio of correlations is large and greatly exceeds the ratio of standard deviations, and hence perceived percentile and overconfidence move together. When observations correspond to group means for difficult and easy versions of the same domain, the ratio of correlations is close to one (because perceived percentile, confidence, and proportion correct consistently move together within topic, so $r_{p'x}$ is high) and the ratio of standard deviations is greater than one ($S_x > S_{x'}$, reflecting the fact that confidence tracks accuracy but does not keep up—see means in Tables 3 and 4). The net result is that the ratio of standard deviations exceeds the ratio of correlations, and so perceived percentile and OC move in opposite directions across levels of difficulty. Finally, when observations correspond to group means on topics for a constant level of the difficulty manipulation, $r_{p'x}$ is low (topics where people place themselves highly are not necessarily the ones on which they are most accurate), the ratio of correlations is high, and perceived percentile and OC move together.

When group means are considered, overplacement is simply perceived percentile minus 50, so we can interpret the above relationships in terms of overplacement and OC moving together or in opposite directions. At the individual level, however, high perceived percentile is not necessarily a bias, because the person may truly be performing better than others. In this case, it is interesting to explore the relationship between overplacement (p´ - p) and overconfidence calculated at the individual level. Again, we start with the covariance.

$$\begin{aligned}\operatorname{cov}\left(p'-p, x'-x\right) &= \operatorname{cov}\left(p', x'\right)-\operatorname{cov}\left(p', x\right)-\operatorname{cov}\left(p, x'\right)+\operatorname{cov}\left(p, x\right)\\ &= S_{p'}S_{x'}r_{p'x'} - S_{p'}S_{x}r_{p'x} - S_{p}S_{x'}r_{px'} + S_{p}S_{x}r_{px}\end{aligned}$$

To simplify this relationship and relate it to the results above, we will make two assumptions. First, we assume that $r_{px} = 1$, which is that the objective measures are perfectly correlated. Strictly speaking, this correlation will be less than one because although there is a monotonic relationship between actual percentile and accuracy it is nonlinear. Second, we assume that $r_{p'x} = r_{px'}$. Empirically these correlations between subjective and objective measures tend to be reasonably similar. After applying these assumptions and performing some simple algebra, we find that the covariance is positive if

$$\frac{r_{p'x'}}{r_{p'x}} > \frac{s_x}{s_{x'}} + \frac{s_p}{s_{p'}}\left(1 - \frac{s_x}{s_{x'}r_{p'x}}\right) .$$

Note that this equation is the same as the result above, with an additional term added to the righthand side. The ratio $S_p / S_{p'}$ is likely to be greater than one, since the numerator is the standard deviation of the standard uniform distribution, and it is unlikely that the subjective measures are more spread out than that. When $r_{p'x}$ is low (i.e., there is a low correlation between perceived percentile and proportion correct) the lefthand side will be large and the righthand side will be small or negative. Consequently, overplacement and overconfidence will tend to move together. It is technically possible to reverse this relationship, but it requires a delicate balance of the relative sizes of the ratios of standard deviations and correlations, which we leave for further study.

Table 1

Regression Equations Predicting Degree of Overconfidence from Perceived Percentile, Difficulty Dummy, and Controls (Study 1).

	Equation			
	1	2	3	4
Constant	-20.4	-33.6	-49.1	-55.49
Perceived percentile	.35	.41	.40	.36
Difficulty dummy		21.7	21.5	21.2
Domain dummies			Incl.	Incl.
Participant dummies				Incl.
Adj. R-sq	.09	.23	.31	.44

Note. All coefficients significant at $p < .001$.

Table 2

Regression Equations Predicting Degree of Overconfidence from Perceived Percentile, Difficulty Dummy, and Controls (Study 2).

	Equation			
	1	2	3	4
Constant	-2.9	-16.9	-13.9	-14.8
Perceived percentile	.39	.45	.39	.46
Difficulty dummy		22.0	21.7	21.9
Domain dummies			Incl.	Incl.
Participant dummies				Incl.
Adj. R-sq	.08	.21	.27	.53

Note. All coefficients significant at $p < .001$.

Table 3

Mean Perceived Percentile, Under-/Overconfidence, Confidence, and Proportion Correct by Domain and by Difficulty (Study 1)

Domain	Difficulty	Perceived Percentile	Under-/Over-Confidence	Confidence	Proportion Correct
University	Easy	51.2	-6.0	65.8	71.7
	Hard	41.1	13.1	36.1	23.0
Nobel	Easy	32.1	-24.9	48.0	72.0
	Hard	21.5	.7	14.2	13.5
Wealth	Easy	33.45	-6.8	38.9	46.4
	Hard	31.0	7.7	14.4	6.8
Pop	Easy	40.8	-13.5	48.7	62.2
	Hard	35.7	-1.3	20.7	22.0
NHL	Easy	41.5	-33.9	62.7	96.5
	Hard	31.4	-13.8	25.7	39.5
Overall Mean	Easy	39.8	-17.0	52.8	69.9
	Hard	32.1	1.3	22.2	21.0

Table 4

Mean Perceived Percentile, Under-/Overconfidence, Confidence, and Proportion Correct by Domain and by Difficulty (Study 2)

Domain	Difficulty	Perceived Percentile	Under-/Over-Confidence	Confidence	Proportion Correct
Demographics	Easy	57.4	10.0	57.4	55.7
	Hard	52.5	29.7	58.2	28.6
Campus Distances	Easy	58.0	5.6	60.9	55.3
	Hard	51.8	22.5	55.6	33.1
Michigan Football	Easy	48.6	-4.4	57.6	62.1
	Hard	43.1	15.8	44.9	29.1
Class Info.	Easy	59.0	23.2	67.5	44.3
	Hard	58.1	43.5	57.2	13.7
Pizza Prices	Easy	58.1	7.8	76.3	68.6
	Hard	50.6	28.7	55.9	27.1
Overall Mean	Easy	56.3	8.5	65.7	57.2
	Hard	51.3	28.1	54.4	26.3

Table 5

Perceived Percentile and Under-/Overconfidence Correlated with Personality Measures, by Study.

	Study 1 (n = 40)		Study 2 (n = 35)	
Personality Measure	Perceived Percentile	Under-/ Overconf.	Perceived Percentile	Under-/ Overconf.
Need for Cognition	-.27	-.14	.01	.03
Need for Closure	.31^{+}	.34*	.02	-.13
Self Esteem	-.23	.01	-.01	-.11
Need for Uniqueness	.05	-.07		
Hypernarcissism			.02	.15
Defensive Pessimism			-.09	.13
Narcissistic Personality Inventory			.01	.09

Note. Perceived percentile and under-/overconfidence are calculated at the individual level by averaging response across 10 subdomains.

* = $p < .05$, $^{+}$ = $p < .1$, two-tailed.

Table 6

Table 3

Correlations between Perceived Percentile, Confidence, Actual Percentile, Proportion Correct, Under-/Overplacement, and Under-/Overconfidence, by Study.

Variable	Confidence	Actual Percentile	Proportion Correct	Under-/Over-Placement	Under-/Over-Confidence
Study 1(n = 40)					
Perceived Percentile	.649**	.490**	.490**	.870**	.565**
Confidence		.282	.310	.587**	.945**
Actual Percentile			.917**	-.003	.167
Proportion Correct				.039	.159
Under-/Overplacement					.556**
Study 2(n = 35)					
Perceived Percentile	.438*	.332*	.348*	.687**	.253
Confidence		.085	.142	.350*	.885**
Actual Percentile			.971**	-.457**	-.375*
Proportion Correct				-.419*	-.335*
Under-/Overplacement					.530**

Note. Easy Under-/Overconfidence and Hard Under-/Overplacement correlated .339 in Study 1 and .495 in Study 2. Hard Under-/Overconfidence and Easy Under-/Overplacement correlated .475 in Study 1 and .194 in Study 2.

Table 7

Average Correlations between Perceived Percentile and Under-/Overconfidence Calculated for Different Combinations of Domain and Difficulty, by Study.

Average Correlations	Study 1	Study 2
Within subdomain (*n* = 10)	.41	.31
Within domain and across difficulty (*n* = 10)	.41	.22
Across domain and within difficulty (*n* = 40)	.25	.06
Across domain and difficulty (*n* = 40)	.22	.08

Note. *n* is the number of correlations that are used to calculate the averages. Correlations were *r*-to-*z* transformed before averaging and then converted back. Significance tests were performed on the transformed values.

Table 8

Average Correlations between Under-/Overplacement and Under-/Overconfidence Controlling for Self-Schema Calculated for Different Combinations of Domain and Difficulty, by Study.

Average Correlations	Study 1	Study 2
Within subdomain (*n* = 10)	.57	.70
Within domain and across difficulty (*n* = 10)	.22	.26
Across domain and within difficulty (*n* = 40)	.13	.07
Across domain and difficulty (*n* = 40)	.14	.09

Note. *n* is the number of correlations that are used to calculate the averages. Correlations were *r*-to-*z* transformed before averaging and then converted back. Significance tests were performed on the transformed values.

Figure Captions

Figure 1. The weak relationship between confidence and proportion correct in the overconfidence litereature

Figure 2. The weak relationship between perceived percentile and actual percentile in the better-than-average literature.

Figure 3. Three hypothesized relationships with perceived percentile: A strong relationship between perceived percentile and confidence (top panel); a weak relationship between perceived percentile and proportion correct (top panel); and a strong relationsihp between perceived percentile and overconfidence (bottom panel).

Figure 4. Hypothesized relationship between perceived percentile and proportion correct (top panel), confidence (top panel), and overconfidence (bottom panel) as task difficulty varies.

Figure 5. Hypothesized reversal between perceived percentile and overconfidence as task difficulty varies within a domain: Easy tasks yield higher perceived percentiles and less overconfidence.

Figure 6. Regressions of confidence, proportion correct, and overconfidence on perceived percentile (Study 1).

Figure 7. Regressions of confidence, proportion correct, and overconfidence on perceived percentile (Study 2).

Figure 8. A plot of the ten domain level averages for overconfidence and perceived percentile ignoring the difficulty manipulation (top panel, with regression line) and including the difficulty manipulation (bottom panel—measures from the same domain are connected by a line).

Figure 1

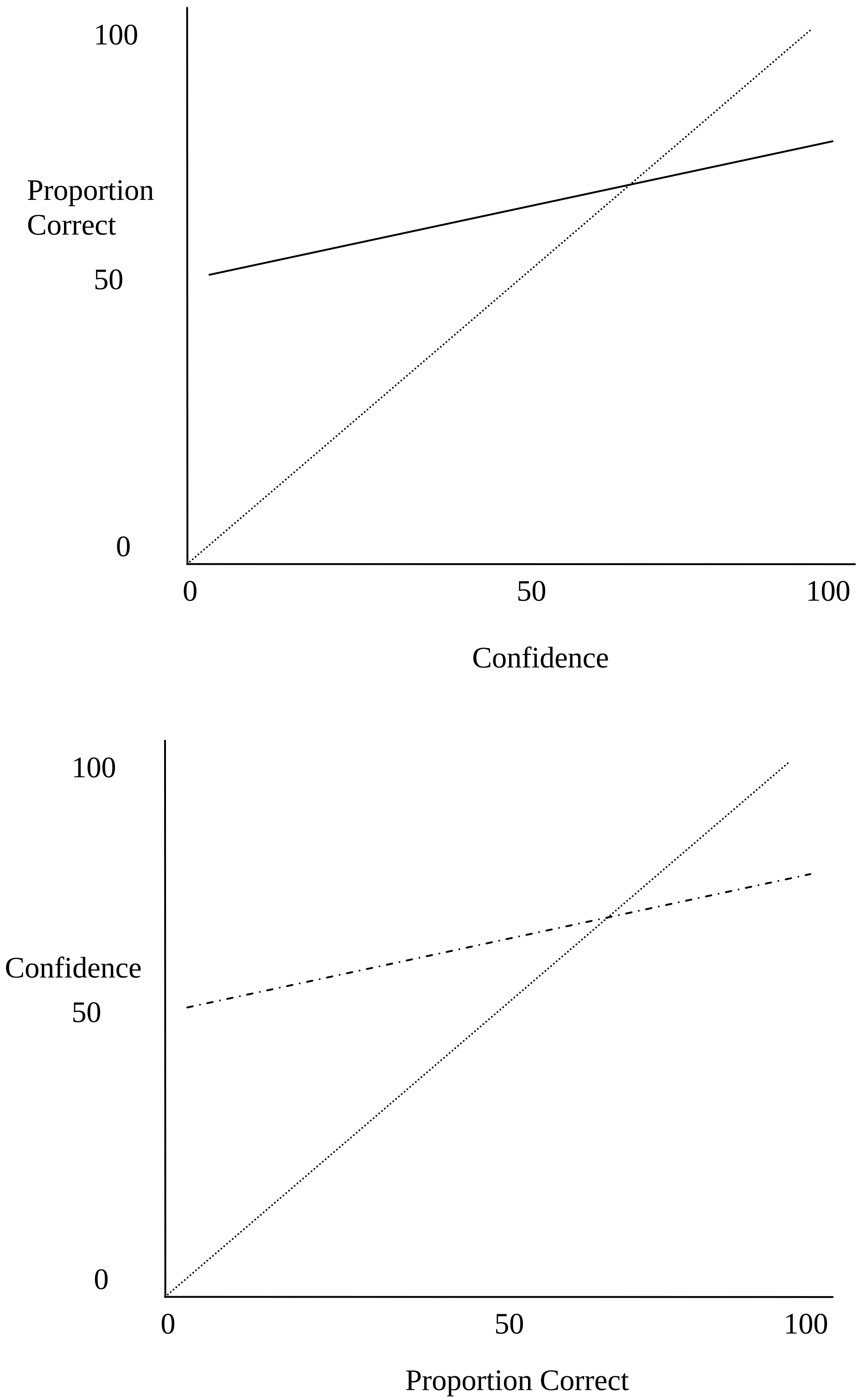

100
Proportion
Correct
50
0
0
50
100
Confidence
100
Confidence
50
0
0
50
100
Proportion Correct

Figure 2

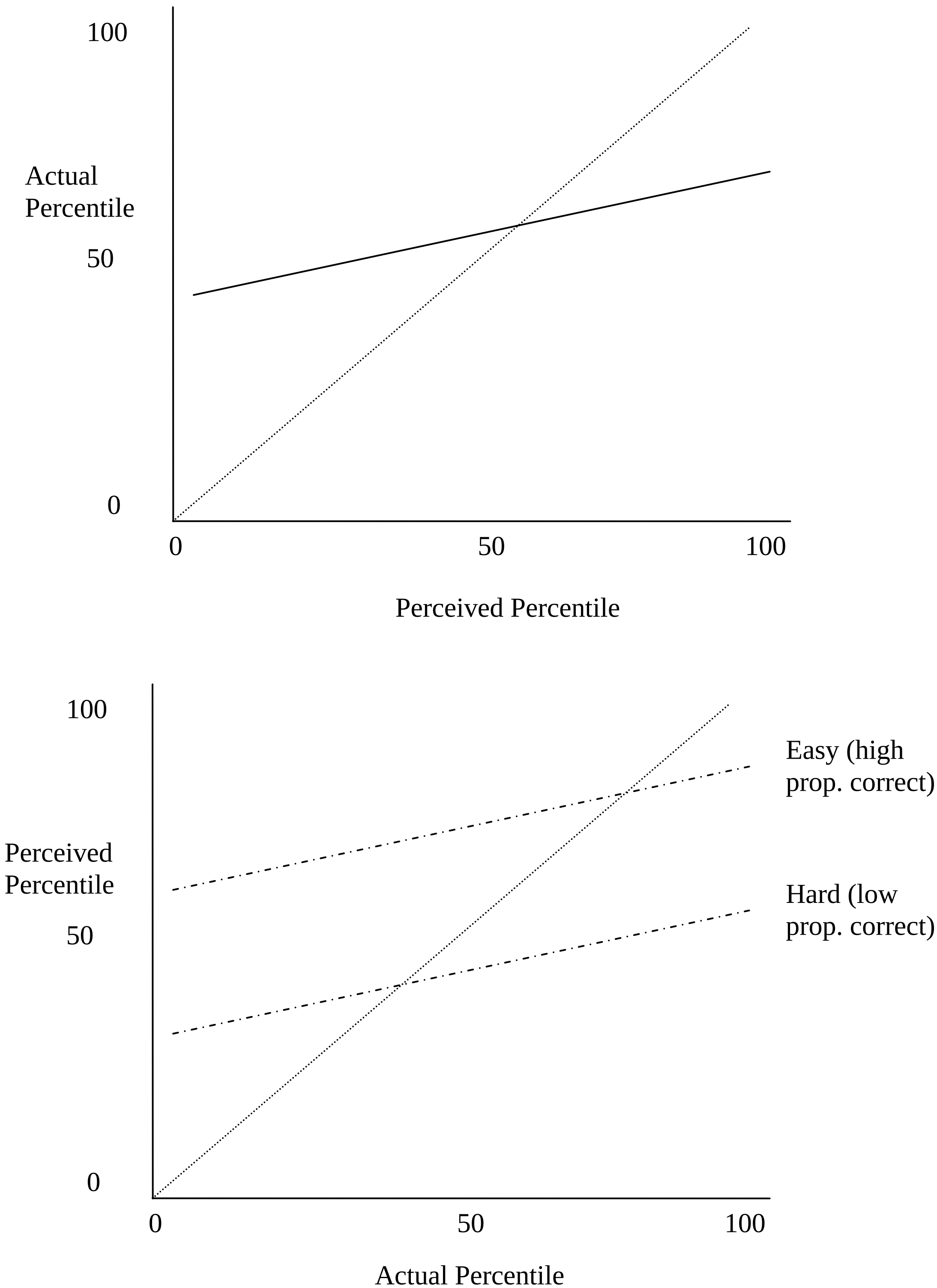
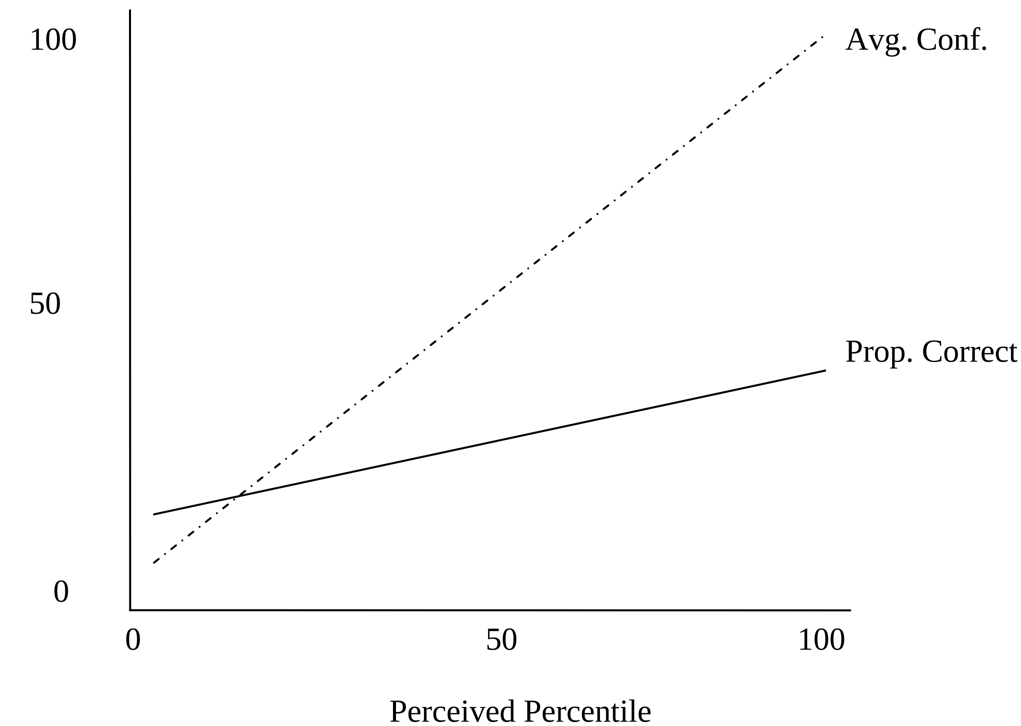

100
Actual
Percentile
50
0
0
50
100
Perceived Percentile
100
Easy (high
prop. correct)
Perceived
Percentile
Hard (low
prop. correct)
50
0
0
50
100
Actual Percentile

Figure 3


100
50
0
Avg. Conf.
Prop. Correct
0
50
100
Perceived Percentile

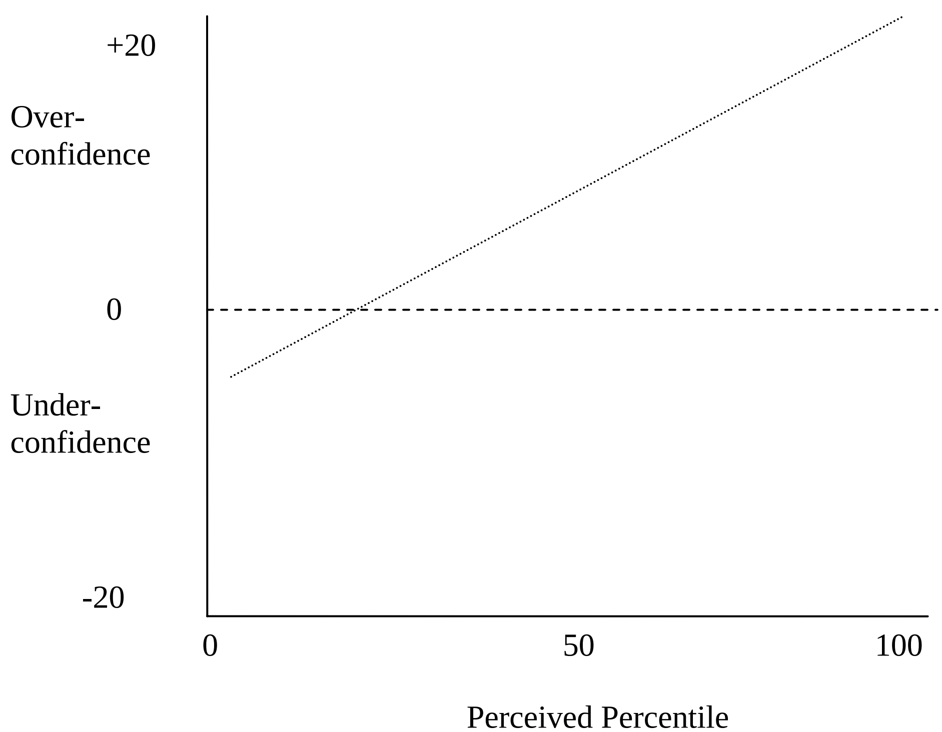

+20
Over-
confidence
0
Under-
confidence
-20
0
50
100
Perceived Percentile

Figure 4

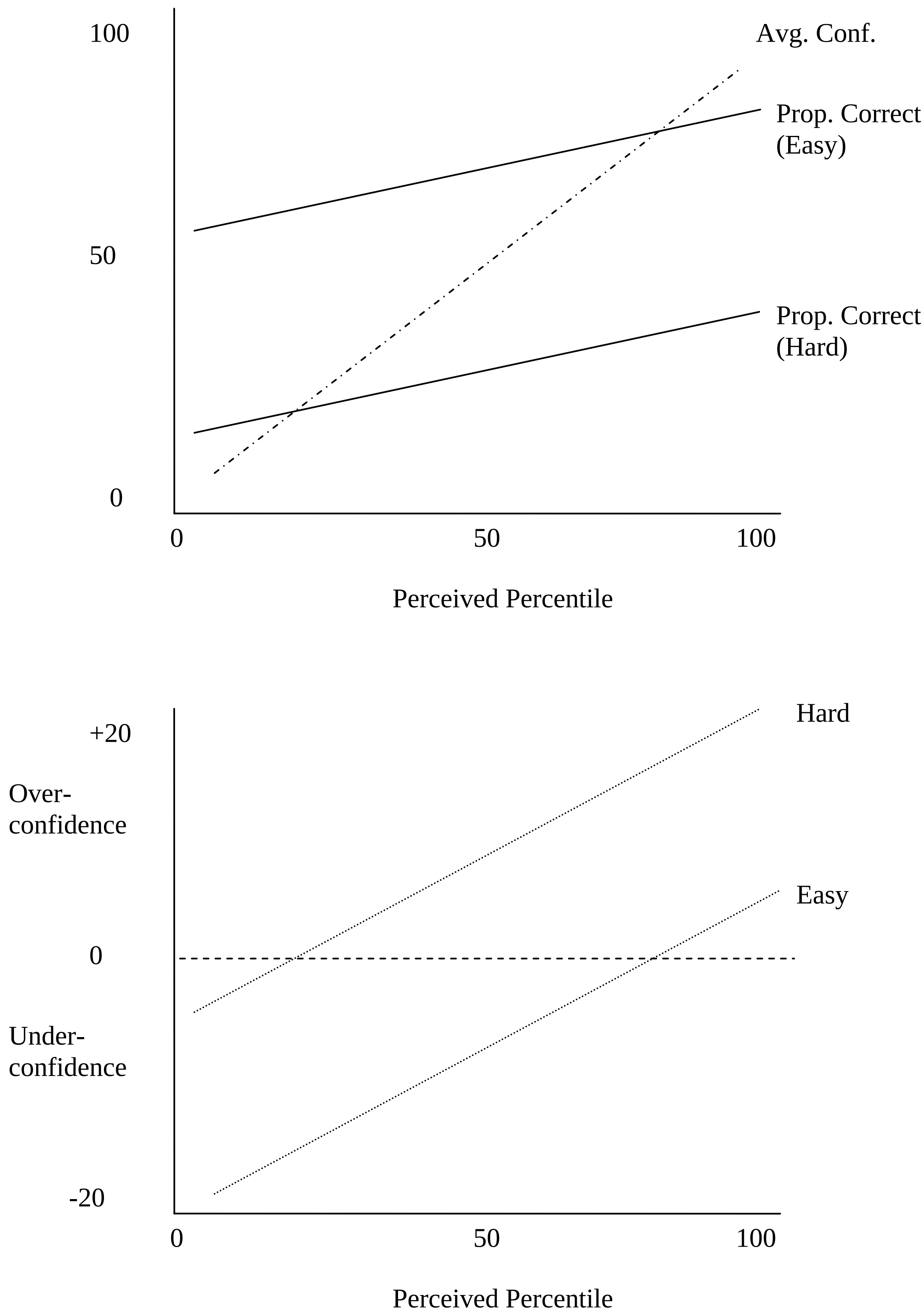

100
50
0
Avg. Conf.
Prop. Correct
(Easy)
Prop. Correct
(Hard)
0
50
100
Perceived Percentile
+20
Over-
confidence
0
Under-
confidence
-20
Hard
Easy
0
50
100
Perceived Percentile

Figure 4

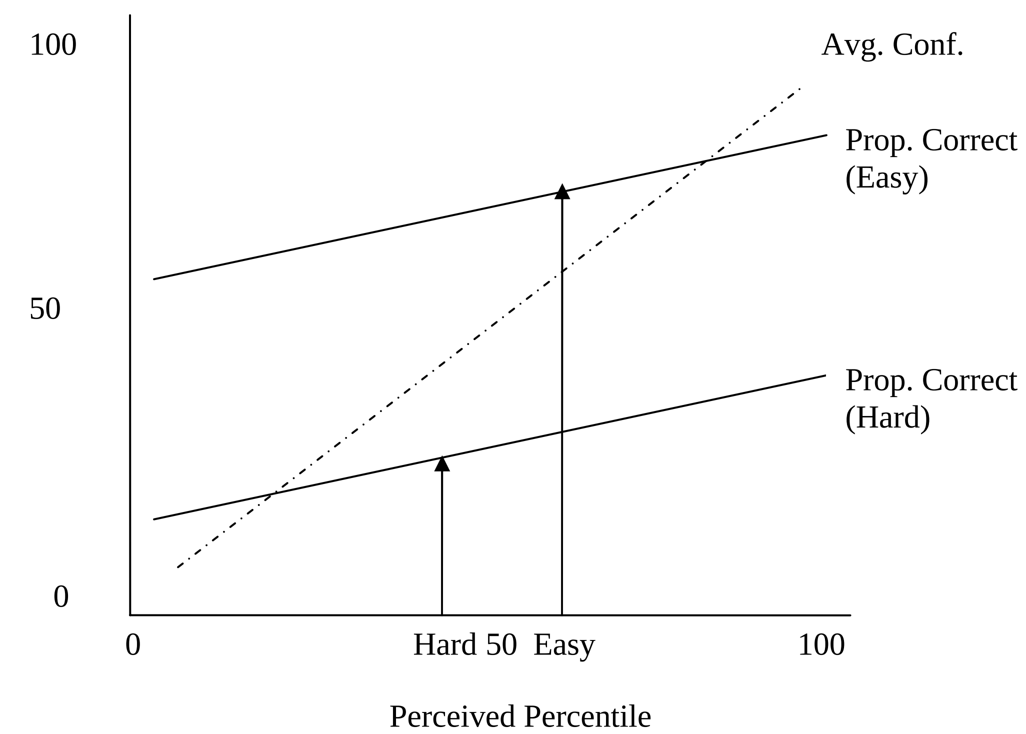


Figure 5

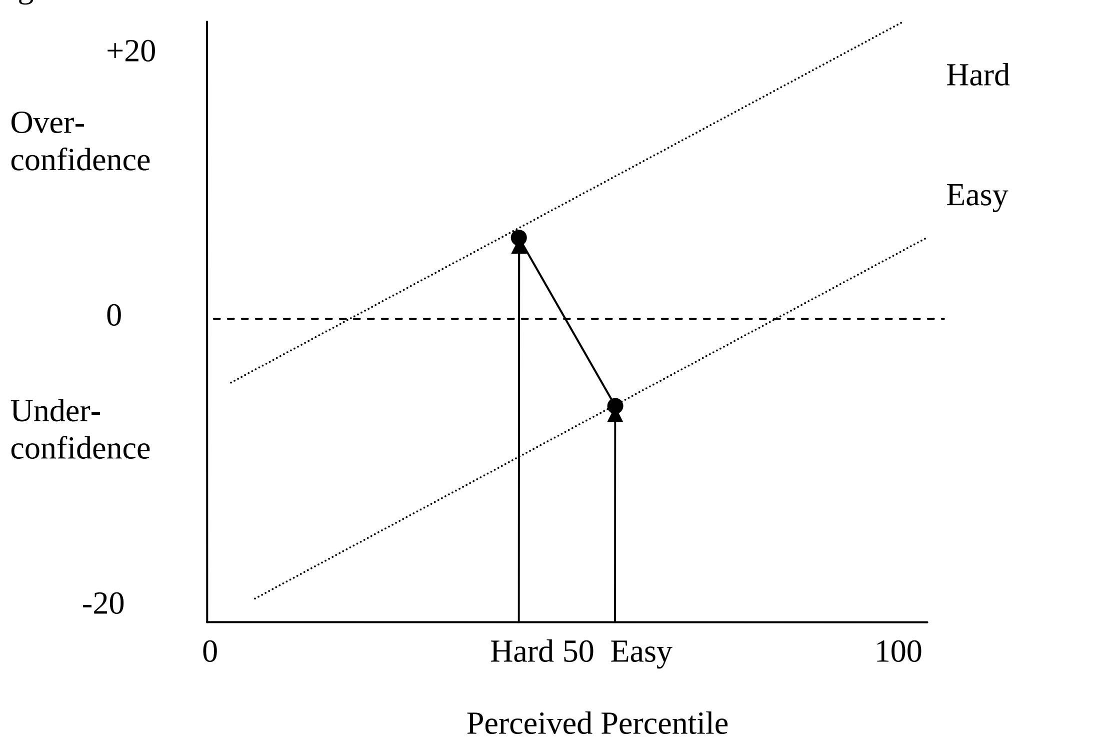
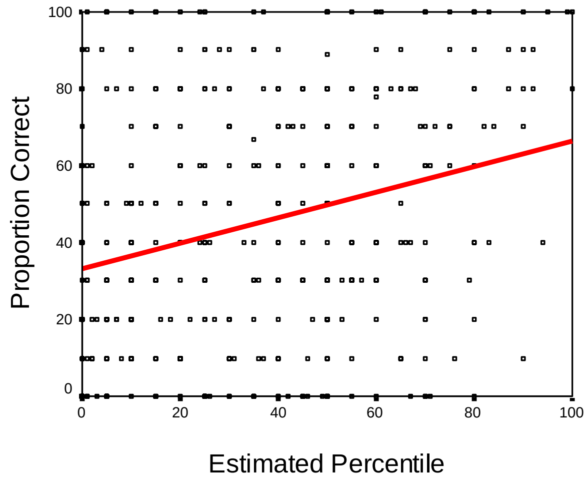
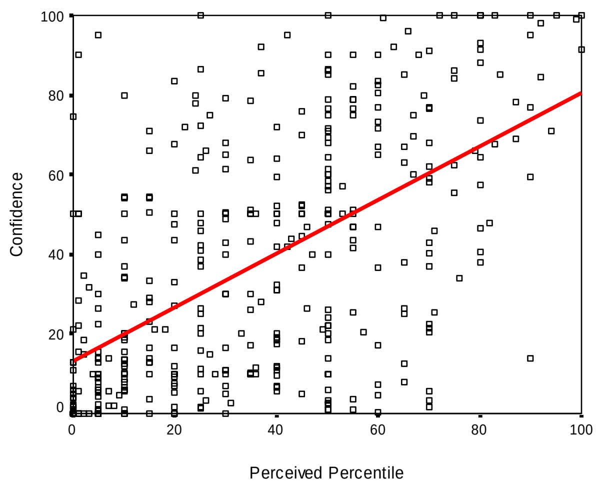


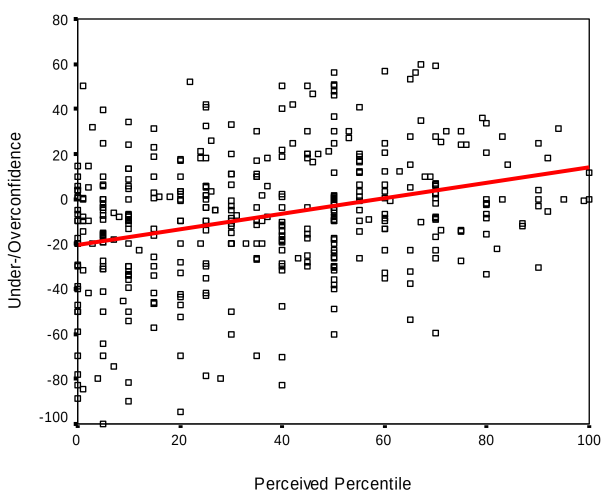
Figure 6



Correct = 33.36 + .33 * Percentile

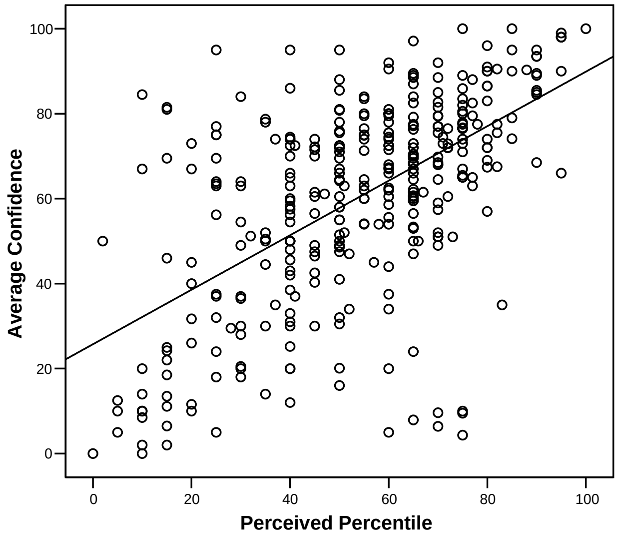


Conf = 13.33 + .67 * Percentile

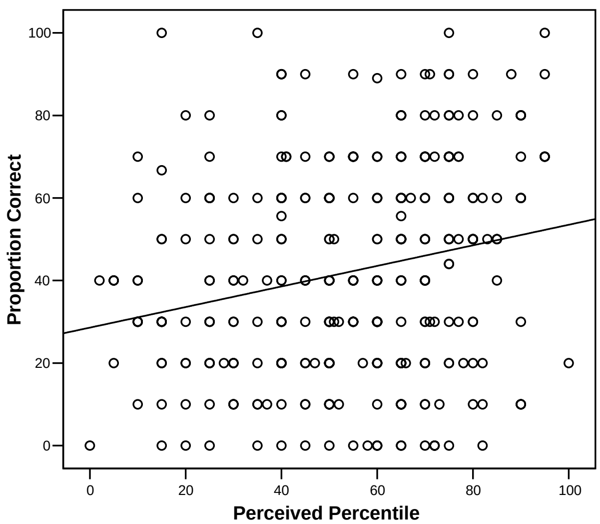


OC = -20.35 + .35 * Percentile

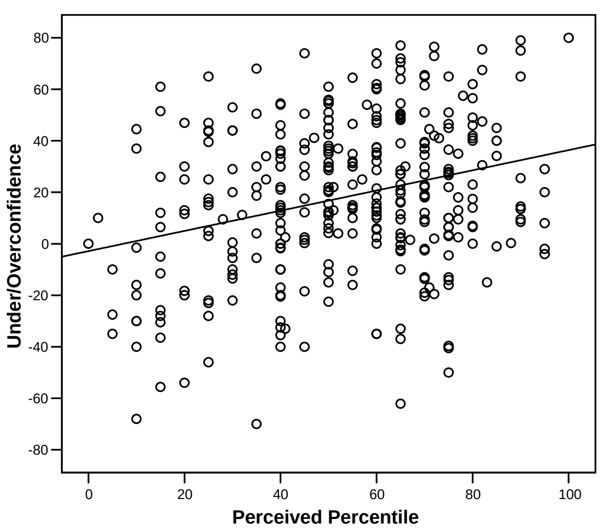
Figure 7



Confidence = 25.7 + .64 * Percentile

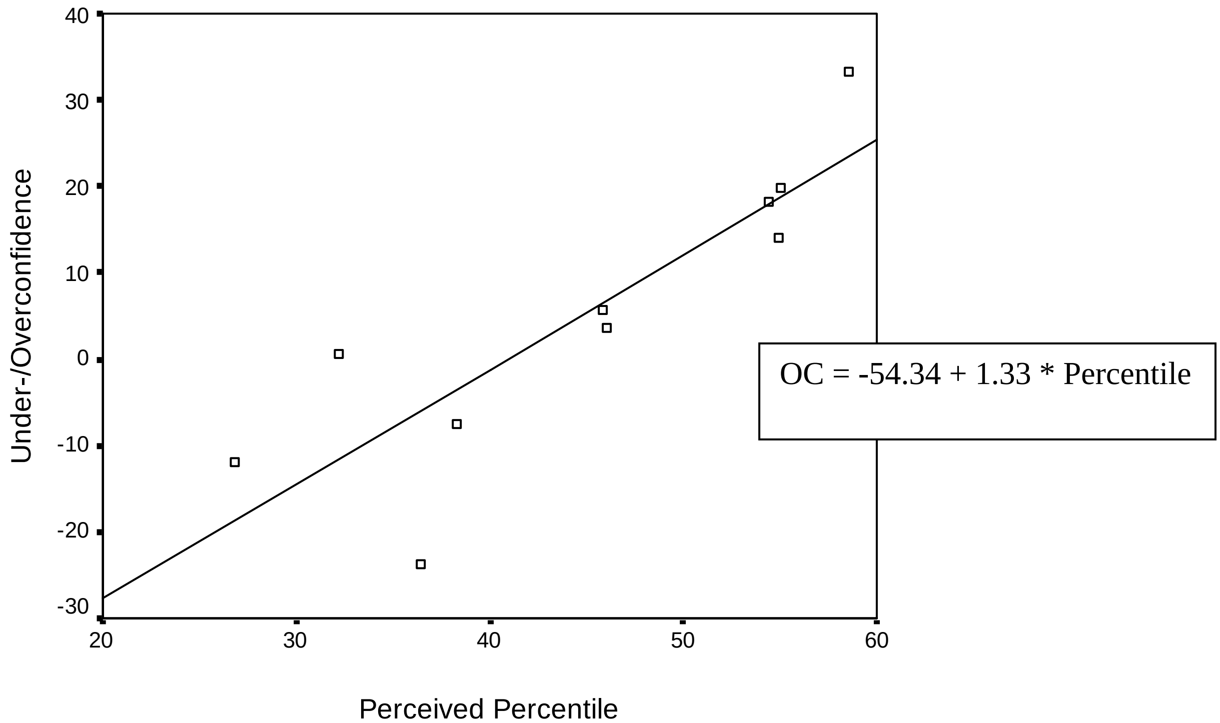


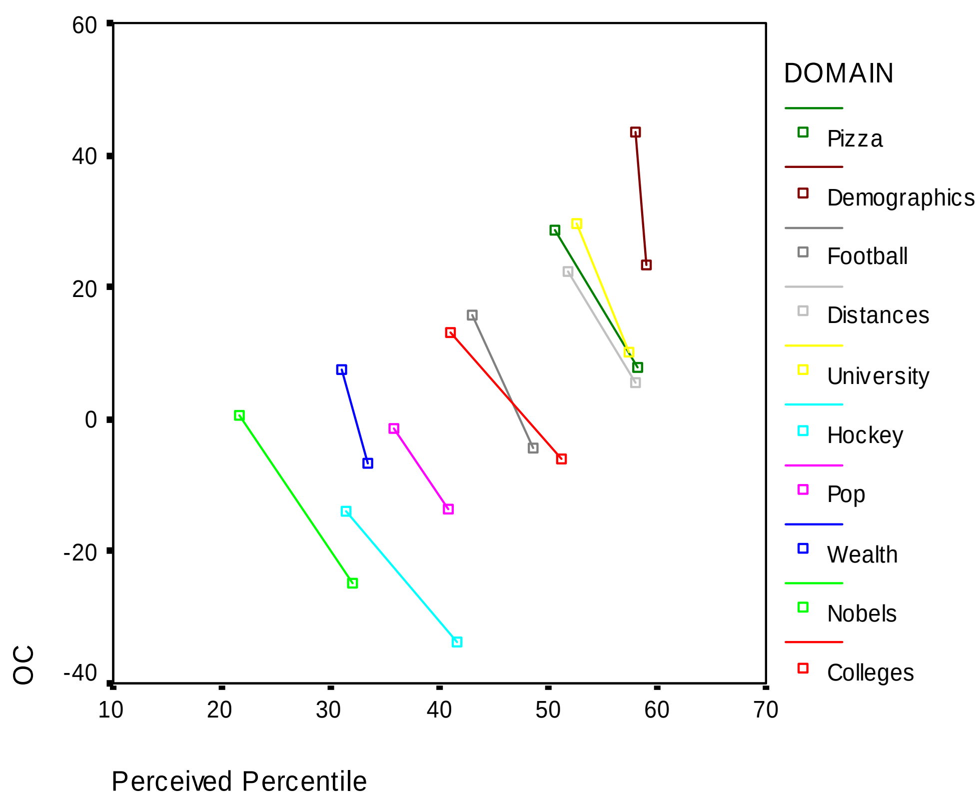
Correct = 28.6 + .25 * Percentile



OC = -2.85 + .39 * Percentile

Figure 8


40
30
20
10
0
-10
-20
-30
Under-/Overconfidence
20
30
40
50
60
Perceived Percentile
OC = -54.34 + 1.33 * Percentile


60
40
20
0
-20
-40
OC
10
20
30
40
50
60
70
Perceived Percentile
DOMAIN
Pizza
Demographics
Football
Distances
University
Hockey
Pop
Wealth
Nobels
Colleges

Footnotes

[1] The hard-easy effect in overconfidence is inevitable for some methods of sorting questions into hard and easy categories (Gigerenzer, Hoffrage, & Kleinbölting, 1991; Juslin et al., 2000; Klayman et al., 1999). Part of the problem is that the independent variable (accuracy) and dependent variable (overconfidence = confidence – accuracy) are bound to be correlated because the same measure of accuracy shows up in both halves of the equation. Error in the accuracy measure guarantees the effect (see Gigerenzer et al., 1991; Juslin et al., 2000; Klayman et al., 1999). Some versions of the hard-easy effect hold up to statistical control. For example, individuals who are less accurate on one set of questions (i.e., the questions are *hard* for them) are more overconfident on a different set of questions on the same topic (Klayman et al., 1999).

[2] It is worth noting that the "hard-easy" effect depicted in the bottom panel of Figure 2 differs from that traditionally discussed in the overconfidence literature because average proportion correct is manipulated for a whole population in an ability domain. Thus, the hard-easy effect on percentiles is manifested as an increase in perceived percentile (i.e., the line in the bottom panel of Figure 2 shifts upward). Kruger and Dunning (1999) conducted an analysis of percentile estimates that more directly parallels the original hard-easy analysis in the overconfidence literature (see Footnote 1) when they sorted participants based on their actual percentile and examined their percentile estimates. They found that the worst performers (as measured by actual percentile) overestimated their percentile, whereas the best performers underestimated theirs. In a direct parallel to the overconfidence literature, this pattern has been reinterpreted as a necessary effect of regression (Ackerman et al., 2002; Burson et al., in press; Krueger & Mueller, 2002).

[3] Mean differences in the correlations were tested after performing an *r*-to-*z* transformation on the individual correlations, taking an average, and then performing a set of planned, non-orthogonal contrasts (df = 96) on the means of the transformed correlations. The contrasts compared Row 1 vs. Row 2, and Row 2 vs. Rows 3 and 4.