

The Market for Overconfidence

Joseph R. Radzevick

A dissertation submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy to the Tepper School of Business at
Carnegie Mellon University

Committee:

Don Moore, Haas School of Business, University of California-Berkeley

Francesca Gino, Harvard Business School, Harvard University

Roberto Weber, Department of Social and Decision Sciences, Carnegie Mellon University

Table of Contents

Acknowledgments	3
Abstract	4
Introduction	6
Part I: Competing To Be Certain (But Wrong): Social Pressure and Overprecision in Judgment	15
Part II: Changes to the Market Landscape	29
Part III: Strategic Interventions in the Market	50
Conclusion.....	63
References	67
Tables	75
Figures	84
Appendices	91

Acknowledgments

I wish to thank the members of my committee Don Moore, Francesca Gino, and Roberto Weber for their advice and suggestions. Additional thanks go out to my advisor and co-author Don Moore for his guidance and financial support for these projects as well as to Mark Fichman for his counsels on some of the nascent ideas that shaped this work. I am grateful for the resources and facilities provided by the Center for Behavioral Decision Research and the Pittsburgh Supercomputer Center. I also acknowledge the financial support of the William Larimer Fellowship and a research grant from the Center for Behavioral Decision Research. I would like to thank Carolina Velez, Jennifer Hwang, and Mingwei Hsu for their help with data collection and Lawrence Rapp for his administrative aid. Thanks to seminar attendees at the Tepper School of Business as well as conference attendees at the Academy of Management Annual Meeting and the Society for Judgment and Decision Making Annual Conference for their thoughtful feedback. I also am grateful for the comments and suggestions of editors and anonymous reviewers at *Management Science*. I appreciate the patience of my family and all the help they provided as I walked this long path. And a special thanks to Griffin Burns for his inspiration and the life lessons he teaches every day. Above all, I appreciate the consummate support of my wife Stephanie through this process, for without her I never could have reached this point (We've come a long way, baby!).

Abstract

Researchers distinguish overprecision as both the most robust and the least understood form of overconfidence. Such overly precise judgments claim more certainty than is objectively warranted and often result in expressions of confidence that far exceed the accuracy of those judgments. In this dissertation, I investigate the social underpinnings of overprecision by examining two key relationships: the relationship between producers (i.e., individuals forming judgments) and receivers (i.e., individuals utilizing those judgments for some purpose) and the relationship between rival producers. Prior research has examined certain elements of these exchanges but has not adequately addressed the exchange system as a whole. To fill this gap, I create an experimental framework in which judgment advisors (producers) interact with guessers (receivers) in a market environment, where producers compete with one another for guessers' attention. I argue that confidence is intensified when competition for influence exists between producers. Moreover overprecision, customarily viewed as a detriment to effective decision making, may actually aid producers since its benefits (such as higher rates of selection) often exceed its costs (such as tarnished reputations).

Two initial studies contrast between the competitive market and a more static producer-receiver exchange. In Study 1, guessers completed an estimation task and chose advice from one of four advisors. Advisors display a high degree of overprecision in their estimates: their expressed confidence levels far exceed their hit rates. The expressed confidence of advisors increased over time with no corresponding gains in accuracy. Guessers did punish advisors for being confident and wrong by selecting them less often in subsequent rounds. However, this cost is outweighed by the benefit of continuing to express high confidence, namely that it attracts more customers. In Study 2, guessers completed the same estimation task with the aid of

advisors, but in this case guessers and advisors were paired for the duration of the session. Advisors again display overprecision in their estimates. Despite similar costs and benefits associated with the advisors' actions, there are no significant changes to their confidence levels and distributions across rounds as had occurred in the first study.

In Studies 3 and 4, I examine how changes in certain features of the market affect these patterns of confidence. In Study 3, I varied the initial information provided to advisors. “High” information advisors received a stronger signal as to the correct answer. These high information advisors possess some inherent advantages over less informed advisors (e.g., greater accuracy). However, they remain overconfident to a similar magnitude, potentially due to the competitive pressure of the other advisors. In Study 4, I demonstrate the importance of behavior transparency for the documented patterns of advisor results. When advisors have information about the decisions made by their competitors, they are more overconfident and escalate their confidence at a higher rate than when this information is undisclosed.

I tested whether certain interventions to the market can help reduce overconfidence and improve performance in Study 5. A *Control* condition mirrored the basic framework of Study 1. In a *Payoff* condition, advisor earnings were based upon both their selection by guessers and their accuracy. For a *Feedback* condition, guessers received enhanced information about the performance of advisors at the time in which they made their selection decisions. The proposed interventions did little to change advisor behavior, as the degree of overprecision and the levels of expressed confidence were similar across all three conditions. However, the interventions did help receivers achieve higher levels of performance than in the *Control* condition.

Combined, these studies help further our understanding of this pervasive yet enigmatic form of overconfidence.

The Market for Overconfidence

Introduction

Overconfidence has been called “perhaps the most robust finding in the psychology of judgment” (DeBondt and Thaler, 1995). In a recent paper, Moore and Healy (2008) distinguish three varieties of overconfidence: (1) overestimation of your own performance, (2) overplacement of your performance relative to those of others, and (3) overprecision, or excessive certainty that you have the right answer. The authors highlight overprecision as both the most robust and the least studied form of overconfidence. Indeed, evidence suggests this kind of overconfidence manifests across a wide variety of decision making contexts (see Koehler, Brenner, and Griffin, 2002).

Financial advisors experience difficulties in accurately predicting short term stock performance, rarely exceeding the performance of more “naïve forecasting” (Malkiel, 1996). Similarly, De Bondt and Thaler (2002) describe the tendencies of securities analysts to overreact to market movement with extreme forecasts and unrealistic optimism. Though fund managers trumpet their skills and the quality of their financial instruments, actively managed mutual funds do not consistently outperform simple indexed funds (Bogle, 1999). Money manager performance consistently fails to meet performance benchmarks and expectations (Wood, 1989). Conversely, their management fees imply that at least someone thinks they are worth paying for. Financial firms rationalize providing high compensation packages for such professionals by claiming they are necessary to “attract and retain the best and brightest talent” (Andrews and Baker, 2009). But the data suggest these claims are dubious.

Despite bold claims to the contrary, political forecasters rarely provide an accurate glimpse of how world events unfold (Tetlock, 2005). Studies often fail to find strong evidence that changes in executive leadership affects key organizational outcomes (Pfeffer, 1977), but firm leadership remains a popular means of attributing performance (Meindl, Ehrlich, and Dukerich, 1985). Along the same lines, management consultants often fail to produce the sweeping improvements they promise their clients, yet they retain a great deal of influence with managers (Mickelthwait and Wooldridge, 1996).

Matters of overprecision emerge on even the grandest of public stages. For example, the presidency of George W. Bush notably was marked by what some have called steady resolve and some have called stubborn inflexibility. Bush's 2004 re-election campaign claimed that he and Vice President Cheney offered "steady leadership in times of change." By contrast, the Bush campaign painted its Democratic opponent, John Kerry, as an unreliable waffler who had a tendency to change his mind. In one telling moment in their first Presidential debate when Bush accused Kerry of indecisiveness, Bush remarked, "I just know how this world works, and that in the councils of government, there must be certainty from the U.S. president" (South Florida Sun-Sentinel, 2004). While voters did elect Bush to a second presidential term, did his statements actually acknowledge a fundamental truth? For decision makers to be effective and prominent, do they have to express excessive certainty in their judgments or chosen courses of action?

Big questions remain about why overprecision occurs and what contributes to it. In this dissertation, I examine the social underpinnings of overprecision in a novel way. In doing so, I hope to develop a more comprehensive representation of this phenomenon while examining how various social forces influence the way individuals express their confidence, and why they so often express too much of it.

Confidence as a Social Construct

Overprecision has been attributed to cognitive causes, such as availability and anchoring (Russo and Schoemaker, 1992), mirroring the attributed causes of overplacement (see Chambers and Windschitl, 2004; Moore, 2007). Though I acknowledge the presence of such cognitive factors, it is equally important to focus on the social nature of confidence as it pertains to expressions of precision. The social aspects of precision judgments are twofold. First, precision judgments are customarily generated by one actor for evaluation and often practical use by other actors. In general, this creates a distinction between producers and receivers of judgment (Yaniv and Foster, 1997). With more formalized exchange relationships, a similar distinction becomes one of advisors and judges (Sniezek and Buckley, 1995). Second, whether the producers of judgments are managers, consultants, forecasters, or even political candidates, they undoubtedly find themselves in competition with other producers in their attempts to influence potential recipients (observers, customers, voters, etc.). These social components involve various elements, chief among them communication and persuasion.

Yaniv and Foster (1995; 1997) describe the inherent communication tradeoff in precision judgments as one of informativeness versus accuracy. Though their work deals exclusively with interval judgments, the arguments they advance provide a useful framework across the broader spectrum of precision estimates (e.g., both interval and percentage estimates). When communicating such judgments, producers must balance the accuracy of the judgment (i.e., the degree to which it approaches the correct answer) with its informativeness (i.e., the degree to which it provides substantive knowledge). So for example, consider someone producing an estimate of when the American Bicentennial occurred. It would be accurate for the person to estimate 1976, the 1970s, the 20th century, or the last 1000 years, yet each of those estimates

includes a decreasing level of informativeness. Yaniv and Foster's experimental evidence also suggests receivers are willing to accept some amount of error to receive more informative judgments. Questionnaire data collected after a similar interval estimation procedure by Plous (1995) provides additional support for producers' mindfulness of this tradeoff in the ways participants explained their tendencies to report overly narrow intervals:

“Some participants reported a fear that unduly wide ranges would make them look ignorant. Others reported that anyone could give enormously wide intervals, but that the trick was to capture the true value as closely as possible...And several participants indicated that extremely large intervals would be meaningless, uninformative, or unrewarding.” (p. 451)

Confidence also may communicate cues about the advisors themselves, such as their knowledge and level of expertise (Price and Stone, 2004; Sniezek and Van Swol, 2001).

In addition to providing information to recipients, confidence also may serve as a means for producers to sway those recipients. Judges follow the advice of more confident advisors (Sniezek and Buckley, 1995; Sniezek and Van Swol, 2001). When interacting with others, expressions of moderate levels of confidence have more persuasive power than lower levels (London, McSeveney, and Tropper, 1971). Members of minority factions in groups exert more influence when they offer consistent, confident belief in their positions (Moscovici, 1976).

Clearly, expressions of confidence can serve constructive functions in social exchanges. But as shown in many of the examples above, excessive confidence can have damaging repercussions. Why then does the system sustain such unbridled overconfidence? I contend that

an important answer centers on how the competitive markets¹ for influence (e.g., leadership, advice, or credibility) push contestants toward expressing higher confidence. For people to benefit from making exaggerated claims about their certainty, the benefits they gain from such exaggeration (in terms of increased influence or credibility) must outweigh the reputational costs of being wrong. The next sections further explore these benefits and costs.

The Benefits of Expressing Certainty

While proving detrimental to certain performance measures (e.g., accuracy), overprecision likely provides some social value for individual producers. There is indeed some compelling evidence that confident others can be persuasive. Financial advisors who insist they know whether stocks will go up or down in the future are seen as more credible and trustworthy than advisors who express modest confidence, even when both predict which way the stock's price goes with equal accuracy (Price and Stone, 2004). Political experts who claim more certainty and make more extreme predictions are in more demand by the media (Tetlock, 2005). Sniezek and Van Swol (2001) found that advisors who expressed more confidence earned greater trust, were more likely to have their advice followed, and engendered more confidence in those receiving their advice. Charismatic and visionary leaders can benefit the organizations they manage by marshalling action both inside (Westley and Mintzberg, 1989) and outside the firm (Flynn and Staw, 2004). Clearly, their motivating influence is more due to the inspiration they provide than the careful calibration of their confidence judgments (Conger and Kanungo, 1987; House, 1977).

The best current theory to account for the effectiveness of overprecision in judgment has to do with its value in communication. As discussed earlier, Yaniv and Foster (1995; 1997)

¹ Throughout these studies, I use the term “markets” in a colloquial sense to describe the exchange between producers and receivers of judgment in a competitive environment. I acknowledge that the exchanges here do not fulfill all of the formal criteria for markets as defined in economics, such as the presence of a price mechanism.

argue that the reason why people express overprecision is that it increases the informativeness of what they say. For example, if pressed to estimate the gross domestic product of the United States in 2007, a person can maximize his or her chances of being right by saying, “Somewhere between zero and infinity.” It would be considerably more informative to estimate that it is between \$14 and \$15 trillion. The second estimate would be wrong—the actual GDP in 2007 was \$13.8 trillion. But it would be nevertheless be a much more useful estimate. So it might be reasonable to expect that consumers of advice, those who look to leaders for guidance or those in search of a credible expert for judgment aid, would place value in having more precise estimates even if they come at the cost of accuracy.

But more precise advice is really only useful if it is closer to the truth. Estimating the GDP at between \$200 and \$210 trillion, while precise, would be extremely misleading. The key question, then, is whether confidence is positively correlated with accuracy. Often it is (Bornstein and Zickafoose, 1999; Lindsay, Read, and Sharma, 1998; Sniezek and Van Swol, 2001). Naturally, there are some important exceptions, in which confidence and accuracy are uncorrelated, such as in eyewitness testimony (Brewer and Wells, 2006; Wells and Olson, 2003) and detecting others’ deception (DePaulo, Charlton, Cooper, Lindsay, and Muhlenbruck, 1997). But it is rare that they are negatively correlated. And so, when it is difficult to get accuracy data (as with predictions of the future) then the advisor’s own confidence that he or she has made the correct prediction may be a useful clue, and it may well be better than nothing. It may therefore be perfectly sensible for people to prefer confident advisors.

The Risks of Being Wrong

Being wrong is, of course, a significant risk created by claims of certainty. Those who make the most confident predictions will have the most egg on their faces when they turn out to

be wrong. John Kerry's rejoinder to Bush's admonition in the 2004 Presidential debate was: "It's one thing to be certain, but you can be certain and be wrong" (South Florida Sun-Sentinel, 2004). After confidently insisting that Saddam Hussein's Iraq possessed weapons of mass destruction, George W. Bush lost a great deal of credibility when he turned out to be wrong (Ricks, 2006).

And there is indeed some evidence that highlights the risk to claiming confidence and being wrong. Tenney, MacCoun, Spellman, and Hastie (2007) showed that eyewitnesses who claimed complete confidence regarding a key fact that later turned out to be false lost credibility. Tenney, Spellman, and MacCoun (2008) showed further that witnesses establish their own credibility best by showing good calibration and knowing when they are correct. Corporations, when selecting leaders, often appear to be willing to pay a premium for managers whose confidence and bravado make them charismatic (Khurana, 2004). However, there is also clear evidence that overconfident CEOs can get their firms into trouble (Hayward and Hambrick, 1997; Malmendier and Tate, 2005).

We ought to expect the risks of overprecision to increase over time, as the chickens come home to roost and people figure out that leaders' bold assurances can be wrong. On the other hand, Pfeffer (1992), for one, was skeptical that feedback about accuracy would catch up with overconfident managers: "People who misuse information and analysis for their own political ends, the argument goes, will eventually be 'uncovered' when decisions or results turn out badly. This learning will ensure that, over time, better information and better analysis are rewarded and incorporated into the organization's standard operating procedures...However, there is little evidence that these assumptions are true, and there are numerous examples of organizations behaving, for quite predictable reasons, in exactly the opposite way. As a consequence, the

opportunity to use information and analysis as potent political weapons is available, and those with the skills and knowledge of how to do so can often...gain substantial power and influence in their organizations.”

One might well ask why it is that these advisors—who are wildly overconfident, and therefore frequently wrong—do not suffer damage to their reputations. The answer is that they do to some degree, but that these costs do not outweigh the clear benefits from claiming confidence. As Tetlock (2005) observed, the lack of clear and immediate feedback is a strong impediment to achieving such a reckoning. The actual decisions made by advisors and other agents customarily are decoupled from the actual outcomes associated with those judgments. Since the feedback takes too long to arrive, people forget (Pfeffer, 1992). During the late 1990’s, a fervent policy debate raged over the possible repeal of the Glass-Steagall Act, which restricted the combination of commercial and investment banking interests. Proponents of the appeal in the legislative and financial sectors contended that removing these barriers would strengthen the capabilities of U.S. financial institutions while opponents warned that the provisions were necessary to minimize risk to the economy. The repeal occurred in 1999 yet it was not until a global financial crisis occurred nearly a decade later that the full consequences of these actions were realized (Vekshin, 2009). By this time, it was difficult to remember the content of the original debate and many of the players already had left public accountability (such as retired Senator Phil Gramm, sponsor of the repeal bill).

Such an environment appears ripe for the overconfident to gain more than they lose in the marketplace.

The Current Studies

I undertake an experimental examination of these markets for influence because field settings contain important limitations with regard to testing predictions. First, it is rare for aspiring leaders, politicians, or advisors to make statements of confidence that are clear enough to test their accuracy. Second, even if unambiguous statements of belief were attainable, it is often difficult to obtain data on outcomes that would allow for assessments of the degree of overconfidence in the initial claims. Without these measures, it is impossible to rule out the possibility that advisors are justified in making confident statements and customers are right to prefer them.

Similar to the real world contexts I wish to approximate, in my experimental paradigms advice and outcomes remain decoupled and the accuracy feedback which receivers draw upon is imperfect. However, this framework is likely to have increased the costs of having been wrong, thus reducing the degree of forgetting. This is because the experimental paradigms make it easier for receivers to keep track of producers over time by providing clarity in their advice, clear feedback about the true answer, and more immediate feedback than usually exists. Removing these kinds of ambiguities allows for a clearer, more conservative test of the theories. Later on, I discuss ways in which constraints of this kind may be relaxed as the research moves beyond these initial studies.

Part 1

Competing To Be Certain (But Wrong):

Social Pressure and Overprecision in Judgment

In this section, I establish an experimental paradigm that attempts to capture many of the essential features of the social exchange between producers and receivers of judgment.

Receivers make decisions with the aid of producers, who must compete with each other to exert their influence on potential customers. I focus on the effect of competition between producers and the effect of market forces on producers' expressions of excessive precision. The importance Bush accorded to certainty from the U.S. President should be especially important in contexts that feature the same kind of rivalry as the presidential campaign: namely, those in which candidates must compete with one another.

I specifically hypothesize that markets in which advisors compete with one another will lead to increases in the overprecision of their advice over time, as marked by increasing confidence without corresponding gains in accuracy. The reasons for these predictions are twofold. First, the degree of one's certainty in the market environment depends upon both personal confidence expressions and the confidence presented by alternative sources of judgment, i.e. rival producers. Thus the system rewards advisors not only for being confident but especially for being more confident than their competitors. If everyone is trying to be more confident than everyone else, escalation is likely to follow. Second, advisors can use competitors as exemplars to infer the behaviors necessary for success in the market. Advisors who succeed by expressing highly precise estimates can recognize that they express higher

confidence than their rivals. Conversely, advisors who express imprecise estimates and fail to attract customers can observe the greater certainty expressed by more in demand advisors. Customers will substantiate these patterns by displaying preferences for confident advice (Yaniv and Foster, 1995) that I expect to overshadow the penalty for providing incorrect advice.

While the behavior of advisors in such a marketplace can be viewed as a proper response to the information and incentives they confront, the pursuit of confidence is more problematic for those seeking advice. Since customers' preference for confidence should encourage excessive confidence among advisors, customers' reliance on such overconfident advice will impair the quality of their own judgments and performance.

These predictions are summarized in the following hypotheses:

Hypothesis 1: People will prefer more confident advisors, *ceteris paribus*.

Hypothesis 2: Customers of advice will reward the confident more than they will punish those who are wrong.

Hypothesis 3: In the competitive market, advisors will increase their expressions of confidence over time.

Hypothesis 4: Selecting more confident advisors will impair customer performance.

STUDY 1

Design

I constructed a laboratory market in which decision makers in the role of guesser (receivers/customers analogous to judges in the advice literature) completed eight rounds of an estimation task. In each round, guessers first had the opportunity to select advice from one of four other participants in the role of advisor. Guessers earned money based on the accuracy of

their estimates in each round. Advisors earned money based on the number of guessers in each round who chose to receive their advice.

The task involved estimating the weights of other people based solely on their pictures, a task adapted from an earlier study (Moore and Klein, 2008). Selected photographs spanned a wide range of weight values (127 to 208 pounds) and represented varying levels of difficulty in identifying the correct weight. Participants first viewed a color picture of the individual. After examining the picture, they filled out a decision sheet (see Appendix A) that listed a series of ten pound weight ranges between 120 and 219 pounds (the series also included equivalent values in kilograms). For each of the ranges, they indicated their confidence level (between 0 and 100 percent) that the target's actual weight fell within that particular range.

Participants

Ninety-eight individuals participated in thirteen sessions of the study (35% female; Mean age = 23.76, $SD = 5.75$). They were recruited from a university research pool of individuals in the community interested in participating in studies for pay. I advertised the study as involving “estimation tasks” in which participants would earn money based on decisions made by themselves and others during the course of the session. Each session consisted of four advisors and a variable number of guessers between two and six ($M = 3.54$, $SD = 1.28$).

Procedures

The experimenter randomly assigned four participants to the advisor role and the remaining participants to the guesser role. All participants read instructions that described the weight guessing task and their specific role in detail. The instructions also briefly described the other role and its incentive structure.

At the start of each round, advisors received the picture of one of the target individuals, with the order of the eight targets chosen randomly for a given session. Advisors then provided confidence levels for the various weight ranges on their decision sheet. After collecting all the confidence estimates, the experimenter publicly posted a subset of these estimates for each of the four advisors to serve as a signal of advisor beliefs. At this time, guessers did not see the corresponding weights to these confidence levels. This procedure captures the imprecise signaling that occurs in many real world contexts. Prospective agents, for instance, do not necessarily provide a full explanation of how they would do their jobs and how they would handle every situation that arises. Instead, they can only attempt to convey confidence in the strategies they would implement if hired.

The posted information consisted of each advisor's confidence for three adjacent intervals. The chosen intervals always included the advisor's peak confidence level and two additional intervals so that they included the largest summed confidence of that advisor. Each advisor was randomly assigned a common color (Blue, Green, Red, or Yellow) that identified him or her and remained the same over all eight rounds. This allowed advisors to maintain consistent identities and form reputations. Table 1 contains an illustration of how this sequence unfolded.

Guessers viewed the public signal of advisor confidence and used a computer chat program to communicate their choice of advisor to the experimenter, who sent back the complete confidence distributions of the chosen advisors (including the corresponding weights). Guessers then received the target individual's picture and filled out their personal estimates for the probability that the target's weight fell within each of the ten pound intervals. At the conclusion of the round, the experimenter announced the correct weight of the target individual and the

number of guessers that chose each advisor. To reduce the likelihood of participants intentionally altering their behavior in the final rounds of the task, I kept them unaware of the duration of the task until announcing the conclusion of the study after the eighth round.

I designed the financial incentives for guessers and advisors to mirror those faced by producers and receivers at large. The earnings for guessers increased with accuracy, and were calculated each round using the following quadratic scoring rule: $4p_c - 2\sum p^2$, where p is the probability assigned for a given interval and p_c is the probability assigned to the correct interval. This function rewards guessers for assigning high probabilities to the correct weight interval and penalizes them for assigning high probabilities to incorrect intervals. Participants' instructions told them truthfully: "This formula may appear complicated, but what it means for you is very simple: You get paid more when you provide accurate estimates of the target person's weight." Earnings for advisors were based on their rate of selection, using the formula $2g$, where g is the percentage of guessers that chose to receive the advisor's estimates. This function rewards advisors when more individuals select them and also allows for similar payoffs across sessions with varying numbers of guessers.

Measures

Confidence. I took *peak confidence* as the maximum confidence level individuals assigned to any of the weight intervals for a given target. I also utilized a second measure, *correct confidence*, based on the confidence level individuals assigned to the weight interval containing a target's actual weight. So if for a 145 pound target someone estimated likelihoods of 30%, 60%, and 10% that the target's weight fell in the respective intervals of 140-149 pounds, 150-159 pounds, and 160-169 pounds, the score for this measure would be 30%. The correct confidence value serves as an initial, superficial indicator of accuracy.

Range.² I computed a simple measure for confidence *range* as the number of intervals to which an individual assigned non-zero confidence levels for a given picture.

Accuracy. As a measure of accuracy, I utilized the quadratic scoring rule used to compute guesser payoffs. Recall that this function produces higher values for assigning greater confidence to the correct weight interval and lower values for assigning greater confidence to incorrect intervals.

Selection. Since sessions consisted of different numbers of guessers, I used the percentage of guessers choosing an advisor as the selection variable. I calculated this by dividing the number of guessers that chose a given advisor by the total number of guessers in the market.

Results

Overprecision. I first examined whether advisors displayed overprecision in their weight estimates by contrasting their peak confidence levels to the actual hit rates of those peak confidence levels for the true weight of the given target. Evidence of overprecision emerges strongly. Advisors provided an average peak confidence of 59%, but this peak confidence corresponded to the accurate weight interval only 15% of the time. This difference is revealed to be significant by a paired t-test, $t(51) = 16.44, p < .001$.

Advisor selection. I investigated how advisors' estimates impacted the rate at which they were favored by guessers, utilizing regression analyses (controlling for session and individual advisor effects) with the selection variable as the dependent variable. Model 1 includes the independent variables of peak confidence, accuracy, the number of guessers, and

² Note that I could have used a more intricate measure to capture the distribution of confidence, such as one based on the variance of estimates. I chose the simpler measure of range since several of the analyses address the ways in which guessers utilized advisor estimates. It is reasonable to assume that guessers could recognize the basic spread of confidence but not necessarily more statistically complex measures. As a precaution, I ran alternative models replacing the range measure with a variance based measure yet found no substantive changes to the results.

round number. Model 2 accounts for previous round (lagged) values of peak confidence, accuracy, and selection. Since guessers that selected an advisor beforehand should be able to assess that advisor's accuracy better, I constructed Model 3 that includes an interaction between the lagged accuracy and lagged selection variables. The results of these analyses are summarized in Table 2.

In the full Model 3, the effect of peak confidence is significant ($\beta = 0.29, p < .01$). This provides support for Hypothesis 1, that higher confidence will help advisors attract guessers. Accuracy is not a significant predictor of selection. However, I find some evidence of reputation effects when accounting for the previous round variables. The interaction between lagged accuracy and lagged selection is significant ($\beta = 0.22, p < .05$). This suggests that advisors may be penalized for poorer accuracy in prior rounds, but only when more guessers were exposed to their inferior estimates. Consistent with Hypothesis 2, the selection benefits of being certain are stronger than the adverse effects of being less accurate in previous rounds ($F(1, 12) = 6.27, p < .05$).

Changes in estimates across rounds. Hypothesis 3 predicts that advisors in the market environment will grow more confident with their estimates over time, as evidenced by narrower distributions and higher peak confidence levels. This indeed was the case. Advisor confidence distributions constricted over time. Advisors used an average range of 4.08 intervals in Round One, but decreased to 3.02 intervals in Round Eight. This negative trend for range is significant ($F(1, 12) = 21.58, p < .001$). As shown in Figure 1, peak confidence levels also displayed increasing precision. In Round One, advisors on average offered peak confidence levels of 52% while in Round Eight, their average peak confidence increased to 65%. The linear trend in confidence is significant ($F(1, 12) = 13.62, p < .01$).

This increase in confidence cannot be attributed to increased calibration with the correct target weights. As shown in Figure 1, advisors made no improvements over time for the confidence they provided in the correct weight interval ($F(1, 12) = 1.05, p = .33$).

The accuracy score provides a more fine grained measure of how their estimates approached the correct answers. Accuracy shows no improvement over time (see Figure 2). In fact, the trend is in the direction of worsening accuracy but it is not significant ($F(1, 12) = 0.21, p = .66$). As a whole, these results show advisors expressing higher degrees of confidence over time but not necessarily providing better estimates.

Guesser performance. Advisors clearly benefit from the unwarranted escalations in their expressions of confidence. This is not necessarily detrimental to the exchange relationship as a whole if these inflated estimates fail to directly impair the performance of the guessers relying on those advisors. However, Hypothesis 4 predicts that such displays of confidence by their chosen advisors will adversely affect guessers. Specifically, I expected that higher advisor confidence expressions would have a negative effect on guesser performance (i.e., their payoffs) and this would be driven by the influence of advisor confidence on guessers' own expressions of confidence.

I ran the necessary mediation analyses to test this prediction, as summarized in Table 3. In Model 1, there is a significant negative relationship between advisor peak confidence and guesser performance. Using the criteria offered by Baron and Kenny (1986), the relationship between advisor confidence and guesser performance is completely mediated by the peak confidence expressed by guessers themselves, as shown in Model 2 and Model 3 (Sobel test: $z = -3.14, p < .01$). Thus Hypothesis 4 was supported.

Discussion

The results of Study 1 show a clear pattern of advisor behavior in the market. Advisors' statements acquire more precision over time, as evidenced by higher peak confidence levels and a narrower range of estimates. Further analyses show that greater peak confidence plays a significant role in the proportion of guessers that advisors are able to attract.

Interestingly, guessers generally did not benefit from choosing confident advisors. Instead, guessers that chose advisors expressing high levels of confidence responded by providing higher confidence levels in their own estimates. These overly confident estimates unfortunately diminished their subsequent accuracy and payoffs.

Taken together, these findings demonstrate that the effectiveness of markets for judgment depends largely upon one's vantage point. For advisors, the marketplace offers some relatively clear incentives to which they can adapt over time. Though they may exhibit sizable degrees of overprecision and inflated confidence, these aspects of advisor judgment are facilitators rather than impediments to success. As other researchers have noted (e.g., Yaniv and Foster, 1995; 1997), it is inappropriate to evaluate producers of judgment solely according to criteria associated with accuracy when such producers may be responding to alternative yet equally viable motivations.

The view for receivers appears less ambiguous. Advisor confidence ideally should provide useful information for guessers to utilize in their own decision making. Instead, such confidence drives receiver decision making toward poorer performance. But these wounds are largely self-inflicted. By rewarding advisors for expressing confidence while not adequately penalizing them for being wrong, customers in the market are essentially "getting what they paid for."

STUDY 2

Though the market based study supported initial hypotheses, I cannot be certain that the market environment differentiates itself from other exchange structures. It may be the case that these patterns would generalize to non-market scenarios without competition between advisors. To investigate this possibility, I conducted a second study in which guessers and advisors did not come together through the market but instead acted in stable pairings. Instead of choosing among advisors, guessers decided whether or not to solicit the estimates of a single available advisor.

Design

The study replicated the essential features of Study 1. Over the course of eight rounds, guessers completed the weight estimation task for the same targets and had the opportunity to receive aid from an advisor prior to making their estimates. The key distinction in Study 2 was that advisors supplied private estimates for use by individual guessers instead of operating through the market environment. Thus guessers also relied upon only a single advisor.

Participants

Eighty individuals participated across thirteen sessions (40% female; Mean age = 23.56, $SD = 5.73$). They were recruited from the same research participant pool as Study 1 using a similar advertisement. Sessions consisted of a variable number of advisor-guesser pairs between one and five.

Procedures

The experimenter randomly assigned participants in each session to the role of either guesser or advisor and then randomly paired participants in each role. Sessions progressed in the same manner as Study 1 with the following exceptions. Instead of selecting from all advisors

based on their publicly available information, guessers decided whether or not to receive the estimates of their paired advisors after viewing private information. This private signal consisted of the advisors' confidence levels for three adjacent intervals (without corresponding weights), constructed identically to the public information in Study 1. In other words, guessers saw a set of responses similar to one of the columns at the bottom of Table 1.

I modified the payoff functions of both guessers and advisors to account for these changes. Guessers still earned money for their weight estimates based on the same function. However, each time guessers chose to receive their advisors' complete estimates, they incurred a cost of \$0.25. This feature mirrors the opportunity cost paid by guessers in the market study (in which choosing one advisor's estimate meant forgoing the advice of the alternative advisors). Advisors earned \$2 for each round in which their guesser chose to receive their estimate.

Results

Overprecision. I again examined whether advisors displayed overprecision in their estimates. Overprecision emerges strongly here in Study 2 as well. Advisors provided an average peak confidence of 54%, but this peak confidence corresponded to the accurate weight interval only 20% of the time. This difference is revealed to be significant by a paired t-test, $t(39) = 11.00, p < .001$.

Advisor selection. I also examined what factors affected the likelihood that guessers would choose to receive the estimates of their advisors, corresponding to the analyses done in Study 1. The selection variable for this study took on values of one if the guesser chose the advice and took on values of zero otherwise. I constructed a logit model utilizing selection as the dependent variable with the following predictors (also controlling for individual advisor effects): advisor peak confidence, accuracy, lagged peak confidence, lagged accuracy, lagged selection,

and round number. Model 3 includes the interaction terms between lagged accuracy and lagged selection to mirror the Study 1 analyses. The results of the analysis are summarized in Table 4.

Of the current round variables, only advisor peak confidence significantly increases the likelihood of guessers selecting their advisor estimates ($p < .01$). This suggests that similar to the market study, guessers respond to higher expressions of confidence made by advisors. Several of the variables concerning the previous round also emerge as significant. Interestingly, as opposed to Study 1, the impact of an advisor's accuracy in the previous round did not depend on whether or not the advice actually had been selected. The lagged accuracy X selection interaction is not significant (Model 3, $p = .20$), while the main effect for lagged accuracy is significant (Model 2, $p < .05$).

Finally, I must note a significant negative effect for round ($p < .05$). In later rounds, guessers are less likely to select their advisors. This may reflect the fact that advisor accuracy is relatively poor, especially in early rounds. While 40% of guessers select their advisors in the first round, only 10% of guessers select their advisors in the last round.

Changes in estimates across rounds. Here I examine how advisor estimates evolve over time in the stable exchange relationship and contrast those patterns with the ones in Study 1.

Differences between the two studies emerge for the span of the confidence estimates. Advisors show no changes to the range of their estimates over time ($F(1, 39) = 0.05, p = .83$), which increase from 3.88 intervals in Round One to only 3.93 in Round Eight. The trends in peak confidence across each round for advisors are summarized in Figure 1. These trends show some striking differences from Study 1. Advisors show no significant changes over time for their peak confidence ($F(1, 39) = 0.34, p = .56$), on average 54% in Round One and only 56% in Round Eight. Nor does their confidence assigned to the correct weight interval change

significantly ($F(1, 39) = 0.09, p = .76$). Trends in accuracy are similar to Study 1 (see Figure 2), as advisors display no changes to their accuracy in later rounds ($F(1, 39) = 0.00, p = .99$).

Guesser performance. I also replicated the mediation analyses for guesser performance, as summarized in Table 5. Model 1 reveals a significant negative relationship between advisor peak confidence and guesser performance. Using the criteria offered by Baron and Kenny (1986), the relationship between advisor confidence and guesser performance is completely mediated by the peak confidence expressed by guessers themselves, as shown in Model 2 and Model 3 (Sobel test: $z = -2.85, p < .01$). Interestingly, the impact of advisor confidence on guesser performance is not dependent on guessers actually viewing the advisors' complete estimates, as I find no significant effects when accounting for an interaction between selection and advisor peak confidence ($\beta = 0.255, p = .25$).

Discussion

The results of Study 2 suggest that not all of the findings from the market study generalize to less competitive advice exchange contexts. Unlike the market environment, advisors here showed no changes over time in their estimates in terms of peak confidence or range. This occurred even though advisor selection still depended on a number of the same factors as it did in the market. Advisors expressing greater peak confidence were more likely to be chosen by their paired guessers, as were those who had been more accurate in previous rounds. So advisor behavior did not evolve over time even though the incentives to be more confident remained similar. This suggests the missing pressure of other advisors in the market played a strong role.

The consequences of advisor confidence for guesser performance remained similar as well. When advisors expressed more confidence for their estimates, their paired guessers also

expressed greater certainty, which subsequently resulted in lower payoffs. This occurred regardless of whether or not advisor confidence was seen in conjunction with its corresponding values, suggesting that the influence of advisor confidence at times may be quite subtle and indirect.

While the initial studies produced some interesting findings, it is unclear how the characteristics of the precisely designed market environment constrained those results. In the next section, I present two additional studies which change substantive features of the market context.

Part 2

Changes to the Market Landscape

STUDY 3

The results of Study 1 suggest that excessive expressions of precision increase over time in competitive “judgment markets.” In that study the advisors were relatively uniform in the sense that they all provided their advice from the same common pool of information (i.e., they gave individual estimates across a certain span of intervals). However, it is often the case that individuals will vary in the quality of information they possess. For example, consider projections for the appreciation of a certain company’s stock. Amateur traders may base their estimates on some set of available information (news reports, personal observations, etc.) while analysts at large financial firms will be able to augment their estimates with the aid of superior information (advanced statistics, models, etc.). It may be the case that the presence of better informed advisors will lead to a more stable and efficient market. Such advisors would express greater confidence, provide better estimates, and garner more influence and esteem than their less informed counterparts.

Prior research, however, gives little hope for this ideal set of outcomes. Informed, expert advisors remain plagued by overconfidence (Koehler, Brenner, and Griffin, 2002; Tetlock, 2005). McKenzie, Liersch, and Yaniv (2008) demonstrate how expert and novice estimates provide fundamentally different estimates that still suffer from similar magnitudes of overconfidence. Compared to novices, more expert advisors produce estimates that are more closely centered on the correct answer but are also overly narrow. When combined, these

characteristics lead to a negligible net benefit as hit rates for experts mirror those of novices.

These prior findings lead to the following predictions:

Hypothesis 5: More informed advisors will exhibit overprecision in their estimates, comparable to that expressed by less informed advisors.

Hypothesis 6: More informed advisors will produce narrower estimates than their less informed counterparts.

Hypothesis 7: More informed advisors will produce more accurate estimates than their less informed counterparts.

In addition to these considerations, I used Study 3 as an opportunity to gain further insights into how producers and receivers utilize expressions of confidence in their respective roles. So far, I have done so only indirectly by drawing inferences from quantitative analyses in Study 1 and Study 2. Study 3 supplements these analyses with qualitative response data provided by both advisors and guessers. I examined this response data within a new categorical framework that addresses the intricate tactics employed by producers and receivers of confidence judgments.

Design and Procedures

The study preserved the basic selection and exchange system of Study 1 with a few notable modifications. The estimation task lasted ten rounds instead of eight to allow for an even greater number of exchanges between advisors and guessers. I also utilized a new estimation task to ensure that the results are not confined to the idiosyncrasies of a single task. In each round, advisors and guessers received a target film and estimated in which decade that movie had won the Academy Award for Best Picture. I randomly selected target Best Picture winners from the complete pool of award recipients from 1928-1999, excluding more recent winners as

potentially too easy. Participants indicated their confidence that a movie won the Oscar in each of the eight decades from the 1920s to the 1990s. Advisors completed hard copy decision sheets similar to those used in the first two studies while guessers recorded their estimates on a computer survey platform (to expedite payoff calculations at the end of the session).

The main distinction for Study 3 involved the degree of information endowed to advisors when they made their estimates. Less informed advisors did not receive any information as to the correct answers while the more information advisors had the possible range of decades narrowed to four (see Appendix B). This four decade range consisted of either 1920s-1950s or 1960s-1990s and always included the correct decade. For example, a *High information* advisor would receive answer choices of 1920s, 1930s, 1940s, 1950s for the 1943 Best Picture winner *Casablanca* but would receive answer choices of 1960s, 1970s, 1980s, 1990s for the 1972 Best Picture winner *The Godfather*. Conversely, a *Low information* advisor would receive answer choices of 1920s, 1930s, 1940s, 1950s, 1960s, 1970s, 1980s, 1990s for both *Casablanca* and *The Godfather*. Advisors maintained their *High* or *Low* status throughout a session so that reputation formation could occur. Each session contained two of each type of advisor and I distributed *High information* across all four color codenames (Blue, Green, Red, Yellow) for different sessions.

Payoff functions took the form of scaled down versions of the payoffs in Study 1. The earnings for guessers were calculated each round using the following quadratic scoring rule: $\$2 * p_c - \$1 * \sum p^2$, where p is the probability assigned for a given interval and p_c is the probability assigned to the correct interval, yielding a maximum payoff of \$1 and a minimum payoff of -\$1. Earnings for advisors were based on the formula $\$1 * g$, where g is the percentage of guessers that chose to receive the advisor's estimates in that round.

After the final round of the estimation task, participants completed questionnaires. In addition to some demographic items, both guessers and advisors responded to a free response question. Advisors described how they decided to complete their confidence estimates and guessers described how they chose their advisors.

Participants

One hundred individuals participated in thirteen sessions of the study (48% female; Mean age = 20.20, $SD = 1.13$). They were recruited from a university research pool of individuals interested in participating in studies for course credit. I advertised the study as involving “estimation tasks” in which participants would have the opportunity to earn money based on decisions made during the course of the session (in addition to their hour of research credit). Each session consisted of four advisors and between two and five guessers ($M = 3.69$, $SD = 1.32$).

Quantitative Results

In general, the measures and analyses paralleled those used in the first two studies. The added dummy variable *High information* accounted for whether or not advisors had *High* or *Low* information and took on values of one for the more informed advisors.

Overprecision. As in the previous studies, advisors displayed overprecision in their estimates, as measured by the differences between their average peak confidence levels and the actual hit rates of those peak confidence levels. Advisors provided an average peak confidence of 59%, but this peak confidence corresponded to the accurate decade interval only 39% of the time. This difference is revealed to be significant by a paired t-test, $t(51) = 5.05$, $p < .001$. No significantly different results emerged when separating high and low information advisors, consistent with Hypothesis 5.

Advisor selection. I constructed regression models for selection similar to those in Study 1 but also incorporated the high information dummy variable. Model 1 includes the independent variables of peak confidence, accuracy, and high information while controlling for round and the number of guessers. It is likely that the benefits of possessing more information will emerge over time, so Model 2 includes a High information*Round interaction. I added additional variables in Model 3 accounting for previous round values of peak confidence, accuracy, and selection. Model 4 includes the interaction between lagged accuracy and lagged selection. The results of these analyses are summarized in Table 6.

I concentrate discussion on the full Model 4. Peak confidence is significant ($\beta = 0.412, p < .001$), once again suggesting that expressing higher confidence is an effective means of attracting customers for advice. The interaction between high information and round attains significance ($\beta = 0.344, p < .05$). This suggests that over time, the advisors that received better signals as to the correct answer were able to attract more guessers.

Similar to the results of Study 1, the interaction between lagged accuracy and lagged selection is positive but only at a marginal level of significance ($\beta = -0.137, p < .10$). Thus in the current study advisors still appear to pay a penalty for supplying less accurate estimates in previous rounds, though this evidence is weaker than in the previous study. Again the effect of peak confidence is greater than that of that of lagged accuracy and selection ($F(1, 12) = 12.01, p < .01$). These results provide additional support for Hypothesis 2.

Differences between high and low information advisors and changes in estimates across rounds. I tested whether advisors would grow more confident over time as in Study 1. This was generally the case, as shown in Figure 3. In Round One, advisors on average offered peak confidence levels of 47% while in Round Ten, their average peak confidence increased to

67%. The linear trend in confidence approaches significance ($F(1, 12) = 4.66, p = .05$). This provides some additional support for Hypothesis 3. Interestingly, no significant differences emerge as a consequence of whether or not advisors were in the high information condition (all p 's $> .46$). As in the previous study, the increases in confidence cannot be attributed to increased calibration with the correct target answers. As shown in Figure 3, advisors made no improvements over time for the confidence they provided in the correct decade ($F(1, 12) = 0.15, p = .71$). In fact, this non-significant trend moves downward, meaning advisors indicated slightly less confidence for the correct decades over time. High information advisors did provide higher confidence in the correct decade ($F(1, 12) = 8.25, p < .05$) but showed no difference in this rate over time ($F(1, 12) = 2.13, p = .17$).

Additionally, confidence distributions again constricted over time. Advisors had an average range of 4.15 in Round One which decreased to 3.04 in Round Ten. The negative trend for range is significant ($F(1, 12) = 6.92, p < .05$). Hypothesis 6 predicts that more informed advisors will use narrower ranges for their estimates. *High information* advisors offered narrower ranges ($M = 3.03, SD = 0.83$) than did the less informed advisors ($M = 3.73, SD = 1.58$) ($F(1, 12) = 10.54, p < .01$), supporting the hypothesis. Both types of advisors decreased the range of their advice at the same rate, as indicated by the lack of a significant interaction between high information and round ($F(1, 12) = 0.29, p = .60$).

Accuracy does not improve (see Figure 4) and actually decreases across rounds ($F(1, 12) = 5.51, p < .05$). *High information* advisors show marginal signs of being more accurate overall ($F(1, 12) = 3.70, p < .10$), providing some support for Hypothesis 7. However, by the final round, high information advisors appear no more accurate than their less informed counterparts.

The rate of decreasing accuracy did not change differently for more informed advisors than it did for less informed advisors ($F(1, 12) = 1.96, p = 0.19$).

Introduction to Qualitative Results

After completing the estimation task, participants answered free response questions related to the critical thought processes associated with their respective roles: advisors' expressions of confidence and guessers' selection of advisors. To classify advisor responses, I created a categorical representation tapping into both the actual confidence levels reported (Expression) and the underlying rationale behind the reported confidence (Rationale). I created a separate scheme to capture guesser selection decisions.

Categories are not mutually exclusive since it is possible for single advisors to employ various means of expression and possess divergent, even conflicting rationale for doing so. These may occur simultaneously (e.g., basing estimates on the best guesses of the actual movie years but inflating those confidence levels; craving to attract a high number of guessers in the current round but also desiring to maintain a good reputation for future rounds) or sequentially (e.g., providing low confidence estimates in early rounds and shifting toward high estimates in later rounds; initially wanting to give as accurate estimates as possible and later worrying only about being chosen by guessers).

Using the developed classification scheme, two independent coders recorded when participants' descriptions of their thought processes fit within a given category. The inter-rater reliability was at an acceptable level for both advisor responses (Cohen's $\kappa = .75$) and guesser responses (Cohen's $\kappa = .78$). In cases of coder disagreement, I used the judgments of a third coder to act as a tiebreaker.

In the next sections, I describe the categories in more detail. I have included representative quotations selected from the actual participant responses. To show the pervasiveness of each technique and strategy, I also included the percentage of total advisors and guessers whose responses fall in a given category. This may represent a conservative estimate of usage since advisors may have employed strategies they did not report. Participants alternatively may have mentioned strategies they did not actually employ in practice because of impression management or demand characteristics. Therefore I offer these numbers as an illustrative rather than a definitive glimpse at these strategies in use.

Advisor Expression

The following categories address different facets of the actual expressions of confidence provided by advisors.

Knowledge-based (87%). At the most basic level, advisor estimates represent their actual beliefs about the judgment at hand (e.g., “I just went with the actual % I was sure” and “An honest reflection of my knowledge”). This may stem from conclusive personal evidence (e.g., “I recognized some titles and had a general idea...what time period it was made in”) or mere heuristics (e.g., “If it sounded like a western, I went 60’s and 50’s. If it sounded old, I went old”). Of note, a large percentage of advisors indicate that they base their estimates on actual knowledge, at least in part. However, subsequent categories show that this is by no means the only consideration they take into account.

High confidence (56%). Advisor estimates can exhibit excessive confidence, appearing relatively certain about their answers even if the expressions do not correspond to their actual beliefs. This will be marked by higher peak confidence levels and narrower distributions (e.g., “Always made it a high estimate...” and “Tried to estimate large numbers” and “I try to make

the peak as high as possible”). Such statements suggest that some advisors do so with full awareness, but it is possible others may do so without consciously realizing it.

Tempered confidence (31%). Conversely, some advisors provide confidence that is more muted. This will be marked by lower peak confidence levels and wider distributions (e.g., “I tried not to be overconfident” and “I did not guess to [sic] extreme”). Such estimates may be even more conservative than their actual beliefs.

Guessing (17%). Some advisors claim to not follow any discernible pattern and base their estimates solely on whims at the time of elicitation (e.g., “Guessed mostly” and “Pretty much randomly”).

Advisor Rationale

The preceding categories represent the actual confidence displayed by advisors. But what provides the impetus for these expressions? The following categories capture some of these underlying interests.

Accurate reflection (6%). Advisors may simply strive for their estimates to represent their true opinions and approach the correct answers to the best of their ability (e.g., “I tried to be as honest as I could with the estimates” and “I decided to put more realistic estimates that were more reflective of my knowledge”).

Selection (29%). Advisors may be concerned with being chosen by potential customers (e.g., “I wanted people to pick me” and “Tried to outbid the other [advisors]”). This category encompasses both current instances as well as selection concerns in general.

Reputation (12%). Conversely, advisors may be concerned with the opinions guessers form about their viability as advisors, putting their focus on selection specifically in future

instances (e.g., “If I was wrong, guessers may not have kept choosing me” and “[Inflating confidence] could potentially lose me the trust of the guessers”).

Reflection (12%). Advisors may look back to earlier instances to guide their actions in current instances (e.g., “In the beginning I tried assigning one decade a high percentage...after no one picked me for a couple of rounds I distributed my numbers more evenly” and “Based on the number of people that chose me last round” and “I reduced all my estimates somewhat after getting one wrong”). This can mean either a continuation of successful techniques or a change in tactics after failure (in terms of accuracy, selection, etc.).

Strategic contrast (13%). Advisors may devise more complex patterns of behavior that attempt to use estimates in one round as a springboard for selection in future instances. This will entail mixing up high and low confidence estimates across instances so that expressions of high confidence will be especially noticeable (e.g., “Throw in a lopsided estimate on one I was kind of sure on...Then downplay for a round or two and then go with a confident answer” and “I varied between spreading my numbers fairly thin and in being absurdly overconfident other times”).

Other regarding (2%). Advisors may concern themselves with how their reported estimates impact the performance of those that select and utilize those estimates (e.g., “Not wanting to take anyone else down with me”).

Links Between Advisor Expression and Rationale

I examined how statements relating to advisor rationale corresponded to their expressions of confidence. This was made more difficult by the fact that while all fifty-two advisors provided codable Expression statements, only twenty-eight of them also gave responses that fit in the Rationale categories (consisting of thirty-eight total statements). So though I cannot form

strong conclusions about these relationships, I do conduct some preliminary analyses for illustrative purposes.

The response data offers some evidence that advisors link their current round selection to expressions of high confidence while they link their reputation to expressions of tempered confidence. Advisors indicating selection concerns are more likely to reference high confidence expressions than those who did not (87% vs. 43%, Fisher's Exact Test $p < .01$). Advisors who indicated reputation concerns are more likely (albeit at a marginal level of significance) to reference tempered confidence expressions than those who did not (67% vs. 26%, Fisher's Exact Test $p < .10$). Conversely, those indicating selection are no more likely to reference tempered confidence (33% vs. 30%, Fisher's Exact Test $p > .52$) and advisors indicating reputation are no more likely to reference high confidence expressions (67% vs. 54%, Fisher's Exact Test $p > .45$).

Guesser Selection

The following categories represent different ways in which guessers chose their advisors.

Track record (38%). Guessers may make their choices based on the estimates provided by the advisors and/or the outcomes associated with choosing certain advisors in previous rounds. This could involve renewing a successful advisor partnership or consciously avoiding an advisor who had proved unreliable (e.g., "After three rounds, I discovered that Blue... was usually right. Thus, I always chose Blue for the rest of the rounds" and "Red or Blue seemed most reliable, Green was always way off"). Trust can be a key factor here. Guessers sometimes settle on one particular advisor with whom they are comfortable (described by 17% of total guessers).

High confidence (79%). Guessers may choose advisors that express a high peak confidence level and provide a narrow distribution for their estimates (e.g., "High and focused"

and “The advisor that was most certain about there [sic] estimate by having the highest percentage in one confidence level” and “Advisors who [sic] confidence was not spread over a large number of decades”).

Tempered confidence (19%). Guessers instead may display a preference for advisors who provide lower peak confidence levels and a wider distribution for their estimates (e.g., “Picking the ones with confidence levels that seemed more spread out” and “Choosing something that is not so extreme”).

Variety (8%). Guessers may sample from several different advisors (e.g., “Trying all” and “I tried to divide it up pretty evenly”). They can do so as part of a long term strategy in which they initially assess the pool of available advisors to help guide their subsequent choices. Alternatively, guessers may simply spread their choices around because they do not migrate towards any particular advisor(s).

The following categories surfaced in a very small number of responses in the current study but they may represent more important distinctions in the larger context of these exchanges.

High information (2%). Guessers may choose advisors that possess more information as to the correct judgment (e.g., “They had different time intervals so I picked the ones that seemed [more informed]”). Even if such advisors themselves do not leverage their knowledge into better estimates, guessers may tap into that knowledge to aid their own performance. For example, consider a choice between equally calibrated advisors A and B, where A is better informed as to the correct answer for the target movie. People may prefer A over B because choosing A allows them definitively to narrow down the range of feasible answers.

Herdning (2%). Guessers may use the selection decisions of other guessers to shape their choices (e.g., “I switched to the advisor that consistently had people choosing them”). They may perceive that those guessers “know something they don’t” and follow their lead accordingly or merely feel some social conformity pressure to go along with the majority.

Contrast (2%). Guessers may compare estimates between advisors either cross-sectionally or a single advisor’s estimates longitudinally. Guessers may be drawn to estimates that stand out in some way from those provided by other advisors or the same advisor previously (e.g., “When many were similar, I sometimes picked the most different advisor from the group”).

Comparison Between Advisor and Guesser Responses

Perhaps the most striking convergence between advisors and guessers is their preferences for high confidence over more tempered confidence. Twenty-nine advisors mentioned high confidence while only sixteen referred to tempered confidence (nine discussed both). Guessers meanwhile show an even greater toward high confidence, with thirty-eight citing high confidence and nine indicating tempered confidence (with five discussing both). This suggests a general tendency people have for higher confidence estimates. It also may reflect that advisors, even if they do not personally feel the need to report high confidence, anticipate (correctly) that guessers have a predilection toward such confident displays.

Both sides similarly exhibit a greater focus on current rounds as opposed to previous or future rounds. While fifteen advisors stated concerns about selection, only six conveyed concerns for reputation (only one person mentioned both). Guessers showed an even greater disparity, as evidenced by the difference between reputation statements and confidence expression statements (either high or tempered). Forty-two guessers described basing their

decisions on the actual confidence levels while only eighteen included past performance as part of their thought process (twelve people mentioned both).

It also is noteworthy that the relatively complex strategic contrast rationale reported by a few advisors did not resonate with potential customers, at least in terms of the guesser self-reports. Not a single guesser referenced such inter-round confidence differences as a reason for choosing one advisor over the others. If such contrast did factor in to guessers' decision making processes, it does not appear to have done so at a conscious level. At the very least, advisors seem to have overestimated guessers' explicit awareness to sensitivity in intra-advisor confidence fluctuations across rounds.

Discussion

Results from the third study generally conform to those of Study 1, despite the presence of more informed advisors. The market environment does little to improve overprecision in advisor estimates. Those made in early rounds are not well-calibrated with the correct answers. Over time, their peak confidence levels increase even further, despite the absence of similar gains in accuracy. However, this should not be too surprising considering that overconfidence is often rewarded rather than punished by those choosing among advisors. As a tactic, expressing high confidence clearly offers much to gain and less to lose.

Interestingly, constructing the markets with advisors of different information levels did not fundamentally change these inherent dynamics. More informed advisors offered narrower, slightly more accurate estimates, which is consistent with prior research (e.g., McKenzie, Liersch, and Yaniv, 2008). However, they also exhibited trends in peak confidence and overprecision similar to their less informed colleagues. The results of Study 3 provide some important insights into why this occurs. It is not enough for more expert and informed advisors

to rest on their inherent information advantages. They must express higher confidence to compete with other informed advisors and novice advisors who themselves are attempting to attract guessers in the marketplace. Guessers do appear somewhat sensitive to the disparity in advisor information, as over time their selection of informed advisors increased. It remains unclear whether this stemmed from a conscious identification of these advisors as more informed or from more subtle, unconscious processes.

The qualitative results of Study 3 further support the theory of confidence as an intricate mechanism in the social exchange between producers and receivers of judgment. Previous studies have shown that producers advance different interests through their confidence estimates (e.g., Yaniv and Foster, 1995; 1997), which in turn provide cues to receivers (e.g., Sniezek and Van Swol, 2001). The results support these earlier arguments while expanding and refining the scope of the exchange dynamics. Though most advisors report that they incorporate their actual knowledge into their estimates, they also admit that these distributions often reflect deliberate manipulation according to some notion of what guessers are looking for. This may depend on how advisors weigh their underlying interests. Expressions of high confidence are linked to immediate selection goals while more tempered confidence is associated with future reputation interests. In most cases, this means that advisors attempt to display an inflated level of confidence to improve their prospects in the near term.

However, this seemingly short sighted strategy often leads to success due to the equally limited vision of the consumers of advice. These receivers suggest that they evaluate advisor estimates more in isolation than they do in combination with previous instances. And they value estimates expressed with higher confidence more than those offered with more tempered confidence.

STUDY 4

Up to this point, I have identified a consistent pattern in which judgment producers escalate their confidence over time in the market environment, accounting for this finding under the premise that advisors respond to both the competitive pressure of their rivals and through observing the actions of those other advisors. However, I have yet to examine the individual contribution of this information about other advisors competing in the marketplace. Will access to such intelligence intensify advisor behavior?

Here in Study 4, I sought to answer this question by varying whether advisors could view the estimates offered by competing advisors. Based on the original reasoning, advisors in such an environment should offer more extreme and less accurate estimates than those who are “blind” to their competitors. Thus I offer the following hypotheses:

Hypothesis 8: When advisors have access to information about the responses of their competing advisors, they will exhibit greater degrees of overprecision.

Hypothesis 9: When advisors have access to information about the responses of their competing advisors, they will make estimates with higher peak confidence levels.

Hypothesis 10: When advisors have access to information about the responses of their competing advisors, they will make estimates with narrower ranges.

Hypothesis 11: When advisors have access to information about the responses of their competing advisors, they will make estimates that are less accurate.

Design and Procedures

The study followed the basic market exchange paradigm employed in Study 1 and Study 3 with the following alterations. The estimation task lasted fifteen rounds to further increase the number of exchanges between advisors and guessers. Guessers in the current study did not

necessarily have to choose an advisor every round and instead had the option to receive no advice. I introduced a third estimation task to further gauge the generalizability of the results. In each round, advisors and guessers received a target company and had to estimate in which decile that company was ranked in the 2009 Fortune 100 (a subset of the larger Fortune 500 ranking of the public corporations in the United States that have produced the highest revenues in a given year). To aid in making their estimates, advisors possessed a list of the revenues for the Fortune 100 companies as well as the correct placement of every tenth company in the rankings (see Appendix C). Advisors received this extra information to help legitimize the potential usefulness of their advice to guessers.

I randomly selected target companies from the remaining ninety companies in the rankings.³ Participants indicated their confidence that a given company appeared in each decile of the Fortune 100 (1-10, 11-20...91-100). Advisors and guessers recorded their estimates on a computer survey platform. I also modified the public signal that guessers received about their potential advisors. Instead of providing them with three intervals worth of confidence levels, I expanded the information to include the complete span of advisor estimates ordered from the highest to lowest confidence. So if an advisor provided confidence estimates of 10, 15, 60, 10, 5 in a given round, the public signal for that advisor would be 60, 15, 10, 10, 5.

Payoff functions were equivalent to Study 1. Earnings for advisors were based on the formula $\$2 * g$, where g is the percentage of guessers that chose to receive the advisor's estimates in that round. The earnings for guessers were calculated each round using the following quadratic scoring rule: $\$4 * p_c - \$2 * \sum p^2$, where p is the probability assigned for a given interval and p_c is the probability assigned to the correct interval. However, here guessers incurred a cost of \$0.01 each round that they chose to receive advice.

³ Each randomly selected set of fifteen companies appeared for one group in both the *Control* and *Public* conditions.

The primary manipulation in Study 4 concerned the knowledge advisors possessed about the estimates of other advisors. Advisors in the *Public* condition had the opportunity to see the same public signals that guessers saw. Conversely, advisors in the *Control* condition did not have access to this information about their competitors.

Participants

One hundred fifty nine individuals participated in twenty sessions of the study (47% female; Mean age = 24.36, $SD = 7.52$). They were recruited from a university research pool of individuals interested in participating in studies for pay. I advertised the study as involving “estimation tasks” in which participants would have the opportunity to earn money based on decisions made during the course of the session. Each session consisted of four advisors and between two and five guessers ($M = 3.95$, $SD = 1.07$).

Results

Overprecision. As before, advisors displayed overprecision in their estimates. Advisors in the *Control* condition provided an average peak confidence of 48%, but this peak confidence corresponded to the correct Fortune 100 decile interval only 18% of the time. This difference is revealed to be significant by a paired t-test, $t(39) = 12.36$, $p < .001$. Advisors in the *Public* condition provided an average peak confidence of 57%, but this peak confidence corresponded to the correct interval only 16% of the time. This difference is revealed to be significant by a paired t-test, $t(39) = 15.37$, $p < .001$. Hypothesis 8 predicts that observing the actions of competing advisors will increase overprecision. The degree of overprecision in the *Public* condition exceeds that of the *Control* condition ($F(1, 18) = 4.79$, $p < .05$), consistent with the hypothesis.

Differences between control and public condition advisors and changes in estimates across rounds. I tested whether advisors in the *Public* condition were more confident than those in the *Control* condition and whether confidence would change across rounds. In general, both sets of advisors increased their peak confidence over time, as this linear trend in confidence is significant ($F(1, 19) = 10.49, p < .01$). Moreover, advisors in the *Public* condition exhibited greater peak confidence than those in the *Control* condition ($F(1, 19) = 5.62, p < .05$), consistent with Hypothesis 9. Interestingly, this difference in confidence evolves in a curvilinear manner across rounds, as summarized in Figure 5. Confidence in the *Public* condition increases faster than in the *Control* condition ($F(1, 19) = 5.47, p < .05$). However, the majority of this increase occurs in earlier rounds as confidence flattens out in later rounds, as suggested in a significant negative effect for a squared term for round ($F(1, 19) = 4.18, p = .05$). Conversely, the increase in confidence for the control condition occurs in later rounds, as suggested in a marginally significant negative effect for the squared term for round ($F(1, 19) = 2.93, p = .10$).

There is some evidence that these increases in confidence are not completely unfounded. Advisors make some modest improvements over time for the confidence they provided in the correct interval ($F(1, 19) = 4.11, p < .10$). *Public* condition advisors provided no higher confidence in the correct interval ($F(1, 19) = 0.33, p = .57$).

Confidence distributions again constricted over time. The negative trend for range is significant ($F(1, 19) = 20.43, p < .001$). *Public* condition advisors offered narrower ranges than did the *Control* advisors, albeit at a marginal level of significance ($F(1, 19) = 3.91, p < .10$). These results provide weak support for Hypothesis 10. I find no differences between conditions in the rate at which the range decreases.

Advisor accuracy (see Figure 6) shows no significant changes across rounds ($F(1, 19) = 0.15, p = .71$). As predicted in Hypothesis 11, *Public* advisors were significantly less accurate ($F(1, 19) = 5.50, p < .05$).

Discussion

The results of Study 4 suggest that advisor behavior in the marketplace is closely linked to knowledge about the actions of competing advisors. Compared to when information about the decision making of other advisors was unavailable, advisors in more “public” markets submitted narrower estimates with higher peak confidence levels. These advisors furthermore exhibited greater levels of overprecision and lower degrees of accuracy in their judgments.

It is interesting to note that Study 4 provides the first evidence that satiation points may exist for advisor confidence. Study 1 and Study 3 featured market settings similar to the public condition in Study 4, yet those studies found steady escalation of advisor peak confidence across rounds. A similar pattern was found in the initial rounds of Study 4, roughly coinciding with the duration of the first two market studies (which lasted eight and ten rounds respectively). However, the final rounds instead display a leveling off of advisor confidence that did not manifest in the previous studies. Further research is needed to help explain why this might occur in later rounds.

The findings of the current study also allow us to begin considering mechanisms that might curtail overconfidence in the judgments made in these markets. When advisors were shielded from their competitors, they made estimates that were less extreme and exhibited lower overprecision and greater accuracy. Private correspondence may provide a means for attenuating overprecision and its consequences in exchanges between producers and receivers. While the effectiveness of such techniques may prove effective, their applications are inherently limited

since many markets for influence (e.g., elections, policy debates, financial forecasting) necessitate open dialogues between parties. Therefore additional interventions will be needed to alter behavior in the marketplace.

In the next section, I explore some potential strategies that address these issues.

Part 3

Strategic Interventions in the Market

The competitive advice markets in the previous studies not only failed to eliminate overprecision but also tended to exacerbate it. Most notably, advisors' peak confidence continued to increase over time despite the fact that their accuracy did not. In this section, I examine whether these inefficiencies are an unavoidable element of these exchange systems or if tactical adjustments can improve calibration and performance in the market.

Prior research has considered a number of techniques with the potential to reduce overprecision. Unfortunately, the evidence for their effectiveness is rather weak. Rigorous feedback and training can be successful (Arkes, Christensen, Lai, and Blumer, 1987; Lichtenstein and Fischhoff, 1980), but as Plous (1995) notes, "their applied value is somewhat limited" (p.444). Koriat, Lichtenstein, and Fischhoff (1980) found that participants were less overconfident when asked to provide reasons for alternative choices. Explicit warnings to restrain confidence appear fruitless (Alpert and Raiffa, 1982; Block and Harper, 1991). Looking specifically at group level estimates, Plous (1995) found little support for a variety of techniques including discussion, warnings, reasons, and devil's advocacy. Together, this prior work does not offer a compelling path upon which to direct interventions in the market context.

Luckily, the qualitative results of Study 3 provide some promising targets for possible intervention. Guessers appear to be swayed more by advisors' current expressions of confidence than by the advisors' prior actions. This may stem from the incomplete information they possessed about advisor estimates. Though in both Study 1 and Study 3 guessers knew what

their chosen advisors estimated and the actual answers for those judgments, it remains unclear the extent to which they used that information to make calculations about the actual quality of the advisors. As boundedly rational individuals (see Simon, 1955), it is possible that they only formed impressionistic profiles of the advisors, if they formed any at all. Moreover, guessers only had accuracy information about one advisor each round, namely the one they had selected. This means that even the most diligent guessers were constrained in their ability to accurately assess the market. Detailed performance feedback about the entire slate of advisors might enable guessers to make more informed choices and thus modify advisor behavior as such feedback would make it easier for guessers to evaluate advisors and then reward and punish them accordingly. Even if guessers fail to “keep score” adequately, the mere opportunity for them to do so may be enough to make advisors wary of overextending themselves. Such changes in accountability have manifested in the realm of political punditry, where bloggers and other activists have tapped into the information repository of the internet to document the chronic inaccuracy of many proclaimed authorities in the field.

Despite their caretaker role for guessers, few advisors indicated that concern for those dependents influenced their decision making process. This is not particularly surprising since advisors’ selection based incentives are largely independent from the accuracy of those judgments. It is only through the indirect rewards and sanctions provided by guesser selection over time that advisors need fret about their own accuracy. In some contexts, advisor payoffs instead may depend on both selection and accuracy. For example, sports agents guiding their clients to certain teams receive a commission based on the success of the deal they broker. A payoff structure that links advisors more closely with their actual performance may compel advisors to improve their estimates beyond what mere selection pressure accomplishes.

STUDY 5

The current study evaluates the effectiveness of these feedback and payoff mechanisms for reducing overprecision and improving the inefficiencies of the basic market environment. I specifically test the following predictions concerning the potential improvements offered by these two interventions:

Hypothesis 12a: When advisors are incentivized partially by their accuracy, they will exhibit lower degrees of overprecision in their judgments.

Hypothesis 12b: When receivers obtain enhanced information about the past performance of potential advisors, advisors will exhibit lower degrees of overprecision in their judgments.

Hypothesis 13a: When advisors are incentivized partially by their accuracy, they will make estimates with lower peak confidence levels.

Hypothesis 13b: When receivers obtain enhanced information about the past performance of potential advisors, advisors will make estimates with lower peak confidence levels.

Hypothesis 14a: When advisors are incentivized partially by their accuracy, receivers will achieve higher levels of performance.

Hypothesis 14b: When receivers obtain enhanced information about the past performance of potential advisors, receivers will achieve higher levels of performance.

To examine these issues, I also turn attention to additional outcomes concerning receivers. For example, Sniezek and Buckley (1995) found that receivers choosing more confident advisors were themselves more confident in their own performance. Sniezek and Van

Swol (2001) additionally found that receivers placed more trust in their advisors when those advisors expressed greater confidence.

Gist and Mitchell (1992) describe self-efficacy as “a person’s estimate of his or her capacity to orchestrate performance on a specific task” (p. 183). The authors further identify an important contributor to these beliefs as the assessment of the resources and constraints available to the individuals. The proposed interventions in the market may increase the efficacy of receivers by providing additional resources for them to tap into during their decision making process. Specifically, when advisor incentives depend on the accuracy of their judgments, receivers may perceive their aid as more useful than the “cheap talk” advice that characterizes the standard market environment. Similarly, receivers who possess more knowledge about advisor performance may feel better equipped to select the right advisor and incorporate that advice in forming their own judgments. Thus the following hypotheses:

Hypothesis 15a: When advisors are incentivized partially by their accuracy, receivers will possess greater self-efficacy.

Hypothesis 15b: When receivers obtain enhanced information about the past performance of potential advisors, receivers will possess greater self-efficacy.

Perhaps even more important to the fundamental relationship between producers and receivers is the level of trust between the two parties. To utilize producer judgments to aid in their own decision making, receivers clearly must engage in trusting behavior, based on “the intention to accept vulnerability based upon positive expectations of the intentions or behavior of another” (Rousseau, Sitkin, Burt, and Camerer, 1998: p. 395). I expect that the market interventions will have a divergent impact on receivers’ trust in their advisors. Receivers should have higher positive expectations of advisors if the advisors’ outcomes are dependent on their

own accuracy. Conversely, receivers should have lower expectations for advisors if they are presented with more intelligence about the likely poor performance of potential advisors. These considerations lead to the following hypotheses:

Hypothesis 16a: When advisors are incentivized partially by their accuracy, receivers will place greater trust in their advisors.

Hypothesis 16b: When receivers obtain enhanced information about the past performance of potential advisors, receivers will place less trust in their advisors.

In conducting Study 5, I chose to employ a prediction based estimation task instead of one based on a predetermined number (i.e., weights, years, rankings). This not only provided further means to test the generalizability of the findings, but it more importantly puts advisors in the role of forecasters, a typical and essential function of advisors in the world at large (e.g., Tetlock, 2005).

Design

Study 5 adapts the basic elements of the established market paradigm to a forecasting task. Over a period of twelve consecutive weekdays (each day corresponding to one round), I asked participants in the roles of advisors and predictors (i.e., receivers analogous to guessers in the previous studies) to predict the daily change in the Dow Jones Industrial Average (“the Dow”) on the New York Stock Exchange. For each of ten intervals (-200 or lower, -199.99 to -150.00, -149.99 to -100, -99.99 to -50.00, -49.99 to 0, 0 to + 49.99, + 50.00 to +99.99, +100 to +149.99, +150 to +199.99, +200 or higher), participants indicated their confidence that the actual change in the Dow would fall within that interval. They completed these predictions using an online survey platform. Within these surveys, guessers also selected advisors and viewed the predictions of their chosen advisors.

Participants received instructions and feedback according to their randomly assigned conditions. The *Control* condition closely mirrored the previous studies. Predictor payoffs were based on the function $\$4 * p_c - \$2 * \Sigma p^2$, where p is the probability assigned for a given interval and p_c is the probability assigned to the correct interval. Advisors earned money according to the formula $\$2 * g$, where g is the percentage of guessers that chose to receive the advisor's predictions. At the end of each round, both advisors and guessers received information via email about the actual change in the Dow and the number of guessers that chose each advisor that day.

In the *Feedback* condition, predictors could access more detailed information concerning advisor performance. Along with the basic information that *Control* participants received, these predictors had the opportunity to see the running hit rates for their advisors' peak confidence levels when making their selection decisions.

In the *Payoffs* condition, advisor rewards were based on both selection and accuracy according to the formula $(\$4 * p_c - \$2 * \Sigma p^2) * r$. The first component of this function is the quadratic scoring rule for accuracy, while the r component represents the percentage of receivers that choose to receive advice from that particular advisor in a given round. This formula means that advisors earn more money when their predictions are more accurate and chosen by more guessers and that they lose money when their predictions are more inaccurate and chosen by more people. Predictors in the *Payoffs* condition got the same information and were compensated in the same manner as predictors in the *Control* condition.

Predictors completed additional measures capturing their beliefs about their own efficacy and the trust they placed in their chosen advisors. I utilized a magnitude measure of efficacy (Bandura, 1986) similar to those used in prior research (e.g., Tolli and Schmidt, 2008; Vancouver and Kendall, 2006). Specifically, predictors gave their expected earnings (based on

their accuracy payoff function) in the given round, ranging between -\$2 and \$2. For trust, I adapted the eight item trust measure developed by Sniezek and Van Swol (2001). These items can be found in Appendix D.

Participants

Participants for the predictor role were recruited from a university research pool of individuals interested in participating in studies for pay. Advisors were recruited from the local population of MBA and undergraduate business students at several universities. I advertised the study as involving “prediction tasks” in which participants would earn money based on decisions made during the course of the activity. The advertisements explained that participants would work online, completing a series of actions at certain times each day over the course of a designated three week period. One hundred twenty three individuals participated in the study as advisors and one hundred thirty participated in the role of predictor. I randomly assigned them to “market sessions” consisting of four to six advisors ($M = 4.92$, $SD = 0.57$) and three to eight guessers ($M = 5.20$, $SD = 1.04$).

There were a total of nine market groups in the *Control* condition and eight market groups in each of the other conditions for a grand total of twenty five. Data was collected across three iterations⁴ with three, six and sixteen groups in each of those respective iterations.

Response rates were high, with advisors providing 93% of possible responses (1367 out of 1476) and predictors providing 92% of possible responses (1434 out of 1560).

Procedures

After enrolling in the study, each participant was assigned to a condition, session, and personal ID code. On the first day of the study (prior to the first trading day of scrutiny), participants received instructions via email describing the general prediction task, their roles in

⁴ Iteration start dates were November, 2 2009; November 30, 2009; and February 1, 2010.

the exchange, and the payoff calculations in effect for their groups. For each day/round, advisors provided their initial predictions via the online survey. I used these responses to create the public information for predictors (equivalent to the ordered full range of estimates used in Study 4), which was then emailed to both advisors and predictors, along with the link to the predictors' survey. Once online, predictors selected their advisors (or chose not to receive any advice), viewed the complete distributions of their chosen advisors, and made their personal predictions. Predictors then completed the additional items for efficacy and trust in their advisors. After a given day's trading had ended at 4:00pm Eastern Standard Time, I emailed feedback to all participants concerning the actual change in the Dow and the number of predictors that chose each advisor.

The process then repeated with the solicitation of advisor predictions for the next day's Dow activity. Participants were told the task could last for up to three full weeks (i.e., 15 weekdays/rounds), but the total number of rounds remained uncertain so as to guard against changes in behavior specifically motivated by an "endgame" scenario. The actual end of the study occurred after the twelfth round of judgments. At the conclusion of the study, participants received notification for collecting their total earnings and were thanked for their involvement in the study.

Results

Overprecision. Consistent with the previous studies, advisors displayed overprecision in their estimates. Advisors in the *Control* condition provided an average peak confidence of 47%, but this peak confidence corresponded to the actual change in the Dow only 17% of the time. This difference is revealed to be significant by a paired t-test, $t(43) = 10.44, p < .001$. Similar patterns also held for the two other conditions. Advisors in the *Payoff* condition provided an

average peak confidence of 44%, but this peak confidence corresponded to the correct interval only 22% of the time. This difference is revealed to be significant by a paired t-test, $t(38) = 8.10, p < .001$. Advisors in the *Feedback* condition provided an average peak confidence of 48%, but this peak confidence corresponded to the correct interval only 22% of the time. This difference is revealed to be significant by a paired t-test, $t(39) = 7.05, p < .001$. Hypothesis 12a and 12b respectively predict that the accuracy incentive and feedback interventions would reduce overprecision in advisor estimates. These were not supported, as there were no significant differences in the degree of overprecision between the three conditions ($F(2, 21) = 1.47, p = .25$).

Differences in confidence between advisor conditions and changes across rounds. I tested whether advisors in the *Control* condition expressed higher confidence than those in the *Payoff* and *Feedback* conditions and whether the difference varied across rounds. These results are summarized in Figure 7. Hypothesis 13a and 13b respectively predict that the accuracy incentive and feedback interventions would decrease the expressed confidence of advisors. The hypotheses were not supported. Peak confidence levels for the *Payoff* ($F(1, 24) = 1.83, p = .19$) and *Feedback* ($F(1, 24) = 0.02, p = .89$) conditions were no different than the confidence expressed by advisors in the *Control* condition. Moreover, in stark contrast to the market studies presented up to this point, advisors failed to increase their peak confidence over time, as the linear trend in confidence is not significant ($F(1, 24) = 0.13, p = .72$). While such a finding superficially appears contrary to the findings of the previous studies, a deeper investigation actually reveals consistency. Utilizing selection models similar to those constructed in the other studies (which are not covered in detail here), I found that here in Study 5 advisor peak confidence does not have a significant positive effect on their selection ($B = .10, p = .14$). Since

these selection benefits should help to cultivate inflated confidence levels in the market, it should not be surprising when this escalation fails to materialize when those incentives are lacking.

Predictor performance. I examined how the intervention treatments impacted the effectiveness of predictor judgments by utilizing regression analyses (controlling for session, iteration, and individual advisor effects) with the predictor accuracy variable as the dependent variable. These results are summarized in Table 7. Model 1 addresses cases in which predictors refused advice and includes independent variables for number of available advisors, round, predictor peak confidence, a payoff condition indicator, and a feedback condition indicator. Model 2 includes identical variables for those cases in which predictors chose to receive advice and additional variables accounting for the peak confidence and accuracy of the predictors' chosen advisors.

Several variables emerge as significant predictors of accuracy for predictors not receiving advice. Round has a positive effect ($\beta = 0.20, p < .01$), indicating that these predictors improve over time in their judgments. However, their best estimates are not generally well calibrated, as predictor peak confidence has a marginally significant negative impact on accuracy ($\beta = -0.36, p < .10$). Interestingly, the feedback variable also is marginally significant ($\beta = 0.48, p < .10$). This suggests that this type of intervention may benefit predictors even when they refrain from choosing an advisor in a given round.

Similar to the results of non-choosers, predictors' own confidence has a significant negative effect on their own performance ($\beta = -0.39, p < .001$) when choosing to receive advice. Predictors conversely show performance gains based on the estimates of their chosen advisors, specifically when those advisors are more confident ($\beta = 0.29, p < .001$) and more accurate themselves ($\beta = 0.69, p < .001$). Hypothesis 14a and 14b respectively predict that the accuracy

incentive and feedback interventions would improve receiver performance. Both the *payoff* ($\beta = 0.21, p < .001$) and *feedback* ($\beta = 0.14, p < .001$) variables are positive and significant, providing support for the hypotheses.

Predictor efficacy. Table 8 summarizes the analyses of predictor efficacy. The model includes the following independent variables (controlling for session, iteration, and individual advisor effects): number of available advisors, round, predictor peak confidence, payoff condition indicator, feedback condition indicator, chosen advisor peak confidence, chosen advisor accuracy, chosen advisor lagged selection, and chosen advisor lagged accuracy. Since the model includes these variables concerning chosen advisors, these analyses only include cases in which predictors decided to receive advice.

Of the advisor variables, only lagged advisor accuracy has a significant effect on predictor efficacy ($\beta = 0.05, p < .05$). Predictors were more confident in their ability to perform when they had chosen advisors who were more accurate in the preceding round. The confidence expressed by their advisors did not affect predictor efficacy ($p = .35$), though this advisor confidence did impact the actual performance of predictors as seen above.

Hypothesis 15a, predicting that the accuracy intervention will increase efficacy, was not supported. The payoff variable is significant but negative ($\beta = -0.13, p < .05$), meaning that predictors in that condition were less confident in their ability to perform. Conversely, Hypothesis 15b, predicting that the feedback intervention will increase efficacy, was supported. The feedback variable was significant and positive ($\beta = 0.28, p < .001$), suggesting predictors in the *Feedback* condition were more confident in their ability to perform.

Predictor trust. The results addressing predictors' trust in their chosen advisors appear in Table 9. The model includes the following independent variables (controlling for session,

iteration, and individual advisor effects): number of available advisors, round, predictor peak confidence, payoff condition indicator, feedback condition indicator, chosen advisor peak confidence, chosen advisor accuracy, chosen advisor lagged selection, and chosen advisor lagged accuracy.

Several of the advisor specific variables emerge as significant. Predictors place more trust in their advisors when those advisors were more accurate in the preceding round ($\beta = 0.09$, $p < .001$) and the advisors were chosen by more predictors in the preceding round ($\beta = 0.05$, $p < .05$). Hypothesis 16a, which predicted that the accuracy intervention would increase receivers' trust in their chosen advisors, was not supported. The payoff variable is significant but negative ($\beta = -0.22$, $p < .001$), meaning that predictors trusted their advisors less when those advisors possessed accuracy-based incentives. Hypothesis 16b predicted that the feedback intervention would reduce receivers' trust in their chosen advisors. The feedback variable was significant and negative ($\beta = -0.48$, $p < .001$), indicating that predictors indeed trusted their advisors less when they possessed more information about the past performance of the pool of available advisors. Thus Hypothesis 16b was supported.

Discussion

The results of Study 5 provide an inconclusive picture as to the effectiveness of the proposed market intervention strategies. The interventions did little to change the ways advisors behaved as compared to the basic market environment (the control condition). Advisors in all conditions displayed similar magnitudes of overprecision and levels of peak confidence.

However, it is unclear whether this pattern reflects poorly on the interventions themselves or instead arises from some other factors that suppressed the estimates in the *Control* condition.

Predictors in Study 5 did not select advisors more when they expressed greater confidence,⁵ removing a strong incentive for advisors to increase the precision of their estimates. It also is possible that advisors possessed some internalized motivation for accuracy that transcended their condition. Since advisors were business students, they may have wanted to offer more moderate estimates (which would be more likely to have higher accuracy) that reflected well on their abilities to analyze the stock market. These concerns may have superseded their selection motivation.⁶ As an interesting finding along these lines, advisor accuracy in Study 5 actually increases across rounds. This marks the first study in the series in which such an improvement occurs.

Despite these shortcomings on the advisor side, the interventions do show some promise as a means to improve the lot of potential advice recipients. Predictors in the *Payoff* and *Feedback* conditions achieved higher performance than those in the standard market. One key reason for this involves the lower trust advisors built among predictors. As expected, predictors confronted with specific information about the deficiencies of their advisors trusted less. Contrary to expectations, predictors also trusted less when their advisors were incentivized based on their accuracy in addition to selection. The reason why this type of intervention would arouse such doubt in the minds of predictors remains an open question. From a utilitarian perspective, it is enough to recognize the potential improvements offered by these interventions to the end consumers.

Clearly, future research should continue to explore their ramifications and continue to develop new strategies with the same objectives.

⁵ The reasons why this might have occurred is an important question in its own right, but outside the scope of the current research.

⁶ Though I did not have the opportunity to collect response data from advisors, informal post-task discussions with advisors suggested a higher engagement with the forecasting aspects of the task.

Conclusion

As economists back to Adam Smith (1776) have pointed out, markets can cure many ills. Some individual biases present in human judgment have less impact in market settings (Gode and Sunder, 1993; Plott, 1995), and markets can certainly concentrate the rewards to popular products, services, or personalities (Frank and Cook, 1995). But markets are not panaceas. There are circumstances under which markets fail entirely (e.g., Akerlof, 1970). Here I present an example in which market competition magnifies a bias present in individual judgments. The individual bias in question is overprecision. Human judgment is prone to overprecision (Soll and Klayman, 2004) and I observed significant degrees of overprecision in the advice offered across the various studies. However, this bias in judgment was magnified by the presence of the competitive markets for advice. Consumers furthermore tended to pick the advice from those who expressed more confidence that they had the right answer, much to their own detriment.

Limitations and Future Directions

The studies presented here examine the interface between producers and receivers of judgment, as manifests across numerous social contexts. I adopted a novel framework that more closely captures the inherent features of these exchanges than has been accomplished in previous research. In this series of studies I chose to focus on the actual quantitative estimates as the sole information by which producers and receivers communicate and evaluate each other. Such streamlining was necessary in these early stages to make clearer tests of the predictions.

As this line of research progresses, researchers should begin to take into account some of the actual interactions between the two parties in the social exchange. Prior studies have shown that communication can play an important role in how partners address unmet expectations and

restore cooperation (e.g., Bottom, Gibson, Daniels, and Murnighan, 2002). As it applies to this environment, producers often rely on several communication strategies to retain their influence in the face of failure.

Advisors may rationalize away their shortcomings by attributing them to some kind of mitigating circumstances. Offering such justifications and counterfactual scenarios allows them to claim that their chosen course of action would have been correct “if only X had (or had not) occurred” (see Tetlock, 2005). Rather than having been wrong, advisors can instead claim that they were “almost right.” For example, proponents of the Iraq War contend that their visions of success would have been realized if not for various unforeseen strategic missteps (e.g., initial shortages of troops on the ground (Cloud and Schmitt, 2006)) and polarizing incidents (e.g., the bombing of the al-Askari Mosque (Weisman and Worth, 2006)).

Alternatively, advisors may attempt to convince customers that “this time it’s different.” In other words, advisors could argue that their previous failures should not be held against them because the situation in which they previously appeared confident but wrong does not apply to the current environment. This claim is, of course, the hallmark of economic bubbles. Stock analysts during the dot-com boom argued that old ways of measuring the value of stocks did not apply any more and the new business environment justified the grandiose valuations common at the time (Glassman and Hassett, 1999). During the real-estate boom of the mid-2000s, real estate agents eagerly provided advice on how to buy and sell homes, confident that the real estate market would continue to go up indefinitely (Roberts and Kraynak, 2006; Kemp, 2007). In the end, these predictions were contradicted by the evidence, but some of these advisors were able to get very rich offering their advice in the mean time.

In the experimental paradigm, advisors had no avenues of communication through which they could articulate any of these arguments. While their persistent claims of greater confidence, despite their unimpressive prior accuracy, have similar implications on their own, it is important to expand these social interactions to their fullest manifestation.

Additional research also is needed to accounts for different types of exchange between producers and receivers. Sniezek and Buckley (1995) highlight three distinct conditions as they apply to advisors and advisees: Independent, Dependent, and Cued. Receivers in Independent exchanges are able to evaluate decision alternatives prior to ever receiving advice while Dependent receivers only can draw upon their advisors' evaluations without drawing upon their own information about the alternatives. Between these two extremes lie Cued advice scenarios in which receivers can draw upon personal information but only after they have received information from their advisors. The current studies clearly fall in this third middle ground category, so future studies could examine how advice markets may differ when one of these alternative exchange structures is in place.

While the contributions of the current studies are noteworthy, clearly they represent the beginning stages of a long line of vital research.

Final Thoughts

Russo and Schoemaker (1992) make the case that estimators can and should curtail overconfidence, which also holds true for the primary decision makers (such as managers) who rely on those estimates. The studies here call into question whether those authors' appeal to "metaknowledge" is the most fitting response on either side. It is appropriate for decision makers to recognize and account for the overconfident information brought forth by their advisors, but it is perhaps more important to address the systems that fuel and sustain this

overconfidence in the first place. As shown here, differences in the composition of these exchange systems can have a significant influence on the estimates they generate. For those selling their advice, overconfidence seems to provide clear benefits. With incentive structures in place that reward such misplaced confidence, there should be no surprise that overconfidence remains such a potent and pervasive force.

If any further evidence is needed, we need only turn our attention back to presidential politics. Many American voters reported that the hubris of the Bush's unfailing self-assurance helped accelerate the collapse in his popularity when the war in Iraq and the U.S. economy fared so poorly. John McCain, the Republican candidate to succeed Bush as President, struggled with how to position his candidacy given McCain's prior support for Bush and Bush's low popularity. McCain's campaign was marked by inconsistencies in his message and reversals in his campaign strategy that stood in marked contrast to his Democratic opponent, Barack Obama. By contrast, political observers noted that Obama displayed an unflappable self-assurance throughout the campaign (Kantor, 2008). And we all know who won the Presidency in 2008.

References

- Akerlof, G. (1970). The market for lemons: Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics*, 89, 488-500.
- Alpert, M., and Raiffa, H. (1982). A progress report on the training of probability assessors. In D. Kahneman, P. Slovic, and A. Tversky (Eds.). *Judgment under uncertainty: Heuristics and biases* (pp. 294-305). Cambridge, England: Cambridge University Press.
- Andrews, E. L., and Baker P. (2009, March 14). A.I.G. planning huge bonuses after \$170 billion bailout. *The New York Times*. Retrieved July 7, 2009 from <http://www.nytimes.com/2009/03/15/business/15AIG.html>.
- Arkes, H. R., Christensen, C., Lai, C., and Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, 39: 133-144.
- Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice Hall.
- Baron, R. M., and Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51(6): 1173-1182.
- Block, R. A., and Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring-and-adjustment hypothesis. *Organizational Behavior and Human Decision Processes*, 49: 188-207.
- Bogle, J. C. (1999). *Common Sense of Mutual Funds*. New York, NY: John Wiley and Sons, Inc.

- Bornstein, B. H., and Zickafoose, D. J. (1999). "I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied*, 5(1), 76-88.
- Brewer, N., and Wells, G. L. (2006). The Confidence-Accuracy Relationship in Eyewitness Identification: Effects of Lineup Instructions, Foil Similarity, and Target-Absent Base Rates. *Journal of Experimental Psychology: Applied*, 12(1), 11-30.
- Bottom, W. P., Gibson, K., Daniels, S. E., and Murnighan, J. K. (2002). When talk is not cheap: Substantive penance and expressions of intent in rebuilding cooperation. *Organization Science*, 13: 497-599.
- Chambers, J. R., and Windschitl, P. D. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, 130(5), 813-838.
- Cloud, D. S., and Schmitt, E. (2006, April 14). More retired generals call for Rumsfeld's resignation. *The New York Times*. Retrieved April 7, 2010 from <http://www.nytimes.com/2006/04/14/washington/14military.html>.
- Conger, J. A., and Kanungo, R. N. (1987). Toward a behavioral theory of charismatic leadership in organizational settings. *Academy of Management Review*, 12(4), 637-647.
- DeBondt, W. F. M., and Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. In R. A. Jarrow, V. Maksimovic and W. T. Ziemba (Eds.), *Finance, Handbooks in Operations Research and management Science* (Vol. 9, pp. 385-410). North Holland, Amsterdam: Elsevier.

- De Bondt, W. F. M., and Thaler, R. H. (2002). Do analysts overreact? In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases*. (pp. 678-685). Cambridge, England: Cambridge University Press
- DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., and Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review*, 1, 346-357.
- Flynn, F. J., and Staw, B. M. (2004). Lend me your wallets: The effect of charismatic leadership on external support for an organization. *Strategic Management Journal*, 25, 309-330.
- Frank, R. H., and Cook, P. J. (1995). *The winner-take-all society: Why the few at the top get so much more than the rest of us*. New York: Penguin.
- Gist, M. E., and Mitchell, T. R. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *The Academy of Management Review*, 17: 183-211.
- Glassman, J. K., and Hassett, K. A. (1999). *Dow 36,000: The new strategy for profiting from the coming rise in the stock market*. New York: Times Business.
- Gode, D. K., and Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101(1), 119-137.
- Hayward, M. L. A., and Hambrick, D. C. (1997). Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Administrative Science Quarterly*, 42, 103-127.
- House, R. J. (1977). A 1976 theory of charismatic leadership. In J. G. Hunt and L. Larson (Eds.), *Leadership: The cutting edge* (pp. 189-204). Carbondale, IL: Southern Illinois University Press.
- Kantor, J. (2008, August 27). For a new political age, a self-made man. *New York Times*.

- Kemp, K. (2007). *Flipping confidential: The secrets of renovating property for profit in any market*. New York: Wiley.
- Khurana, R. (2004). *Searching for a corporate savior: The irrational quest for charismatic CEOs*. Princeton, NJ: Princeton University Press.
- Koehler, D. J., Brenner, L., and Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. In T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and biases*. (pp. 686-715). Cambridge, England. Cambridge University Press.
- Koriat, A., Lichtenstein, S., and Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6: 107-118.
- Lichtenstein, S., and Fischhoff, B. (1980). Training for calibration. *Organizational Behavior and Human Performance*, 26: 149-171.
- Lindsay, D. S., Read, J. D., and Sharma, K. (1998). Accuracy and Confidence in Person Identification: The Relationship Is Strong When Witnessing Conditions Vary Widely. *Psychological Science*, 9(3), 215-218.
- London, H., McSeveney, D., and Tropper, R. (1971). Confidence, overconfidence and persuasion. *Human Relations*, 24 (5): 359-369.
- Malkiel, B.G. (1996). *A random walk down wall street*. New York, NY: W. W. Norton and Company, Inc.
- Malmendier, U., and Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*(60), 6.

- McKenzie, C. R. M, Liersch, M. J., and Yaniv I. (2008). Overconfidence in interval estimates: What does expertise buy you?" *Organizational Behavior and Human Decision Processes*, 107: 179-191.
- Meindl, J. R., Ehrlich, S. B., and Dukerich, J. M. (1985). The romance of leadership. *Administrative Science Quarterly*, 30: 78-102.
- Micklethwait, J., and Woolridge, A. (1996). *The witch doctors: Making sense of the management gurus*. New York: Times Books.
- Moore, D. A. (2007). Not so above average after all: When people believe they are worse than average and its implications for theories of bias in social comparison. *Organizational Behavior and Human Decision Processes*. 102: 42-58.
- Moore, D. A., and Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, 115(2), 502-517.
- Moore, D. A., and Klein, W. M. P. (2008). The use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior and Human Decision Processes*, 107, 60-74.
- Moscovici, S. (1976). *Social influence and social change*. London: Academic Press.
- Pfeffer, J. (1977). The ambiguity of leadership. *Academy of Management Review*, 2: 104-112.
- Pfeffer, J. (1992). *Managing with power: Politics and influence in organizations*. Boston: Harvard Business School Press.
- Plott, C. (1995). Industrial organization: A survey of laboratory research. In J. H. Kagel and A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

- Plous, S. (1995). A comparison of strategies for reducing interval overconfidence in group judgments. *Journal of Applied Psychology*, 80: 443-454.
- Price, P. C., and Stone, E. R. (2004). Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *Journal of Behavioral Decision Making*, 17, 39-57.
- Ricks, T. E. (2006). *Fiasco: The American military adventure in Iraq*. New York: Penguin.
- Roberts, R. R., and Kraynak, J. (2006). *Flipping houses for dummies*. New York: Wiley.
- Rousseau, D. M., Sitkin, S. B., Burt, R. S., and Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review*, 23: 393-404.
- Russo, J. E., and Schoemaker, P. J. H. (1992). Managing overconfidence. *Sloan Management Review*, 33(2), 7-17.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.
- Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London: Strahan and Cadell.
- Sniezek, J. A., and Buckley, T. (1995). Cueing and cognitive conflict in judge-advisor decision-making. *Organizational Behavior and Human Decision Processes*, 62: 159-174.
- Sniezek, J. A., and Van Swol, L. M. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, 84(2), 288-307.
- Soll, J. B., and Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(2), 299-314.

- South Florida Sun-Sentinel. (2004, October 1). *Read the complete first presidential debate*. Retrieved November 25, 2008, from <http://www.dailypress.com/topic/sfl-930debatetranscript,0,7938101.story?page=1>
- Tenney, E. R., MacCoun, R. J., Spellman, B. A., and Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*(1), 46-50.
- Tenney, E. R., Spellman, B. A., and MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology, 44*, 1368-1375.
- Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton, NJ: Princeton University Press.
- Tolli, A. P., and Schmidt, A. M. (2008). The role of feedback, causal attributions, and self-efficacy in goal revision. *Journal of Applied Psychology, 93*: 692-704.
- Vancouver, J. B., and Kendall, L. N. (2006). When self-efficacy negatively relates to motivation and performance in a learning context. *Journal of Applied Psychology, 5*: 1146-1153.
- Vekshin, A. (2009, December 16). U.S. senators propose reinstating Glass-Steagall Act. Retrieved April 7, 2010 from <http://www.bloomberg.com/apps/news?pid=20601103&andsid=aQfRyxBZs5uc>.
- Weisman, S. R., and Worth, R. F. (2006, February 24). Iraqi violence strains U.S. political and military strategy and imperils pullout plans. *The New York Times*. Retrieved April 7, 2010 from <http://www.nytimes.com/2006/02/24/politics/24diplo.html>.
- Wells, G. L., and Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295.

- Westley, F., and Mintzberg, H. (1989). Visionary leadership and strategic management. *Strategic Management Journal*, 10, 17-32.
- Wood, A. S. (1989). Fatal attractions for money managers. *Financial Analysts Journal*, 45: 3-5.
- Yaniv, I., and Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, 124(4), 424–432.
- Yaniv, I., and Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making*, 10(1), 21-32.

Tables

Table 1. Outline of the Market Interface in Study 1

Weight	Advisor Blue Confidence %	Advisor Green Confidence %	Advisor Red Confidence %	Advisor Yellow Confidence %
120-129 pounds		5		
130-139 pounds		5		
140-149 pounds		20		
150-159 pounds		50	20	
160-169 pounds	15	10	20	
170-179 pounds	70	5	40	
180-189 pounds	15	5	10	100
190-199 pounds			10	
200-209 pounds				
210-219 pounds				

Example public information for guessers	Advisor Blue	Advisor Green	Advisor Red	Advisor Yellow
	15	20	20	
	70	50	20	100
	15	10	40	

Table 2. Regression results for advisor selection in Study 1

	Model 1	Model 2	Model 3
Round	-0.077*	-0.078†	-0.077†
Number of guessers	0.115*	0.117	0.046
Peak confidence	0.300**	0.294**	0.293**
Accuracy	-0.014	-0.023	-0.026
Lagged selection		-0.201**	-0.152*
Lagged peak confidence		0.063	0.045
Lagged accuracy		-0.016	-0.172†
Lagged accuracy*Lagged selection			0.221*
R^2	0.224	0.281	0.302

Note: Standardized β weights for independent variables. Controls for individual advisors are included in all models but not shown. Standard errors clustered by market session.

† $p < .10$

* $p < .05$

** $p < .01$

*** $p < .001$

Table 3. Mediation test for guesser payoffs in Study 1

	Model 1	Model 2	Model 3
Dependent variable	Guesser performance	Guesser peak confidence	Guesser performance
Advisor peak confidence	-0.169*	0.353***	-0.051
Guesser peak confidence			-0.337**
R^2	0.029	0.124	0.128

Note: Standardized β weights for independent variables.

* $p < .05$

** $p < .01$

*** $p < .001$

Table 4. Logit regression results for advisor selection in Study 2

	Model 1	Model 2	Model 3
Round	-0.106***	-0.037*	-0.179*
Peak confidence	5.780*	7.313**	4.522**
Accuracy	0.367	0.406	0.099
Lagged selection		1.535	0.650
Lagged peak confidence		0.683	-1.421
Lagged accuracy		0.739*	0.217
Lagged accuracy*Lagged selection			0.513
<i>Pseudo-R</i> ²	0.067	0.093	0.100

* $p < .05$

** $p < .01$

*** $p < .001$

Table 5. Mediation test for guesser payoffs in Study 2

	Model 1	Model 2	Model 3
Dependent variable	Guesser performance	Guesser peak confidence	Guesser performance
Advisor peak confidence	-0.121*	0.287***	-0.063***
Guesser peak confidence			-0.201
R^2	0.015	0.083	0.052

Note: Standardized β weights for independent variables.

* $p < .05$

** $p < .01$

*** $p < .001$

Table 6. Regression results for advisor selection in Study 3

	Model 1	Model 2	Model 3	Model 4
Round	-0.077†	-0.205**	-0.152*	-0.156*
Number of guessers	0.415***	0.408***	0.244*	0.252*
Peak confidence	0.447***	0.446***	0.417***	0.412***
Accuracy	-0.034	-0.025	-0.012	-0.019
Informed	0.217***	-0.034	0.004	0.025
Informed*Round		0.309*	0.334*	0.344*
Lagged selection			-0.182**	-0.191**
Lagged peak confidence			-0.036	-0.047
Lagged accuracy			0.049	-0.052
Lagged accuracy*Lagged selection				0.137†
R^2	0.281	0.298	0.348	0.355

Note: Standardized β weights for independent variables. Controls for individual advisors are included in all models but not shown. Standard errors clustered by market session.

† $p < .10$

* $p < .05$

** $p < .01$

*** $p < .001$

Table 7. Regression results for predictor performance in Study 5

	Predictors refusing advice	Predictors choosing advice
	Model 1	Model 2
Round	0.199**	0.034
Number of advisors	0.060	-0.012
Predictor peak confidence	-0.357†	-0.385***
Payoff condition	-0.277	0.215***
Feedback condition	0.478†	0.137***
Chosen advisor peak confidence		0.294***
Chosen advisor accuracy		0.692***
R^2	0.384	0.605

Note: Standardized β weights for independent variables. Controls for individual predictors and iterations are included in all models but not shown. Standard errors clustered by market session.

† $p < .10$

* $p < .05$

** $p < .01$

*** $p < .001$

Table 8. Regression results for predictor efficacy in Study 5

	Model
Round	-0.038
Number of advisors	-0.101
Predictor peak confidence	0.100
Payoff condition	-0.126*
Feedback condition	0.277***
Chosen advisor peak confidence	0.036
Chosen advisor accuracy	0.015
Chosen advisor lagged selection	-0.009
Chosen advisor lagged accuracy	0.045*
R^2	0.577

Note: Standardized β weights for independent variables. Controls for individual predictors and iterations are included in all models but not shown. Standard errors clustered by market session.

† $p < .10$

* $p < .05$

** $p < .01$

*** $p < .001$

Table 9. Regression results for predictor trust in their chosen advisors in Study 5

	Model
Round	0.020
Number of advisors	0.062
Predictor peak confidence	0.029
Payoff condition	-0.217***
Feedback condition	-0.476***
Chosen advisor peak confidence	-0.026
Chosen advisor accuracy	-0.045†
Chosen advisor lagged selection	0.049*
Chosen advisor lagged accuracy	0.087***
R^2	0.728

Note: Standardized β weights for independent variables. Controls for individual predictors and iterations are included in all models but not shown. Standard errors clustered by market session.

† $p < .10$

* $p < .05$

** $p < .01$

*** $p < .001$

Figures

Figure 1. Peak and correct interval confidence across rounds for advisors in Study 1 and Study 2

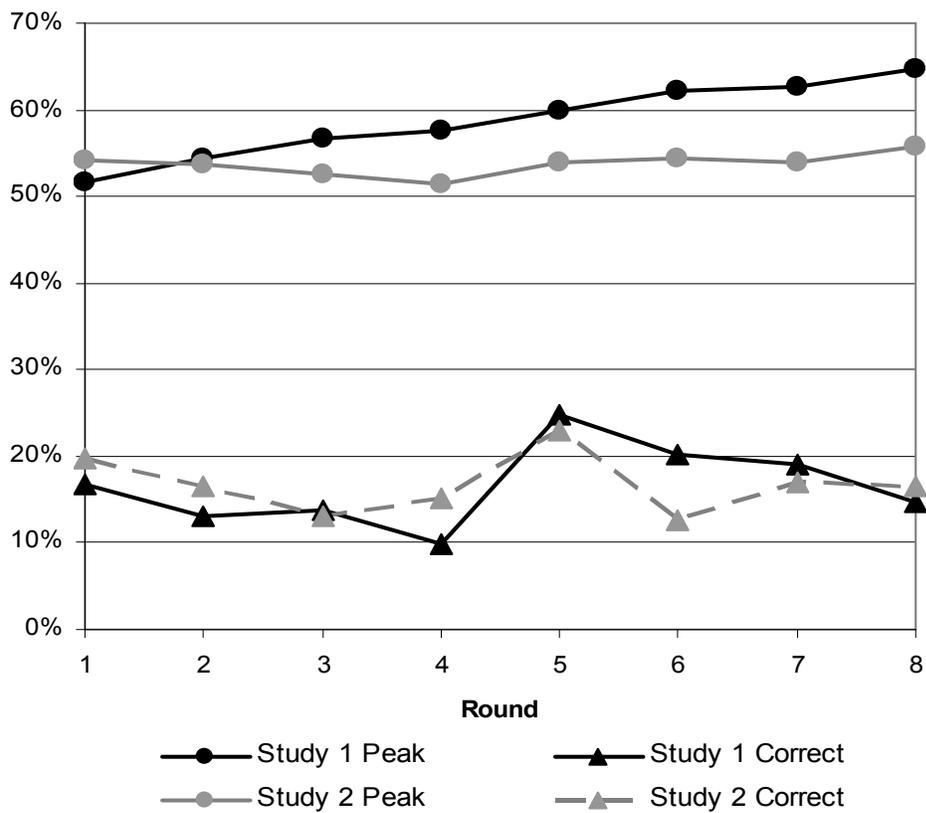


Figure 2. Accuracy across rounds for advisors in Study 1 and Study 2

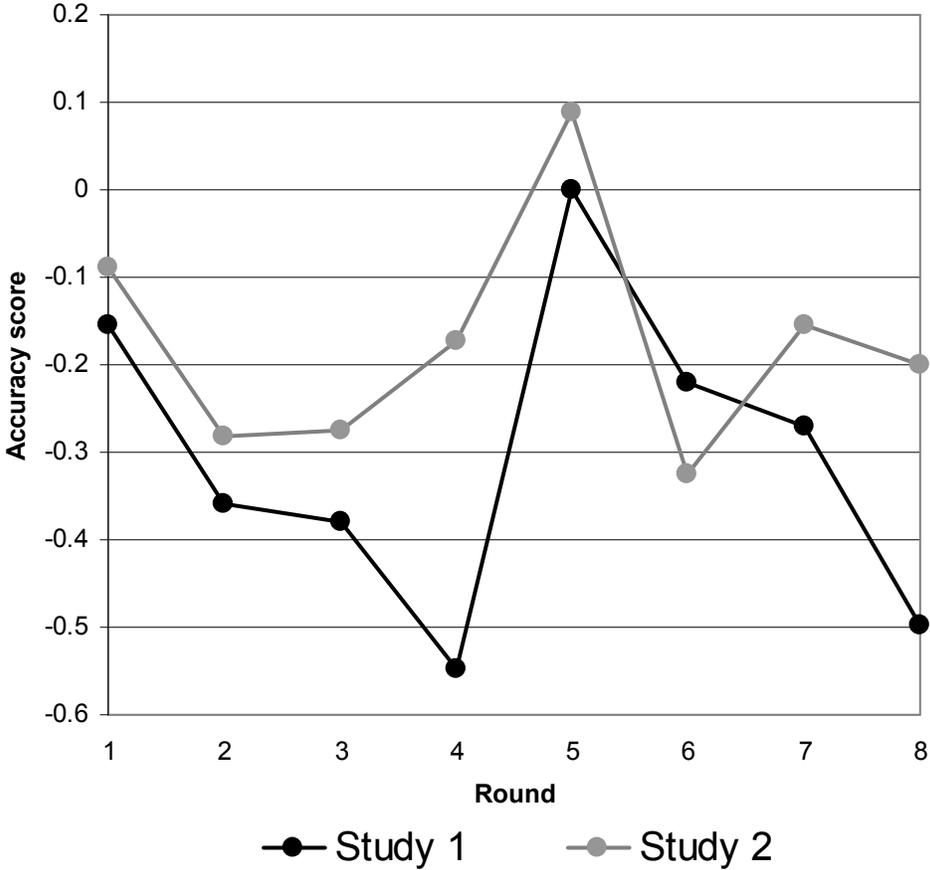


Figure 3. Peak and correct interval confidence across rounds for advisors in Study 3

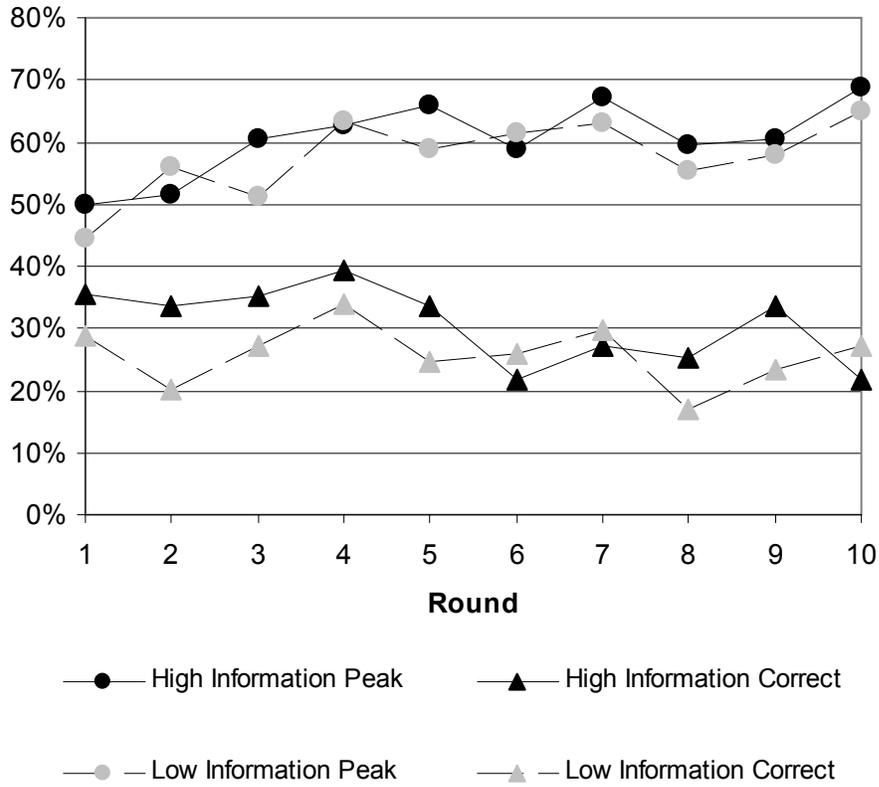


Figure 4. Accuracy across rounds for advisors in Study 3

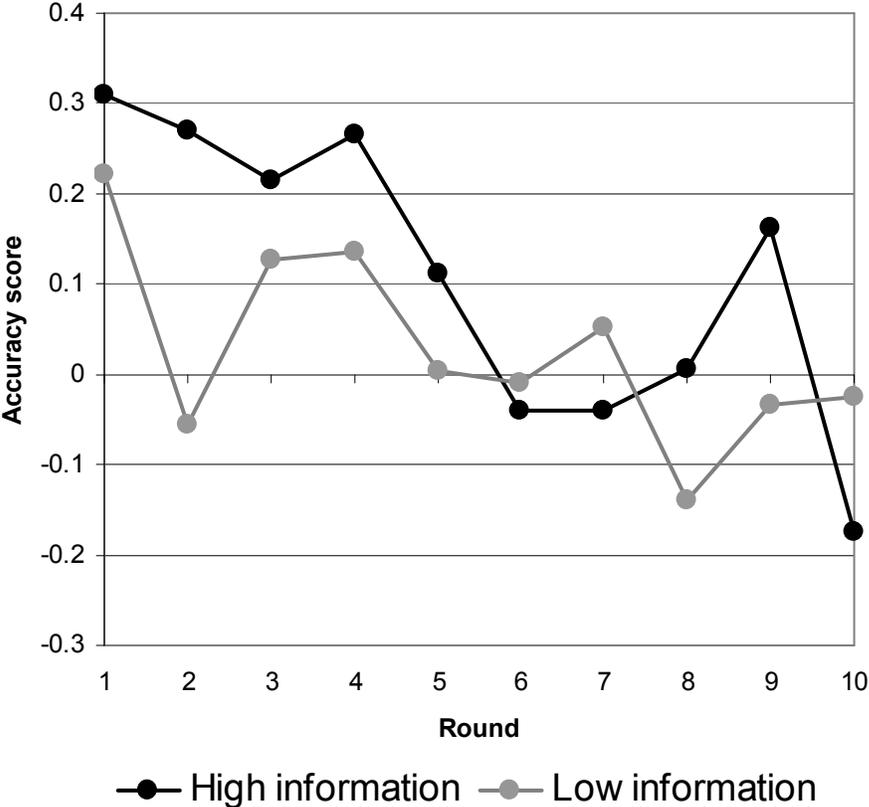


Figure 5. Peak confidence across rounds for advisors in Study 4

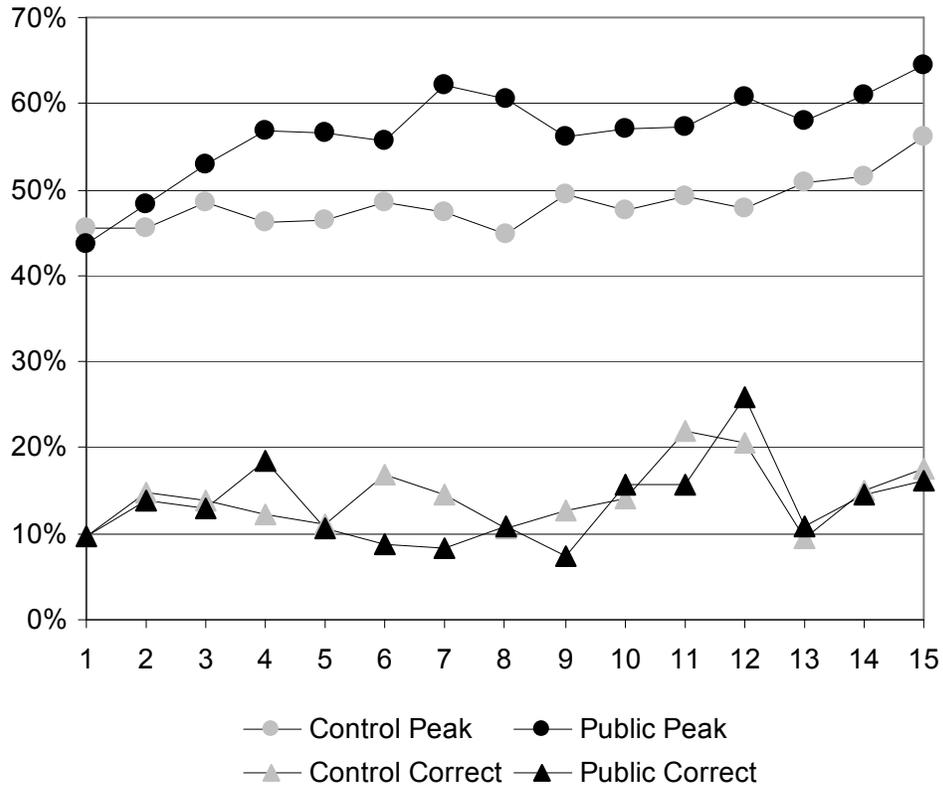


Figure 6. Accuracy across rounds for advisors in Study 4

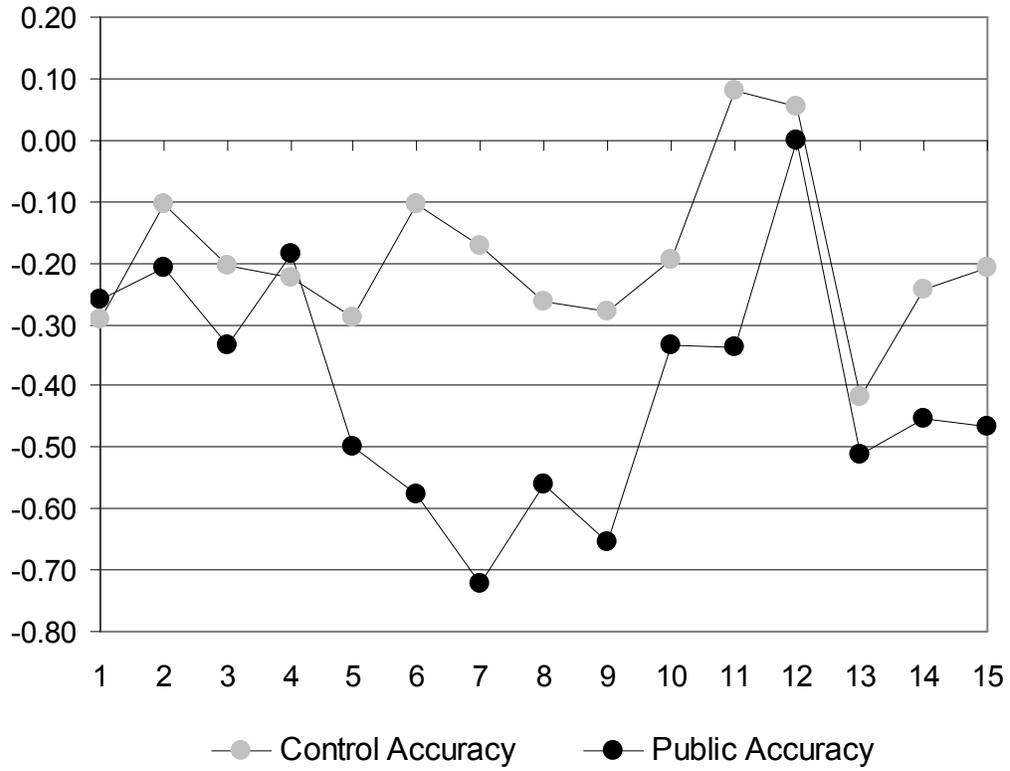
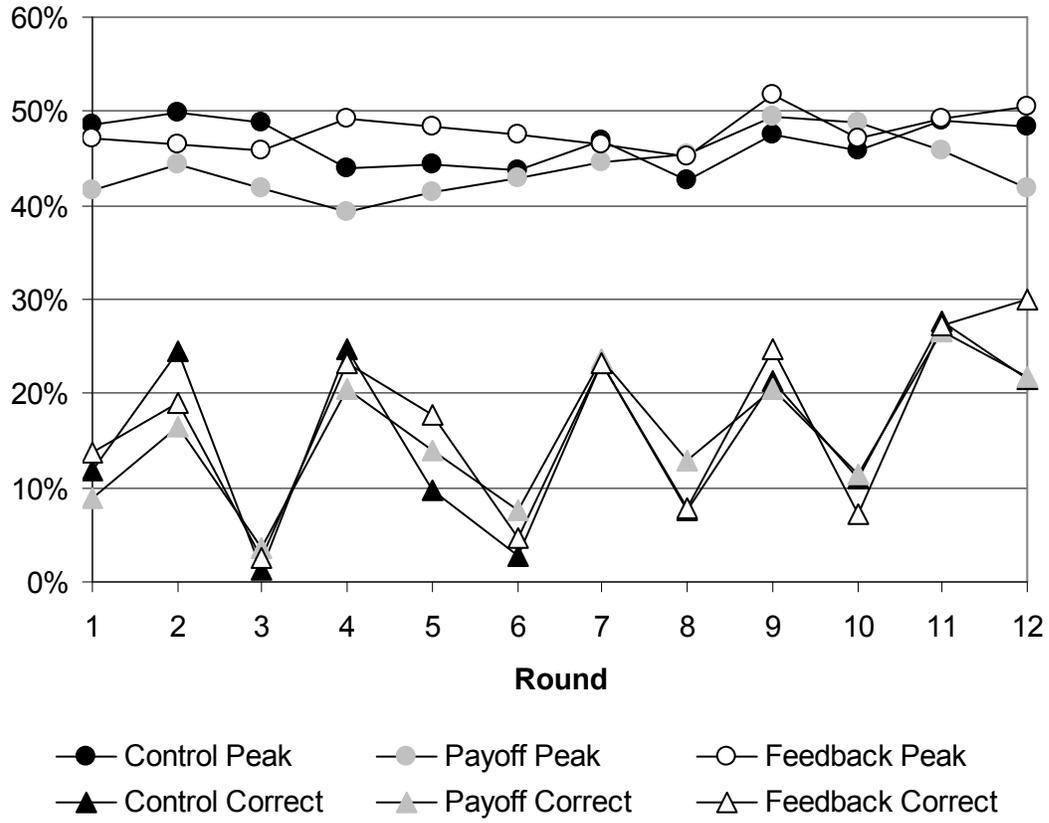


Figure 7. Peak confidence across rounds for advisors in Study 5



Appendices

Appendix A. Advisor decision sheets used in Study 1 and Study 2

Round: _____

Advisor Name _____

For each of the weight ranges below, indicate your confidence level (from 0-100%) that the target person's actual weight falls within that range.

Weight in pounds	Confidence %	Weight in kilograms
120-129 pounds	_____	54.4-58.6 kilograms
130-139 pounds	_____	58.9-63.1 kilograms
140-149 pounds	_____	63.5-67.6 kilograms
150-159 pounds	_____	68.0-72.2 kilograms
160-169 pounds	_____	72.5-76.7 kilograms
170-179 pounds	_____	77.1-81.2 kilograms
180-189 pounds	_____	81.6-85.8 kilograms
190-199 pounds	_____	86.1-90.3 kilograms
200-209 pounds	_____	90.7-94.9 kilograms
210-219 pounds	_____	95.2-99.4 kilograms

Appendix B. Sample advisor decision sheets used in Study 3

Low information

Round: _____

Advisor Name _____

For each of the decades below, indicate your confidence level (from 0-100%) that the given movie won Best Picture during that decade.

Ben-Hur

Decade	Confidence %
1920s	_____
1930s	_____
1940s	_____
1950s	_____
1960s	_____
1970s	_____
1980s	_____
1990s	_____

High information

Round: _____

Advisor Name _____

For each of the decades below, indicate your confidence level (from 0-100%) that the given movie won Best Picture during that decade.

Ben-Hur

Decade	Confidence %
1920s	_____
1930s	_____
1940s	_____
1950s	_____

Appendix C. Additional company information provided to advisors in Study 4

Fortune 100 Revenues

Position	Company Revenues (in billions)	Position	Company Revenues (in billions)	Position	Company Revenues (in billions)
1	\$442.85	41	\$51.65	81	\$29.39
2	\$405.61	41	\$51.65	82	\$29.36
3	\$263.16	43	\$51.49	83	\$29.30
4	\$230.76	44	\$51.32	84	\$29.28
5	\$183.21	45	\$51.26	85	\$28.95
6	\$148.98	46	\$48.30	86	\$28.86
7	\$146.28	47	\$48.23	87	\$28.44
8	\$124.03	48	\$46.98	88	\$28.37
9	\$118.36	49	\$46.77	89	\$28.13
10 Valero Energy	\$118.30	50 Safeway	\$44.10	90 Alcoa	\$28.12
11	\$113.11	51	\$44.05	91	\$28.03
12	\$112.37	52	\$43.25	92	\$27.51
13	\$107.79	53	\$42.87	93	\$25.71
14	\$103.63	54	\$42.73	94	\$25.28
15	\$101.70	55	\$41.09	95	\$25.27
16	\$101.49	56	\$40.02	96	\$24.89
17	\$97.35	57	\$39.54	97	\$24.83
18	\$91.09	58	\$38.06	98	\$24.48
19	\$87.47	59	\$37.95	99	\$24.48
20 Procter and Gamble	\$83.50	60 Walt Disney	\$37.84	100 Rite Aid	\$24.42
21	\$81.19	61	\$37.59		
22	\$76.00	62	\$37.52		
23	\$73.50	63	\$36.56		
24	\$72.48	64	\$35.64		
25	\$71.29	65	\$35.47		
26	\$70.59	66	\$35.45		
27	\$69.82	67	\$34.36		
28	\$64.95	68	\$34.26		
29	\$63.75	69	\$33.94		
30 Morgan Stanley	\$62.26	70 News Corp	\$33.00		
31	\$61.34	71	\$32.48		
32	\$61.25	72	\$32.17		
33	\$61.10	73	\$31.94		
34	\$60.91	74	\$31.88		
35	\$60.42	75	\$31.84		
36	\$59.03	76	\$31.42		
37	\$58.68	77	\$30.95		
38	\$57.51	78	\$30.15		
39	\$55.09	79	\$30.06		
40 Goldman Sachs	\$53.58	80 Abbott Laboratories	\$29.53		

Appendix D. Questions relating to trust used in Study 5

1. Compared to the other advisors, I trusted my chosen Advisor.
2. I would be willing to lend my Advisor almost any amount of money because he or she would pay it back as soon as he or she could.
3. I could expect my Advisor to tell me the truth.
4. I would expect my Advisor to play fair.
5. My Advisor consistently gave his or her best advice.
6. My Advisor was honest in this task.
7. My Advisor was competent to serve as an advisor in this task.
8. My Advisor was loyal to me and had motives that were in my best interest.