MANAGEMENT SCIENCE
Vol. 57, No. 1, January 2011, pp. 93–106
ISSN 0025-1909 | EISSN 1526-5501 | 11 | 5701 | 0093

informs

DOI 10.1287/mnsc.1100.1255

Competing to Be Certain (But Wrong): Market Dynamics and Excessive Confidence in Judgment

Joseph R. Radzevick
Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, jrr2@cmu.edu

Don A. Moore
Haas School of Business, University of California, Berkeley, Berkeley, California 94720, dmoore@haas.berkeley.edu

In this paper, we investigate how market competition contributes to the expression of overconfidence among those competing for influence. We find evidence that market competition exacerbates the tendency to express excessive confidence. This evidence comes from experiments in which advisors attempt to sell their advice. In the first, advisors must compete with other advice sellers. In the second, advisors and their customers are paired. Advisors are overconfident in both studies and it helps advisors sell their advice. However, competition between advisors in the market further exacerbates overconfidence. In a third study, we demonstrate that the market competition drives overconfidence even when advisors vary in quality. We also investigate the strategic expressions and interpretations of confidence by both sides in the exchange.

Key words: overconfidence; advice; competition; markets; judgment
History: Received May 26, 2010; accepted August 9, 2010, by Peter Wakker, decision analysis. Published online in *Articles in Advance* November 19, 2010.

1. Introduction

The presidency of George W. Bush was marked by what some have called steady resolve and others have called stubborn inflexibility. President Bush's 2004 reelection campaign claimed that he and Vice President Richard 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*). Was Bush right? For decision makers to be effective, do they have to inflate their expressed certainty in a chosen course of action? Or is it possible that Bush exaggerated his certainty about "how this world works?"

These issues are not confined to presidential politics. Instead, there exist a wide variety of contexts in which overly confident individuals wield disproportionate influence (see Koehler et al. 2002). It is, for example, the most confident political forecasters who are in the greatest demand by the press (Tetlock 2005). Nevertheless, despite bold claims to the contrary, political forecasters, as a group, do not have a particularly strong track record of accurately predicting how world events will unfold (Tetlock 2005). Although confidence and charisma help aspiring leaders attain influence (Anderson and Kilduff 2009), there are ample grounds to question whether these leaders are actually more effective (Waldman et al. 2001). Although the confidence leaders express often gives people faith that they can affect organizational performance (Meindl et al. 1985), big changes in executive leadership produce only small changes in key organizational outcomes (Pfeffer 1977). Management consultants, too, are better at generating faith in their solutions than in achieving actual results (Mickelthwait and Wooldridge 1996).

In this paper, we focus on the strategic dynamics between those aspiring to be influential and the audiences they try to win over. Thus we distinguish between the producers and receivers of judgment (Yaniv and Foster 1995, 1997). Whether the producers of judgments are managers, consultants, forecasters, or political candidates, these would-be advisors frequently find themselves in competition with others in their attempts to influence potential recipients (employees, customers, voters, etc.). Some prior research has focused on producers (e.g., Russo and Schoemaker 1992). Other research has focused on receivers (e.g., Price and Stone 2004). Some research has utilized both producers and receivers (e.g., Sniezek and Buckley 1995), but not in a way that allows for actual competition between producers, as exists in most organizations and markets. Thus

previous studies have captured some but not all of the key features of these relationships.

Here we examine the exchange between producers and receivers within a dynamic market environment. Specifically, we investigate the possibility that these competitive markets for influence (e.g., advice, leadership, credibility) push contestants toward expressing excessive confidence in the quality of their advice. For producers 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. We begin by examining these benefits and costs.

2. Benefits of Expressing Certainty and Risks of Being Wrong

There is indeed some evidence that people are more persuaded by confident others. 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 (Flynn and Staw 2004) the firm. Clearly, their motivating influence is more due to the inspiration they provide than the careful calibration of their confidence judgments (House 1977, Conger and Kanungo 1987).

Overconfidence has been called "perhaps the most robust finding in the psychology of judgment" (DeBondt and Thaler 1995, p. 389). Overprecision, the excessive certainty that one has the right answer, is the most robust variety of overconfidence (Moore and Healy 2008). The best current theory to account for the ubiquity of overprecision in judgment has to do with its value in communication. Yaniv and Foster (1995, 1997) argue that people express overprecision because it increases the informativeness of what they say. For example, if pressed to estimate the gross domestic product (GDP) of the United States in 2009, individuals can maximize their chances of being right by saying, "Somewhere between zero and infinity." It would be considerably more informative for them to estimate that it is between \$15 and \$16 trillion. The second estimate would be wrong—the actual GDP in 2009 was, according to the Central Intelligence Agency (CIA) *World Factbook* (2010), \$14.4 trillion. But it would 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, would place value in having more precise estimates even if they came 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 \$208 and \$210 trillion, while precise, would be misleading. The key question, then, is whether confidence is positively correlated with accuracy. Often it is (Lindsay et al. 1998, Bornstein and Zickafoose 1999, Sniezek and Van Swol 2001). Naturally, there are some important exceptions, in which confidence and accuracy are uncorrelated, such as in eyewitness testimony (Wells and Olson 2003, Brewer and Wells 2006) and detecting others' deception (DePaulo et al. 1997). An advisor's own confidence that he or she has made the correct prediction may be the only clue available, and it may well be better than nothing. It may therefore be perfectly sensible for people to prefer confident advisors. This reasoning underlies our first hypothesis.

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

Being wrong is, of course, the 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. 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 insisting that Saddam Hussein's Iraq possessed weapons of mass destruction, Bush lost a great deal of credibility when none were found (Ricks 2006). And there is indeed some good research evidence that highlights the risk of claiming confidence and being wrong. Tenney et al. (2007) showed that eyewitnesses who claimed complete confidence regarding a key fact that later turned out to be false lost credibility. Tenney et al. (2008) showed further that witnesses establish their own credibility best by showing good calibration and knowing when they are correct.

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) wrote that there was little evidence that overconfident managers were often "uncovered" and held to account when their decisions turn out badly for their organizations. Indeed, he wrote, "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" (p. 249).

One might well ask why it is that these advisors—who may be 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 of asserting confidence. As Tetlock (2005) observed, the lack of clear and immediate feedback is a strong impediment for such a reckoning. The actual decisions made by advisors and other agents are customarily decoupled from the actual outcomes associated with those judgments. Because the feedback takes too long to arrive, people forget (Pfeffer 1992). During the late 1990s, for instance, 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 repeal contended that the removal of these barriers would strengthen the capabilities of U.S. financial institutions whereas opponents warned that repeal would put the economy at risk. The repeal occurred in 1999 yet it was not until nearly a decade later that the global financial crisis confirmed the fears of those who spoke against repeal (Vekshin 2009). By this time, it was difficult to remember the content of the original debate and many of the players had already left public accountability (such as retired Senator Phil Gramm, sponsor of the repeal).

Such an environment appears ripe for the overconfident to gain more than they lose in the marketplace, leading to our second hypothesis.

Hypothesis 2. *The selection decisions of customers of advice will depend more on the advisors' expressed confidence than on the advisors' previous accuracy.*

3. Further Predictions

In this paper, we focus on the effects of competition between advisors and other market forces on advisors' motivation to express excessive confidence in the accuracy of their judgment. We predict 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 in excess of any gains in accuracy. Our reasoning for these predictions is twofold. First, the certainty advisors express in the market environment depends upon both personal confidence and the confidence expressed by rival producers of judgment. 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 will likely follow (Lichtendahl and Winkler 2007). Second, advisors can use competitors 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 successful advisors. Customers will substantiate these patterns by displaying a preference for confident advice (Yaniv and Foster 1995). This leads us to our third hypothesis.

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

Although the behavior of advisors in such a marketplace can be viewed as a proper response to the information and incentives at hand, the pursuit of confidence is more problematic for those seeking advice and leadership. 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, Hayward et al. 2006, Malmendier and Tate 2005). Because we expect customers' preference for confidence to encourage excessive confidence among advisors, we expect that customers' reliance on such overconfident advice will impair the quality of their own judgments and performance.

Hypothesis 4. *When customers select more confident advisors, their performance will be impaired.*

We present an experimental test of our hypotheses because field data include important limitations with regard to testing the predictions. First, aspiring leaders, politicians, or advisors rarely make statements of confidence that are clear enough that we could test their accuracy. Second, even if we could obtain such unambiguous statements of belief, it is often difficult to obtain data on outcomes that would allow us to estimate the degree of overconfidence in the initial claim. Without these measures, it is impossible to assess the possibility that advisors are justified in making confident statements and customers are right to prefer them. Our experimental setup removes many of these ambiguities and allows for a clearer test of our theory.

4. Study 1

4.1. Design

We constructed a laboratory market in which decision makers in the role of guesser (i.e., receiver) 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 (i.e., producer). 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. We used photographs from a previous study (Moore and Klein 2008) that 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 that listed a series of 10-pound weight ranges between 120 and 219 pounds (we also provided equivalent ranges in kilograms). For each of the ranges, they indicated their confidence level (between 0% and 100%) that the target's actual weight fell within that particular range.

Table 1 Outline of the Market Interface Utilized in Study 1

Example advisor responses				
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	

4.2. Participants

Ninety-eight individuals participated in 13 sessions of the study (35% female; mean age = 23.8, SD = 5.8). They were recruited from a university research pool of community members interested in participating in studies for pay. We advertised the study as involving "estimation tasks" in which participants would earn money based on decisions made during the course of the session. Each session consisted of four advisors and a variable number of guessers (between two and six).

4.3. Procedures

Upon arriving at a session, four participants were randomly assigned to the advisor role while the remaining participants were assigned 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. The order of the eight targets was randomly determined for each experimental session. Advisors then estimated the likelihood that the target's weight fell in each of the 10-pound intervals from 120 to 220 pounds. The narrowness of this subjective probability distribution reflected the advisors' confidence in the accuracy of his or her estimate. 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 confidence. At this point in the experiment, guessers only saw how confident advisors claimed to be but not the weights corresponding to those confidence levels. We used this procedure to capture the imprecise signaling that occurs in many real-world contexts. Prospective agents (advisors, leaders, etc.) 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, such as promising "steady leadership" without a great deal of specificity about what that means.

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 over all eight rounds. This allowed advisors to form reputations with guessers. Table 1 illustrates this procedure.

Guessers viewed the public signal of advisor confidence and used a computer chat program to communicate their choice of advisor to the experimenter, who then sent each guesser the complete confidence distribution of the advisor he or she chose (including the corresponding weights). Guessers then received the target individual's picture and filled out their own confidence estimates that the target's weight fell within each of the 10-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 as the endgame approached and reputations became worth less, we kept them unaware of the duration of the task until announcing the conclusion of the study after the eighth round.

Guessers and advisors faced different financial incentives, intended to mirror those faced by many real producers and receivers of advice. The earnings for guessers increased with accuracy and were calculated each round using the following quadratic scoring rule (Selten 1998): $\$4 * p_c - \$2 * \Sigma p^2$, where p is the probability assigned for a given interval and p_c is the probability assigned to the correct interval. 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 $\$2 * g$, 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.

4.4. Measures

4.4.1. Confidence. We took *peak confidence* as the maximum confidence level individuals assigned to any of the weight intervals for a given target. We 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 a guesser 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.

4.4.2. Range. We computed a simple measure for confidence range as the total number of intervals to which an individual assigned non-zero confidence levels for a given picture. We could have used a more intricate measure to capture the distribution of confidence, such as one based on the variance, but chose the simpler measure, assuming that it is reasonable to expect that guessers could recognize the basic spread of confidence but not necessarily more statistically complex measures. As a precaution, we ran alternative models replacing the range measure with a variance-based measure yet found no substantive changes to our results.

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

4.4.4. Selection. Because sessions consisted of different numbers of guessers, we utilized the percentage of guessers choosing an advisor as the selection variable. We calculated this by dividing the number of guessers that chose a given advisor by the total number of guessers in the market.

4.5. Results

4.5.1. Overprecision. We first examined whether advisors displayed overprecision in their weight estimates. To do so, we contrasted their average 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 correct weight interval only 15% of the time. A paired t-test reveals this difference to be significant, $t(51) = 16.44$, $p < 0.001$. A similar pattern held for guessers, though not quite to the same level of severity. Guessers provided an average peak confidence of 56%, but this peak confidence corresponded to the correct weight interval only 29% of the time, $t(45) = 8.04$, $p < 0.001$.

4.5.2. Advisors Selection. We investigated how advisors' estimates impacted the rate at which they were favored by guessers. To test this, we utilized regression analyses (controlling for session and individual advisor effects) using the selection by guessers as the dependent variable. Peak confidence, accuracy, the number of guessers, and round number served as independent variables in Model 1. We added variables in Model 2 accounting for previous round values of peak confidence, accuracy, and selection. Because guessers who selected an advisor previously should be able to better assess that advisor's accuracy, our third model includes an interaction between previous accuracy and previous selection. The results of these analyses are summarized in Table 2.

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
Previous selection		−0.201**	−0.152*
Previous peak confidence		0.063	0.045
Previous accuracy		−0.016	−0.172†
*Previous accuracy * previous selection*			0.221*
R^2	0.224	0.281	0.302

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

†$p < 0.10$; *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

In the full Model 3, the effect of peak confidence is significant ($\beta = 0.29$, $p < 0.01$). This provides support for Hypothesis 1, that increased confidence would help advisors attract guessers. Accuracy, by contrast, is not a significant predictor of selection. However, we do find some evidence of reputation effects when accounting for the previous round variables. The interaction between previous accuracy and previous selection is significant ($\beta = 0.22$, $p < 0.05$). This suggests that advisors may be penalized for poorer accuracy in prior rounds, but only among those guessers who chose them and were therefore exposed to their overconfident estimates. Consistent with Hypothesis 2, we find that the benefits of claiming certainty are stronger than the reputational costs of inaccuracy in previous rounds ($F(1, 12) = 6.27$, $p < 0.05$).

4.5.3. Changes in Estimates over Time. Hypothesis 3 predicted that advisors in the market environment would 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 1, but decreased to 3.02 intervals in Round 8. This negative trend for range is significant ($F(1, 12) = 21.58$, $p < 0.001$). As shown in Figure 1, peak confidence levels also displayed increasing precision. In Round 1, advisors on average offered peak confidence levels of 52% whereas in Round 8, their average peak confidence increased to 65%. The linear trend in confidence is significant ($F(1, 12) = 13.62$, $p < 0.01$).

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 ($F(1, 12) = 1.05$, $p = 0.33$). Moreover, our measure of accuracy using the quadratic scoring rule reveals a nonsignificant trend toward decreasing accuracy ($F(1, 12) = 0.21$, $p = 0.66$). As a whole, these results show advisors expressing higher degrees of confidence over time but not necessarily providing better estimates.

Figure 1 **Peak and Correct Interval Confidence for Advisors in Study 1 and Study 2**

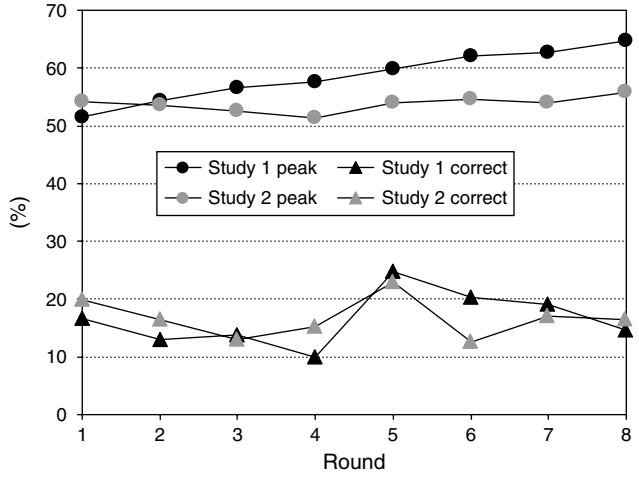


Table 3 **Mediation Test for Guesser Payoffs in Study 1**

	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.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

4.5.4. 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 do not directly impair the performance of the guessers relying on the advisors. However, Hypothesis 4 predicts that such displays of confidence by their chosen advisors will adversely affect guessers. Specifically, we 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.

We ran the necessary mediation analyses to test this prediction, as summarized in Table 3. In Model 1, we find that 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 Models 2 and 3 (Sobel test: $z = -3.14$, $p < 0.01$). Thus Hypothesis 4 is supported.

4.6. Discussion

The results of Study 1 show a clear pattern of advisor behavior in the market. Advisors' statements become more precise over time, as evidenced by higher peak confidence levels and a narrower range of estimates. Further analyses show that greater peak confidence helps attract guessers' business. Interestingly, guessers did not benefit from choosing confident advisors. Instead, guessers who 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 payoffs.

Taken together, we see that evaluating the effectiveness of markets for judgment depends largely upon one's vantage point. For advisors, markets exacerbated their tendencies toward overprecision and inflated confidence. Yet as previous 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 other important motivations. The view for receivers appears less ambiguous. Advisor confidence ideally should provide useful information for them to utilize in their own decision making. Instead, confidence influences receiver decisions though ultimately leads to poorer performance. But we see these wounds as 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."

5. Study 2

Although the market-based study supported our hypotheses, we cannot be certain that the market environment differentiates itself from other exchange structures. It should be the case that these patterns do not generalize to nonmarket scenarios in which competition between advisors is absent. To investigate this possibility, we conducted a second study in which guessers and advisors did not come together through the market but were instead paired for the duration of the task. Instead of choosing among advisors, guessers decided whether to solicit the estimates of a single available advisor.

5.1. 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.

5.2. Participants

Eighty individuals participated in 13 sessions (40% female; mean age = 23.6, SD = 5.7). They were recruited from the same research participant pool as in Study 1 using a similar advertisement. Sessions consisted of a variable number of advisor-guesser pairs (between one and five).

5.3. Procedures

We randomly assigned participants in each session to the role of either guesser or advisor and then created random advisor-guesser pairings. Sessions progressed in the same manner as in 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 based on private information they viewed. 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.

We 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. We included this feature to mirror 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.

5.4. Results

5.4.1. Overprecision. We again examined whether advisors and guessers displayed overprecision in their estimates. Overprecision emerges strongly 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. A paired t-test reveals this difference to be significant, $t(39) = 11.00$, $p < 0.001$. Guessers provided an average peak confidence of 56%, but this peak confidence corresponded to the accurate weight interval only 27% of the time. This difference also is revealed to be significant by a paired t-test, $t(39) = 8.87$, $p < 0.001$.

5.4.2. Advisor Selection. We 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 1 if the guesser chose the advice and took on values of 0 otherwise. We constructed a logit model utilizing selection as the dependent variable with the following predictors (also controlling for individual advisor effects): advisor peak confidence, accuracy, previous peak confidence, previous accuracy, previous selection, and round number. In Model 3, we included the interaction terms between previous accuracy and previous 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

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
Previous selection		1.535	0.650
Previous peak confidence		0.683	−1.421
Previous accuracy		0.739*	0.217
Previous accuracy ∗ *previous selection*			0.513
Pseudo-R^2	0.067	0.093	0.100

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

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.
*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

guessers selecting their advisor's estimates ($p < 0.01$). This suggests that, as in Study 1, guessers rewarded advisors' expressions of confidence. Several of the variables concerning the previous round also emerge as significant. Interestingly, as opposed to Study 1, the impact of previous advisor accuracy did not depend on whether the advice actually had been selected. The accuracy X selection interaction is not significant (Model 3, $p = 0.20$), whereas the main effect for accuracy is significant (Model 2, $p < 0.05$).

Finally, we must note a significant negative effect for round ($p < 0.05$). 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.

5.4.3. Changes in Estimates over Time. Here we examine how advisor estimates evolve over time in the stable exchange relationship and contrast those patterns with the ones shown earlier 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 ($F(1, 39) = 0.05$, $p = 0.83$), which increase from 3.88 intervals in Round 1 to only 3.93 in Round 8. 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 ($F(1, 39) = 0.34$, $p = 0.56$), on average 54% in Round 1 and only 56% in Round 8. Nor does their confidence assigned to the correct weight interval change significantly ($F(1, 39) = 0.09$, $p = 0.76$). Trends in accuracy are similar to Study 1, in that advisors display no changes to their accuracy in later rounds ($F(1, 39) = 0.00$, $p = 0.99$).

5.4.4. Guesser Performance. We also replicated the mediation analyses for guesser performance, as summarized in Table 5. In Model 1, we find that 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 Models 2 and 3 (Sobel test: $z = -2.85$, $p < 0.01$). Interestingly, the impact of advisor confidence on guesser performance is not dependent on guessers actually viewing the advisors' complete estimates, as we find no significant effects if we account for an interaction between selection and advisor peak confidence ($\beta = 0.255$, $p = 0.25$).

5.5. 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 pressure of other advisors in the market played a strong role in the escalating confidence in the first study.

The consequences of advisor confidence for guesser performance remained similar as well. When advisors expressed more confidence in their estimates, their paired guessers also expressed greater certainty, which subsequently resulted in lower payoffs. This occurred regardless of whether 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. Consumers appeared not just to follow their advisors' lead as to the specific value of the correct answer but also took cues as to the appropriate level of confidence to express even when choosing to ignore the specific content of the available advice.

6. 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.) whereas 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 lesser informed counterparts.

Prior research, however, gives us little hope for this ideal set of outcomes. Informed, expert advisors remain plagued by overconfidence (Koehler et al. 2002, Tetlock 2005). McKenzie et al. (2008) demonstrate how expert and novice estimates may 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 greater overprecision in their estimates than less informed advisors.*

Hypothesis 6. *More informed advisors will produce more accurate estimates than their less informed counterparts.*

Hypothesis 7. *More informed advisors will produce narrower estimates than their less informed counterparts.*

In addition to these considerations, we 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, we have done so only indirectly through our quantitative analyses. In Study 3 we supplemented these analyses with qualitative response data provided by both advisors and guessers. We examined this response data within a new categorical framework that addresses the intricate tactics employed by producers and receivers of confidence judgments.

6.1. Design and Procedures

We preserved the basic selection and exchange system of Study 1 but made a few notable modifications. The estimation task lasted 10 rounds instead of eight to allow for more exchanges between advisors and guessers. We also utilized a new estimation task to ensure that our results are not confined to the idiosyncrasies of a single task. In each round, advisors and guessers received a target film and had to estimate in which decade that movie had won the Academy Award for Best Picture. We randomly selected target Best Picture winners from the complete pool of award recipients from 1928 to 1999, thereby excluding winners in the last decade 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 paper decision sheets similar to those used in the previous studies whereas 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 that advisors had when they made their estimates. *Low-information* advisors did not receive any hints as to the correct answers. *High-information* advisors had the possible range of decades narrowed to four. 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 we distributed high across all four color code names (blue, green, red, yellow) for different sessions.

Payoff functions took the form of scaled-down versions of the payoffs from previous studies. The earnings for guessers were calculated each round using the following quadratic scoring rule: $\$2 * p_c - \$1 * \Sigma p^2$, where p is the probability assigned to a given interval and p_c is the probability assigned to the correct interval. 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. We asked advisors to describe how they decided to complete their

confidence estimates and asked guessers to describe how they chose their advisors.

6.2. Participants

One hundred individuals participated in 13 sessions of the study (48% female; mean age = 20.2, SD = 1.13). They were recruited from a university research pool of individuals interested in participating in studies for course credit. We 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 (mean = 3.69, SD = 1.32).

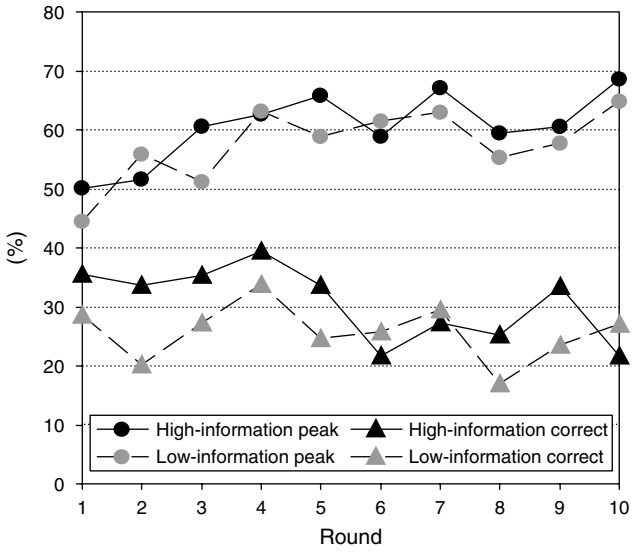
6.3. Quantitative Results

The measures and analyses paralleled those used in the previous studies. We also created the dummy variable *high information*, which took on values of 1 for the more informed advisors and 0 otherwise.

6.3.1. 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. A paired t-test reveals this difference to be significant, $t(51) = 5.05$, $p < 0.001$. High- and low-information advisors displayed similar levels of overprecision, consistent with Hypothesis 5. A similar pattern of overprecision also held for guessers. Guessers provided an average peak confidence of 65%, but this peak confidence corresponded to the correct interval only 44% of the time. This difference also is revealed to be significant by a paired t-test, $t(47) = 7.84$, $p < 0.001$.

6.3.2. Differences Between High- and Low-Information Advisors and Changes in Estimates over Time. We tested whether advisors would grow more confident over time, as in Study 1. This was generally the case, as shown in Figure 2. In Round 1, advisors offered peak confidence levels averaging 47%, whereas in Round 10, their peak confidence increased to 67%. The linear trend in confidence approaches significance ($F(1, 12) = 4.66$, $p = 0.05$), providing some additional support for Hypothesis 3. Interestingly, no significant differences emerge as a consequence of whether advisors were in the high-information condition (all $ps > 0.46$). As in the previous study, the increases in confidence cannot be attributed to increased calibration with the correct target answers. As shown in Figure 2, advisors made no improvements over time for the confidence they provided in the correct decade ($F(1, 12) = 0.15$, $p = 0.71$).

Figure 2 **Peak and Correct Interval Confidence for High- and Low-Information Advisors in Study 3**



In fact, this nonsignificant trend moves downward, meaning advisors became nonsignificantly less accurate over time. Naturally, high-information advisors did provide higher confidence in the correct decade ($F(1, 12) = 8.25$, < 0.05) but showed no difference in this rate over time ($F(1, 12) = 2.13$, $= 0.17$).

Additionally, confidence distributions again constricted over time. Advisors had an average range of 4.15 in Round 1 which decreased to 3.04 in Round 10. The negative trend for range is significant, $F(1, 12) = 6.92$, $p < 0.05$. High-information advisors also offered narrower ranges (mean = 3.03, SD = 0.83) than did the less informed advisors (mean = 3.73, SD = 1.58), $F(1, 12) = 10.54$, $p < 0.01$). These results support Hypothesis 6. 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, $F(1, 12) = 0.29$, $p = 0.60$.

Accuracy, as measured by the quadratic scoring rule, does not improve, and actually decreases across rounds, $F(1, 12) = 5.51$, $p < 0.05$. High-information advisors show some signs of being more accurate, $F(1, 12) = 3.70$, $p < 0.10$, providing some support for Hypothesis 7. This rate did not change significantly for less informed advisors, $F(1, 12) = 1.96$, $p = 0.19$.

6.4. Qualitative Results

After completing the 10 rounds of the estimation task, participants answered free-response questions relating to two important thought processes: advisors' expressions of confidence and guessers' selection of advisors. Using the classification scheme described below, two independent coders recorded when participants' descriptions of their thought processes fit within the defined categories. The inter-rater reliability achieved an acceptable level for both advisor

responses (Cohen's $\kappa = 0.75$) and guesser responses (Cohen's $\kappa = 0.78$). In cases of coder disagreement, the judgments of a third coder served as tiebreaker. For the sake of brevity, we have listed only those categories used in our subsequent analyses. We provide a description of each category and include a representative quotation from the actual participant responses.

6.4.1. Advisor and Guesser Categories. For advisors, we first coded whether they indicated a preference for expressing *high confidence* (appearing more confident in their answers than warranted by their actual beliefs, marked by higher peak confidence levels and narrower distributions: "Tried to estimate large numbers"), *tempered confidence* (providing confidence that is more muted, marked by lower peak confidence levels and wider distributions: "I did not guess to [sic] extreme"), or *knowledge-based responses* (basing estimates on their true beliefs: "I just went with the actual percent I was sure"). We also examined whether advisors held underlying motives based on their *selection* (concern for being chosen by potential customers: "I wanted people to pick me") or their *reputation* (concern for the opinions guessers form about their viability as advisors in the future: "If I was wrong, guessers may not have kept choosing me").

For guessers, we coded whether they described choosing advisors according to the advisors' *high confidence* (advisors that express a high peak confidence level and provide a narrow distribution for their estimates: "The advisor that was most certain about there [sic] estimate by having the highest percentage in one confidence level"), *tempered confidence* (advisors who provide lower peak confidence levels and a wider distribution for their estimates: "Picking the ones with confidence levels that seemed more spread out"), or *track record* (estimates provided by the advisors and/or the outcomes associated with choosing certain advisors in previous rounds: "Red or blue seemed most reliable, green was always way off").

6.4.2. Comparison Between Advisor and Guesser Responses. Seventy-nine percent of advisors indicated that they constructed estimates based on their actual knowledge, at least in part. However, evidence from additional categories shows that this is by no means the only consideration they took into account.

Perhaps the most striking convergence between advisors and guessers is the preferences for high confidence over more tempered confidence. Fifty-eight percent of advisors mentioned high confidence whereas only 29% referred to tempered confidence (17% discussed both). Guessers show an even greater propensity toward valuing high confidence than did advisors, with 83% citing a preference for high confidence and only 15% favoring tempered confidence (with 10% discussing both). This suggests a general tendency people have for giving and receiving more confident advice. It also may indicate 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.

Similarly, both sides exhibit a greater focus on current rounds as opposed to previous or future rounds. Whereas 31% of advisors mentioned concerns about selection, only 10% conveyed concerns for reputation (and no advisors mentioned both). Guessers showed an even greater disparity, as evidenced by the difference between reputation statements and confidence expression statements (either high or tempered). Eighty-eight percent of guessers described basing their decisions on the actual confidence levels whereas only 35% included past performance as part of their thought process (25% mentioned both).

Also of note, advisors indicating selection concerns were more likely to mention high confidence expressions than those who did not (88% versus 44%, Fisher's Exact Test $p < 0.01$). Conversely, we find that those expressing selection concerns were no more likely to reference tempered confidence than those who did not (31% versus 28%, Fisher's Exact Test $p = 0.52$).

6.5. Discussion

Results from the third study generally replicate those of the first study, despite the presence of more informed advisors. The market environment does little to reduce overprecision in advisor estimates. Advisor estimates do not start out well-calibrated. Over time, their confidence increases further, despite the absence of similar gains in accuracy. However, this should not be too surprising considering how often overconfidence is rewarded by guessers. Expressing high confidence clearly offers many benefits and few costs.

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 et al. 2008). However, they also exhibited trends in peak confidence and overprecision similar to their less informed colleagues. Our results 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.

The qualitative results of Study 3 shed further light on the role of confidence 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). Our results support these earlier arguments while expanding and refining the scope of the exchange dynamics. Although most advisors report that they incorporate their actual knowledge into their estimates, they also admit that these distributions often reflect deliberate shaping according to some notion of what guessers may desire. These may depend on how advisors weigh their underlying interests. Expressions of high confidence are linked to immediate selection goals, suggesting 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.

7. General Discussion

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 they are not panaceas, and markets can fail (e.g., Akerlof 1970). Here we present an example in which market competition magnifies a bias present in individual judgments. Human judgment is prone to overprecision (Soll and Klayman 2004), and we observe substantial overprecision in the advice offered across the various studies. This carried over for the estimates offered by other participants in the studies who had clear incentives to make accurate judgments. 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.

7.1. Limitations and Future Directions

Important questions remain about how expressions of confidence are used to gain credibility in competitive market settings. Our experimental markets included two features that we believe drive the escalation of precision in the market environment: competition between advisors and the ability to witness the actions of other competitors. Both are normal features of markets for advice and influence, so it was sensible to include them both in our experimental framework. However, future research would benefit from determining the singular effects of each of these components.

In this series of studies, we chose to focus on the actual quantitative estimates as the sole information by which producers and receivers communicate and evaluate each other. Future research should begin to take into account some of the more complex 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 et al. 2002). For instance, 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, Iraq War supporters contended that their visions of success would have been realized relatively quickly 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 rich offering their advice in the meantime.

In our experimental paradigm, advisors had no avenues of communication through which they could articulate any of these arguments. Although their persistent claims of greater confidence, despite their unimpressive prior accuracy, have similar implications on their own, it would be worth studying these persuasive communications more deeply.

We also should not presume that unfettered confidence always remains unpunished while more restrained confidence goes unrewarded. Some would-be financial gurus receive a significant blow to their reputations when their stock picks prove more miss than hit (Alpert 2009) or their sweeping market forecasts fail to materialize (e.g., Glassman and Hassett 1999). Conversely, shrewd forecasters, often vilified when making their less bombastic claims, may ultimately gain credibility when their predictions prove out (Mihm 2008, Sah et al. 2010). One potential catalyst for these alternative outcomes is the availability of enhanced information for potential customers. For example, a vast collection of online resources (e.g., story archives, databases, blogs, etc.) has developed to provide useful data to discerning consumers of politics, economics, sports, and numerous other domains. Investigating the impact of these information disparities and other potential moderators would provide important additional insights in to the types of judgment markets we describe here.

7.2. Conclusion

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 the 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 we have shown, 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, it should be no surprise that overconfidence remains potent and pervasive.

If any further evidence is needed, we need only turn our attention back to presidential politics. Many American voters reported that the hubris of 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 opponent, Barack Obama. Political observers noted that, unlike McCain, Obama displayed an unflappable self-assurance throughout the campaign (Kantor 2008). And we all know who won the Presidency in 2008.

Acknowledgments

The authors are grateful for the helpful comments and suggestions made by Francesca Gino, Roberto Weber, Mark Fichman, the editors, two anonymous reviewers, and participants at the 2009 conferences for the Academy of Management in Anaheim, the Society for Judgment and Decision Making in Boston, and the Association for Psychological Science in Las Vegas. The authors thank the Center for Behavioral Decision Research, the Carnegie Bosch Institute, and the Pittsburgh Supercomputer Center for the use of their resources and facilities. The authors also thank Carolina Velez, Jennifer Hwang, and Mingwei Hsu for their help with data collection.

References

Akerlof, G. 1970. The market for lemons: Qualitative uncertainty and the market mechanism. *Quart. J. Econom.* **84**(3) 488–500.

Alpert, B. 2009. Cramer's star outshines his stock picks. *Barron's* (February 9), http://online.barrons.com/article/SB123397107399659271.html#articleTabs_panel_article%3D1.

Anderson, C., G. J. Kilduff. 2009. Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *J. Personality Soc. Psych.* **96**(2) 491–503.

Baron, R. M., D. A. Kenny. 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Personality Soc. Psych.* **51**(6) 1173–1182.

Bornstein, B. H., D. J. Zickafoose. 1999. I know it, I know I saw it: The stability of the confidence-accuracy relationship across domains. *J. Experiment. Psych.: Appl.* **5**(1) 76–88.

Bottom, W. P., K. Gibson, S. E. Daniels, J. K. Murnighan. 2002. When talk is not cheap: Substantative penance and expressions of intent in rebuilding cooperation. *Organ. Sci.* **13**(5) 497–513.

Brewer, N., G. L. Wells. 2006. The confidence-accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *J. Experiment. Psych.: Appl.* **12**(1) 11–30.

Central Intelligence Agency (CIA). 2010. *World Factbook*. CIA, Washington, DC. https://www.cia.gov/library/publications/the-world-factbook/.

Cloud, D. S., E. Schmitt. 2006. More retired generals call for Rumsfeld's resignation. *The New York Times* (April 14), http://www.nytimes.com/2006/04/14/washington/14military.html.

Conger, J. A., R. N. Kanungo. 1987. Toward a behavioral theory of charismatic leadership in organizational settings. *Acad. Management Rev.* **12**(4) 637–647.

DeBondt, W. F. M., R. H. Thaler. 1995. Financial decision-making in markets and firms: A behavioral perspective. R. A. Jarrow, V. Maksimovic, W. T. Ziemba, eds. *Finance: Handbooks in Operations Research and Management Science*. Elsevier, North Holland, Amsterdam, 385–410.

DePaulo, B. M., K. Charlton, H. Cooper, J. J. Lindsay, L. Muhlenbruck. 1997. The accuracy-confidence correlation in the detection of deception. *Personality Soc. Psych. Rev.* **1**(4) 346–357.

Flynn, F. J., B. M. Staw. 2004. Lend me your wallets: The effect of charismatic leadership on external support for an organization. *Strategic Management J.* **25**(4) 309–330.

Frank, R. H., P. J. Cook. 1995. *The Winner-Take-All Society: Why the Few At the Top Get So Much More Than the Rest of Us*. Penguin, New York.

Glassman, J. K., K. A. Hassett. 1999. *Dow 36,000: The New Strategy For Profiting From the Coming Rise in the Stock Market*. Times Business, New York.

Gode, D. K., S. Sunder. 1993. Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. *J. Political Econom.* **101**(1) 119–137.

Hayward, M. L. A., D. C. Hambrick. 1997. Explaining the premiums paid for large acquisitions: Evidence of CEO hubris. *Admin. Sci. Quart.* **42**(1) 103–127.

Hayward, M. L. A., D. A. Shepherd, D. W. Griffin. 2006. A hubris theory of entrepreneurship. *Management Sci.* **52**(2) 160–172.

House, R. J. 1977. A 1976 theory of charismatic leadership. J. G. Hunt, L. Larson, eds. *Leadership: The Cutting Edge*. Southern Illinois University Press, Carbondale, 189–204.

Kantor, J. 2008. For a new political age, a self-made man. *New York Times* (August 28), http://www.nytimes.com/2008/08/28/world/americas/28iht-28obama.15695182.html.

Kemp, K. 2007. *Flipping Confidential: The Secrets of Renovating Property for Profit in Any Market*. John Wiley & Sons, New York.

Khurana, R. 2004. *Searching for a Corporate Savior: The Irrational Quest for Charismatic CEOs*. Princeton University Press, Princeton, NJ.

Koehler, D. J., L. Brenner, D. Griffin. 2002. The calibration of expert judgment: Heuristics and biases beyond the laboratory. T. Gilovich, D. Griffin, D. Kahneman, eds. *Heuristics and Biases*. Cambridge University Press, Cambridge, UK, 686–715.

Lichtendahl, K. C., R. L. Winkler. 2007. Probability elicitation, scoring rules, and competition among forecasters. *Management Sci.* **53**(11) 1745–1755.

Lindsay, D. S., J. D. Read, K. Sharma. 1998. Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psych. Sci.* **9**(3) 215–218.

Malmendier, U., G. Tate. 2005. CEO overconfidence and corporate investment. *J. Finance* **60**(6) 2661–2700.

McKenzie, C. R. M., M. J. Liersch, I. Yaniv. 2008. Overconfidence in interval estimates: What does expertise buy you? *Organ. Behav. Human Decision Processes* **107**(2) 179–191.

Meindl, J. R., S. B. Ehrlich, J. M. Dukerich. 1985. The romance of leadership. *Admin. Sci. Quart.* **30**(1) 78–102.

Micklethwait, J., A. Woolridge. 1996. *The Witch Doctors: Making Sense of the Management Gurus*. Times Books, New York.

Mihm, S. 2008. Dr. Doom. *New York Times* (August 15), http://www.nytimes.com/2008/08/17/magazine/17pessimist-t.html?scp=1&sq=&st=nyt.

Moore, D. A., P. J. Healy. 2008. The trouble with overconfidence. *Psych. Rev.* **115**(2) 502–517.

Moore, D. A., W. M. P. Klein. 2008. The use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organ. Behav. Human Decision Processes* **107**(1) 60–74.

Pfeffer, J. 1977. The ambiguity of leadership. *Acad. Management Rev.* **2**(1) 104–112.

Pfeffer, J. 1992. *Managing with Power: Politics and Influence in Organizations*. Harvard Business School Press, Boston.

Plott, C. 1995. Industrial organization: A survey of laboratory research. J. H. Kagel, A. E. Roth, eds. *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.

Price, P. C., E. R. Stone. 2004. Intuitive evaluation of likelihood judgment producers: Evidence for a confidence heuristic. *J. Behav. Decision Making* **17**(1) 39–57.

Ricks, T. E. 2006. *Fiasco: The American Military Adventure in Iraq*. Penguin, New York.

Roberts, R. R., J. Kraynak. 2006. *Flipping Houses for Dummies*. John Wiley & Sons, New York.

Russo, J. E., P. J. H. Schoemaker. 1992. Managing overconfidence. *Sloan Management Rev.* **33**(2) 7–17.

Sah, S., D. A. Moore, R. J. MacCoun. 2010. Expertise, confidence, and persuasion. Unpublished manuscript, Carnegie Mellon University, Pittsburgh.

Selten, R. 1998. Axiomatic characterization of the quadratic scoring rule. *Experiment. Econom.* **1**(1) 43–61.

Smith, A. 1776. *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Strahan and Cadell, London.

Sniezek, J. A., T. Buckley. 1995. Cueing and cognitive conflict in judge-advisor decision making. *Organ. Behav. Human Decision Processes* **62**(2) 159–174.

Sniezek, J. A., L. M. Van Swol. 2001. Trust, confidence, and expertise in a judge-advisor system. *Organ. Behav. Human Decision Processes* **84**(2) 288–307.

Soll, J. B., J. Klayman. 2004. Overconfidence in interval estimates. *J. Experiment. Psych.: Learn., Memory, Cognition* **30**(2) 299–314.

South Florida Sun-Sentinel. 2004. Read the complete first presidential debate. (October 1), http://www.dailypress.com/topic/sfl-930debatetranscript,0,654507,full.story.

Tenney, E. R., B. A. Spellman, R. MacCoun. 2008. The benefits of knowing what you know (and what you don't): How calibration affects credibility. *J. Experiment. Soc. Psych.* **44**(5) 1368–1375.

Tenney, E. R., R. MacCoun, B. A. Spellman, R. Hastie. 2007. Calibration trumps confidence as a basis for witness credibility. *Psych. Sci.* **18**(1) 46–50.

Tetlock, P. E. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, Princeton, NJ.

Vekshin, A. 2009. U.S. senators propose reinstating Glass-Steagall act. *Bloomberg* (December 16), http://www.bloomberg.com/apps/news?pid=20601103&sid=aQfRyxBZs5uc.

Waldman, D. A., G. G. Ramirez, R. J. House, P. Puranam. 2001. Does leadership matter? CEO leadership attributes and profitability under conditions of perceived environmental uncertainty. *Acad. Management J.* **44**(1) 134–143.

Weisman, S. R., R. F. Worth. 2006. Iraqi violence strains U.S. political and military strategy and imperils pullout plans. *New York Times* (February 24), http://www.nytimes.com/2006/02/24/politics/24diplo.html.

Wells, G. L., E. A. Olson. 2003. Eyewitness testimony. *Annual Rev. Psych.* **54**(1) 277–295.

Westley, F., H. Mintzberg. 1989. Visionary leadership and strategic management. *Strategic Management J.* **10**(S1) 17–32.

Yaniv, I., D. P. Foster. 1995. Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *J. Experiment. Psych.: General* **124**(4) 424–432.

Yaniv, I., D. P. Foster. 1997. Precision and accuracy of judgmental estimation. *J. Behav. Decision Making* **10**(1) 21–32.