Competing To Be Certain (But Wrong): Social Pressure and Overprecision in Judgment

Joseph R. Radzevick

Don A. Moore

Carnegie Mellon University

Corresponding author:
Joseph Radzevick
CMU/Tepper
5000 Forbes Avenue
Pittsburgh, PA 15213
jrr2@andrew.cmu.edu

Draft of January 14, 2009

We wish to thank the staff and facilities of the Center for Behavioral Decision Research and the Carnegie Bosch Institute at Carnegie Mellon University. We also would like to thank the Pittsburgh Supercomputer Center and to Carolina Velez for her help with data collection. Please address correspondence to jrr2@andrew.cmu.edu or don.moore@alumni.carleton.edu.

Abstract

Overprecision in judgment is both the most robust and the least understood form of overconfidence. Overly precise judgments claim more certainty than is objectively warranted. In this paper, we investigate whether the competitive social pressure of a market contributes to overprecision among those competing for influence. We find evidence that markets do indeed exacerbate overprecision. This evidence comes from two experiments in which advisors attempt to sell their advice. In the first experiment, advisors must compete with other advice sellers. In the second, advisors and decision makers are paired. Overprecision exists in both studies, and it helps advisors' sell their advice. However, the market also exacerbates overprecision. We discuss the strategic implications of these results.

Competing To Be Certain (But Wrong): Social Pressure and Overprecision in Judgment

The presidency of George W. Bush has been marked by what some have called steady resolve and some have called stubborn inflexibility. Bush's 2004 re-election campaign claimed that he and 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?"

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. This paper will focus on the third type. Overprecision is both the most robust and the least studied form of overconfidence. Big questions remain about why it occurs and what contributes to it. In this paper, we examine the possibility that competitive markets for influence (e.g., leadership, advice, or credibility) push contestants toward expressing overprecision. 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. What do we know about these benefits and costs?

The Benefits of Expressing Certainty

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 & 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 & Mintzberg, 1989) and outside the firm (Flynn & Staw, 2004). Clearly, their motivating influence is more due to the inspiration they provide than the careful calibration of their confidence judgments (Conger & Kanungo, 1987; House, 1977).

Overconfidence has been called "perhaps the most robust finding in the psychology of judgment" (DeBondt & Thaler, 1995), and overprecision is the most robust variety of overconfidence (Moore & 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 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, I can maximize my chances of being right by saying, "Somewhere between 0 and infinity." It would be considerably more informative for me 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 those who are the 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 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 & Zickafoose, 1999; Lindsay, Read, & Sharma, 1998; Sniezek & Van Swol, 2001). Naturally, there are some important exceptions, in which confidence and accuracy are uncorrelated, such as in eyewitness testimony (Brewer & Wells, 2006; Wells & Olson, 2003) and detecting others' deception (DePaulo, Charlton, Cooper, Lindsay, & 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 the only clue available, 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, 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. 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 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 good research 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 & Hambrick, 1997; Malmendier & 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."

In this paper, we focus on the effect of competition between advisors and the effect of market forces on advisors' motivation to express excessive precision. We speculate that the importance Bush accorded to certainty from the U.S. President will 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. We also hypothesize that markets in which advisors compete with one another will lead to increases in the overprecision of their advice over time. We expect this to be marked by increasing confidence without corresponding gains in accuracy. This increase in overprecision will be driven by customers' preference for confident advice and the consequent competition between advisors. Where appropriate, we test these predictions at both the individual level as well as the market session level.

Study 1

Design

We constructed a laboratory market in which decision makers in the role of "guesser" (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 that 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 & 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 ten 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.

Participants

Ninety-eight individuals participated in thirteen sessions of the study (35% female; Mean age = 23.8, *S.D.* = 5.8). They were recruited from a Carnegie Mellon University research pool of individuals 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).

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, 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. 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 most aggregate 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 unfolds.

Guessers viewed the public information of advisor confidence before they selected an advisor. At this time, the guessers did not see the corresponding weights to these confidence levels. Guessers used a computer chat program to communicate their choices to the experimenter, who sent back the complete confidence distributions of the chosen advisors (including 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.

Guessers and advisors faced different financial incentives. The earnings for guessers were calculated each round using the following quadratic scoring rule: $\$4*p_c - \$2*\Sigma\, p^2$, 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 (Selten, 1998). 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 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 varying numbers of guessers.

Measures

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 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%.

Range. We computed a simple measure for confidence range as the number of intervals individuals used to construct their confidence distributions.

Accuracy. We formed a measure of accuracy by taking the absolute difference between a participant's weighted estimate for the target and the target's true weight. To derive the weighted estimate, we created a summation of the confidence estimates multiplied by the midpoint of the corresponding weight interval. So if someone estimated likelihoods of 30%, 60%, and 10% for the target's weight falling in the respective intervals of 140-149 pounds, 150-159 pounds, and 160-169 pounds, the weighted average estimate is 152.5 pounds. Smaller difference scores correspond to greater accuracy in the estimates.

Selection. Since 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.

Retention. This occurred when guessers chose the same advisor in a given round as they did in the previous round. We created a dummy variable set equal to one if the guesser retained their previous advisor and set to zero otherwise.

Results

Overprecision

We first examined whether advisors and guessers displayed overprecision in their weight estimates. To do so, we contrasted 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, $t\ (51) = 16.44$, $p < .001$. 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 accurate weight interval only 29% of the time. This difference also is revealed to be significant by a paired t-test, $t\ (45) = 8.04$, $p < .001$.

Confidence

Advisors. We predicted that advisors in the market environment would grow more confident over time. This indeed was the case, as shown in Figure 1. 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 at both the session level ($F\ (1, 12) = 13.62$, $p < .01$) and the individual advisor level ($F\ (1, 51) = 31.18$, $p < .001$). 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 at either the session level ($F\ (1, 12) = 1.05$, $p = .33$) or at the individual advisor level ($F\ (1, 51) = 1.75$, $p = .19$).

While the market environment failed to attenuate advisor overconfidence, there is also some evidence that it actually exacerbated it. We can see this in the differences between advisors' peak confidence and the confidence they assigned to the correct weight interval. The difference grows from 34.81% in Round One up to 50.13% in Round Eight. This upward trend is not significant at the session level (F (1, 12) = 3.10, p = .11) but is significant at the individual advisor level (F (1, 51) = 5.19, p < .05).

Guessers. We also examined trends in the peak confidence displayed by guessers. As shown in Figure 1, guesser confidence exhibits no significant trend across rounds at either the session (F (1, 12) < 0.01, p = .97) or individual guesser level (F (1, 45) = 0.05, p = .83). Guesser confidence for the correct weight interval likewise shows no significant changes over time (F (1, 12) = 0.57, p = .47; F (1, 45) = 2.45, p = .12). Unlike advisors, guessers display consistent magnitudes of overconfidence, starting at 34.46% in Round One and remaining relatively steady at 38.59% in Round Eight. Neither trend is significant (F (1, 12) = 0.51, p = .49; F (1, 45) = 1.49, p = .23).

Chosen advisors. In addition to the activities of advisors as a whole, it is perhaps more important to examine what occurs for the specific instances in which advisors are selected by guessers. This approach has the additional feature of weighting advisors more heavily when they are chosen by a larger number of guessers. These results also appear in Figure 1. There is a significant trend in the peak confidence of the chosen advisors at both the session (F (1, 12) = 9.58, p < .01) and individual level (F (1, 45) = 27.04, p < .001). This suggests that over time, guessers display a preference for more confident advisors. However, we see that this is not completely unfounded since these chosen advisors also display increases in their confidence for the correct weight intervals. This trend is only marginally significant at the session level (F (1, 12) = 4.27, p < .10) but is significant at the individual level (F (1, 45) = 11.12, p < .01).

Range

Advisors. We predicted that advisor confidence distributions would constrict over time in the market. This is indeed what occurred. Advisors had an average range of 4.08 in Round One which decreased to 3.02 in Round Eight. The negative trend for range is significant (F (1, 51) = 35.14, $p < .001$).

Guessers. Conversely, guessers show no such changes to the span of their distributions over the course of the eight rounds (F (1, 45) = 0.03, $p = .87$). Guessers used an average of 3.35 intervals in Round One and 3.17 intervals in Round Eight.

Chosen advisors. While guessers did not decrease the ranges of their own estimates over time in the market, the distributions of their selected advisors more closely matched the larger pool of available advisors. Guessers chose advisors with narrower distributions as sessions progressed, from 3.83 in Round One to only 3.11 in Round Eight. This negative trend is significant (F (1, 45) = 15.66, $p < .001$).

Accuracy

The difference between the weighted average estimate and the correct weight of the target provide us with a measure of accuracy for advisors, guessers, and chosen advisors. The results are summarized in Figure 2.

Advisors. Accuracy shows little improvement as evidenced by decreases in the deviation between estimated and actual weight over time. The trend is in the direction of decreasing deviation (i.e. increasing accuracy) but it is not significant (F (1, 12) = 2.22, $p = .16$) at the session level nor is it significant at the individual advisor level (F (1, 51) = 2.12, $p = .15$).

Guessers. Conversely, this measure of accuracy shows a non-significant increasing deviation signaling worsening accuracy at both the session (F (1, 12) = 1.42, $p = .26$) and individual guesser level (F (1, 45) = 1.30, $p = .26$).

Chosen advisors. Guessers did not improve over time in selecting advisors with more accurate estimates at the session (F (1, 12) = 2.90, p = .11) but did at the individual guesser level (F (1, 45) = 9.80, p < .01).

Advisor Selection

We further investigated how these factors impacted the rate at which advisors received favor from guessers. To test this, we utilized regression analysis (controlling for session effects) using the S*election* variable as the dependent variable. We included the independent variables of peak confidence, accuracy, correct interval confidence, range, and round in Model 1. We added additional variables in Model 2 accounting for previous round values of peak confidence, accuracy, correct interval confidence, and selection. Since guessers that selected an advisor beforehand will be able to assess that advisor's accuracy more acutely, we constructed Model 3 that includes an interaction between previous accuracy and previous selection as well as Model 4 which includes an interaction between previous correct interval confidence and previous selection. The results of these analyses are summarized in Table 2.

In Model 3, the effect of peak confidence is significant, β = 0.21, p < .01. This supports the prediction that increased confidence would attract guessers. Neither of the other current round variables are significant predictors. However, we find some evidence of reputation effects in the previous round variables. The interaction between previous accuracy and previous selection is significant, β = -0.16, 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. Similar results hold for the alternative specification in Model 4. The interaction between previous correct interval confidence and previous selection is significant, β = 0.25, p < .01. Peak confidence remains significant as well (β = 0.20, p < .01).

Advisor Retention

We constructed a probit model with *Retention* as the dependent variable. Independent predictors included previous round values of peak confidence, accuracy, and correct interval confidence for both guessers and advisors. We also controlled for round. The results of this analysis are shown in Table 3.

None of the peak confidence or accuracy variables significantly affected the likelihood of guessers retaining their advisors from the previous round. However, both correct interval confidence variables emerge as significant. Guessers are more likely to retain their advisors when they themselves indicated higher confidence for the interval containing the target's true weight ($p < .05$). This suggests that guessers attribute at least some of their success to the advisors or at the very least feel reluctance to depart from a previously effective strategy. Moreover, guessers are more likely to retain advisors when those advisors previously had indicated greater confidence for the correct interval ($p < .01$). Combined with the non-signficant effect of the more complex measure of accuracy, this suggests that the formation of advisor reputation is based on a fairly rudimentary impression.

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. This is likely driven by competition among advisors for attracting guessers. The advisors chosen by guessers show a similar concentration of confidence over time in the market. Further analyses show that greater peak confidence plays a significant role in the proportion of guessers obtained by advisors.

However, unqualified confidence is not the lone factor. Both simple (using the confidence level for the correct interval) and more complex (using the weighted average estimate) measures of accuracy influence advisors' proportion of guessers selecting them. We also found that both of these increase over time for the specific advisors chosen by guessers. Only the simple measure, confidence for the correct interval, affected whether or not guessers retained their previous round advisor.

Interestingly, guessers did not capitalize on the improvements of their chosen advisors. Their correct interval confidence and weighted average accuracy showed no improvements over time. In line with this, their peak confidence and range of estimates did not change in later rounds.

Though the market based study supported our predictions, we 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 in which competition between advisors is absent. To investigate this, we 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 their single available advisor.

Study 2

To investigate whether the results of Study 1 are limited to the environment of the market, we ran a follow-up study in which advisors did not have to compete with others to provide their services. It would support our claims if those previous results dissipated in this more restricted exchange relationship.

Design

The study replicated the essential features of Study 1. Over the course of eight rounds, guessers completed the weight estimation task and had the opportunity to receive aid from an advisor. 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 in thirteen sessions of the study (40% female; Mean age = 23.6, *S.D.* = 5.7). They were recruited from the same research participant pool as Study 1 using a similar advertisement. Sessions consisted of a variable number of advisors and guessers (between one and five).

Procedures

Participants in each session were randomly assigned to the role of either guesser or advisor and then paired. 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 estimate of their paired advisors based on private information they viewed. This private information 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.

Results

Overprecision

We again examined whether advisors and guessers 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, $t(39) = 11.00$, $p < .001$. 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, $t(39) = 8.87$, $p < .001$.

Confidence

The trends in confidence across each round for both advisors and guessers are summarized in Figure 3. These trends show some striking differences with 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 ($F(1, 39) = 0.09$, $p = .76$). Conversely, guessers display significant decreases over time in their peak confidence (61% in Round One and 52% in Round Eight; $F(1, 39) = 8.35$, $p < .01$) and their confidence assigned to the correct weight interval ($F(1, 39) = 5.07$, $p < .05$).

Range

Differences between the two studies emerge for the span of the confidence estimates as well. 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. It is instead guesser ranges that change by increasing significantly and indicating less confidence over the eight rounds (moving from 3.13 to 3.45), F (1, 39) = 6.04, $p < .05$).

Accuracy

As summarized in Figure 4, trends in accuracy diverge less from Study 1. Advisors display a non-significant increase to their accuracy in later rounds, F (1, 39) = 0.84, p = .36. Meanwhile guessers show a significant decrease in accuracy over time, F (1, 39) = 4.48, $p < .05$. This means that guessers in this study performed slightly worse over time compared to the market study.

Advisor Selection

We also examined what factors affected the likelihood that guessers would choose to receive the estimates of their advisors. The selection variable for this study took on values of one if the guesser chose the advice and took on values of zero otherwise. We constructed a probit model utilizing selection as the dependent variable with the following predictors: advisor peak confidence, advisor correct interval confidence, advisor range, advisor previous peak confidence, advisor previous correct interval confidence, advisor previous accuracy, advisor previous selection, guesser previous peak confidence, guesser previous correct interval confidence, guesser previous accuracy, and round. 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 confidence expressed by advisors. Several of the variables concerning the previous round also emerge as significant. Whereas higher confidence in the current round enhances advisors' prospects, higher advisor confidence in the previous round decreases the likelihood of selection ($p < .01$). This may reflect guessers' identification of prior demonstrations of overconfidence from their advisors. Similarly, guessers appear attuned to the confidence their advisors estimate for the correct weight interval. This variable significantly increases the likelihood of selection ($p < .001$). Prior round selection has a marginally significant impact on selection in the current round ($p < .10$). This may reflect a tendency for guessers to make a relatively long term commitment to receive advice.

Guessers' own success contributes to their selection process, as evidenced by a significant decrease in selection likelihood associated with guesser confidence for the correct interval in previous rounds ($p < .001$). Interestingly, neither this correct interval confidence nor the previously discussed effect for advisors expands to the more complex assessments of accuracy (p's $> .65$). As in the market study, guessers rely on more superficial metrics of accuracy. Finally, we must note a strong negative effect for round ($p < .001$). 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.

Discussion

The results of Study 2 suggest that the findings from the market study do not necessarily generalize to a less competitive advice exchange context. Unlike in 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. Likewise, advisors who had indicated a higher confidence level for the correct interval in the previous round had a greater chance of being chosen by their guessers. 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.

We must also note that guessers' confidence moderated itself in Study 2's static exchange context. Their peak confidence fell and ranges increased over time, showing decreasing certainty as the rounds progressed. Likewise, confidence for the correct interval and weighted average accuracy decreased over time. In the market environment, these measures did not improve but at least remained unchanged. So while we earlier noted the failings of the market to improve guesser performance over time, the results of the second study suggest that the failings of the market are not as severe as other potential arrangements.

GENERAL DISCUSSION

As economists all the way 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 & Sunder, 1993; Plott, 1995). And markets can certainly concentrate the rewards to popular products, services, or personalities (Frank & Cook, 1995). But markets are not panaceas. There are circumstances under which markets fail entirely (e.g., Akerlof, 1970). We present an example of an instance 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 & Klayman, 2004) and we observed substantial overprecision in the advice offered in both Study 1 and Study 2. This carried over for the estimates offered by other participants in both studies who had clear incentives to make accurate judgments. However, this bias in judgment was magnified by the presence of the competitive market for advice we constructed in Study 1. We observe that consumers tend to pick the advice from those who express more confidence that they have the right answer.

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. Why not? There are three possibilities, each of which deserves further research. First, it is possible that people forget. This was Pfeffer's (1992) claim. In many circumstances, the feedback just takes too long to arrive. When, in 2008, investors saw their investment portfolios sinking in value, it is likely many of them could not remember exactly who had said what to them about investing their money in the market back when they made the decision years earlier. While this sort of forgetting may often reduce the costs of having been wrong, our experimental context is likely to have minimized its role. We made it extremely easy for guessers to keep track of advisors by providing clarity in their advice, clear feedback about the true answer, and more immediate feedback than usually exists (Tetlock, 2005).

Advisors also may rationalize away their failures 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). For example, proponents of the Iraq War contend that their visions for an end to post-war sectarian violence would have been realized if not for various unforeseen strategic missteps (e.g., initial shortages of troops on the ground) and polarizing incidents (e.g., the bombing of the al-Askari Mosque). Rather than having been wrong, advisors can instead claim that they were "almost right."

The third possibility is that customers believe that "this time it's different." In other words, advisors could argue that their previous failures should not be held against them because the regime, environment, or paradigm in which they previously appeared confident but wrong does not apply to the current situation. 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 & 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 (Kemp, 2007; Roberts & Kraynak, 2006). 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 our experimental paradigm, we did not allow our advisors an avenue of communication through which they could articulate such arguments. Nevertheless, their persistent claims of great confidence, despite their unimpressive prior accuracy, have the same implication.

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 we present 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 show here, differences in the composition of these exchange systems can have a significant influence on the estimates they generate. As 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 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.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know it, I know I saw it": The stability of the confidenc-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5*(1), 76-88.

Brewer, N., & 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.

Conger, J. A., & 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., & Thaler, R. H. (1995). Financial decision-making in markets and firms: A behavioral perspective. In R. A. Jarrow, V. Maksimovic & W. T. Ziemba (Eds.), *Finance, Handbooks in Operations Research and management Science* (Vol. 9, pp. 385-410). North Holland, Amersterdam: Elsevier.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*, 346-357.

Flynn, F. J., & 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., & 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.

Glassman, J. K., & 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., & 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., & 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 & 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.

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and Confidence in Person Identification: The Relationship Is Strong When Witnessing Conditions Vary Widely. *Psychological Science, 9*(3), 215-218.

Malmendier, U., & Tate, G. (2005). CEO overconfidence and corporate investment. *Journal of Finance*(60), 6.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517.

Moore, D. A., & Klein, W. M. P. (2008). The use of absolute and comparative performance feedback in absolute and comparative judgments and decisions. *Organizational Behavior & Human Decision Processes, 107*, 60-74.

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 & A. E. Roth (Eds.), *The handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Price, P. C., & 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., & Kraynak, J. (2006). *Flipping houses for dummies*. New York: Wiley.

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

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics, 1*(1), 43-61.

Smith, A. (1776). *An inquiry into the nature and causes of the wealth of nations*. London: Strahan & Cadell.

Sniezek, J. A., & 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., & 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., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science, 18*(1), 46-50.

Tenney, E. R., Spellman, B. A., & 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.

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295.

Westley, F., & Mintzberg, H. (1989). Visionary leadership and strategic management. *Strategic Management Journal, 10*, 17-32.

Yaniv, I., & 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., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making, 10*(1), 21-32.

TABLE 1

Outline of the Market Interface

	Weight	Advisor Blue Confidence %	Advisor Green Confidence %	Advisor Red Confidence %	Advisor Yellow Confidence %
Example advisor responses	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

Results of Regression Analyses for Advisor Selection (Study 1)

(Standardized *β* weights for independent variables)

	Model 1	Model 2	Model 3	Model 4
Round	-0.0600	-0.0361	-0.0313	-0.0349
Peak confidence	0.1997**	0.2095**	0.2070**	0.1958**
Correct interval confidence	-0.0173	-0.0073	0.0132	-0.0104
Accuracy	-0.0640*	-0.0647	-0.0194	-0.0546
Range	-0.0236	-0.0366	-0.0434	-0.0394
Previous selection		0.0372	0.1319	-0.0781
Previous peak confidence		-0.0710	-0.0714	-0.0673
Previous correct interval confidence		0.0731	0.0402	-0.0864
Previous accuracy		0.0931	0.1324*	0.0760
Previous accuracy*Previous selection			-0.1631*	
Previous correct*Previous selection				0.2458*
R^2	0.0456	0.0493	0.0607	0.0724

* $p < .05$

** $p < .01$

TABLE 3

Probit Regression Results for Advisor Retention (Study 1)

	Model
Previous round guesser peak confidence	0.0037
Previous round guesser accuracy	-0.0008
Previous round guesser correct interval confidence	0.0075*
Previous round advisor peak confidence	0.0028
Previous round advisor accuracy	0.0018
Previous round advisor correct interval confidence	0.0099**
Round	0.0606
Pseudo-R^2	0.0722

* $p < .05$

** $p < .01$

TABLE 4

Probit Regression Results for Advisor Selection (Study 2)

	Model
Advisor peak confidence	0.0269**
Advisor correct interval confidence	.0028
Advisor range	-.0781
Advisor previous peak confidence	-.0169**
Advisor previous correct interval confidence	.0161***
Advisor previous accuracy	.0003
Advisor previous selection	.4536†
Guesser previous peak confidence	-.0041
Guesser previous correct interval confidence	-.0182***
Guesser previous accuracy	.0024
Round	-0.1634***
Pseudo-R^2	0.1608

† $p < .10$

* $p < .05$

** $p < .01$

*** $p < .001$

FIGURE 1

Peak and Correct Interval Confidence for Advisors and Guessers (Study 1)

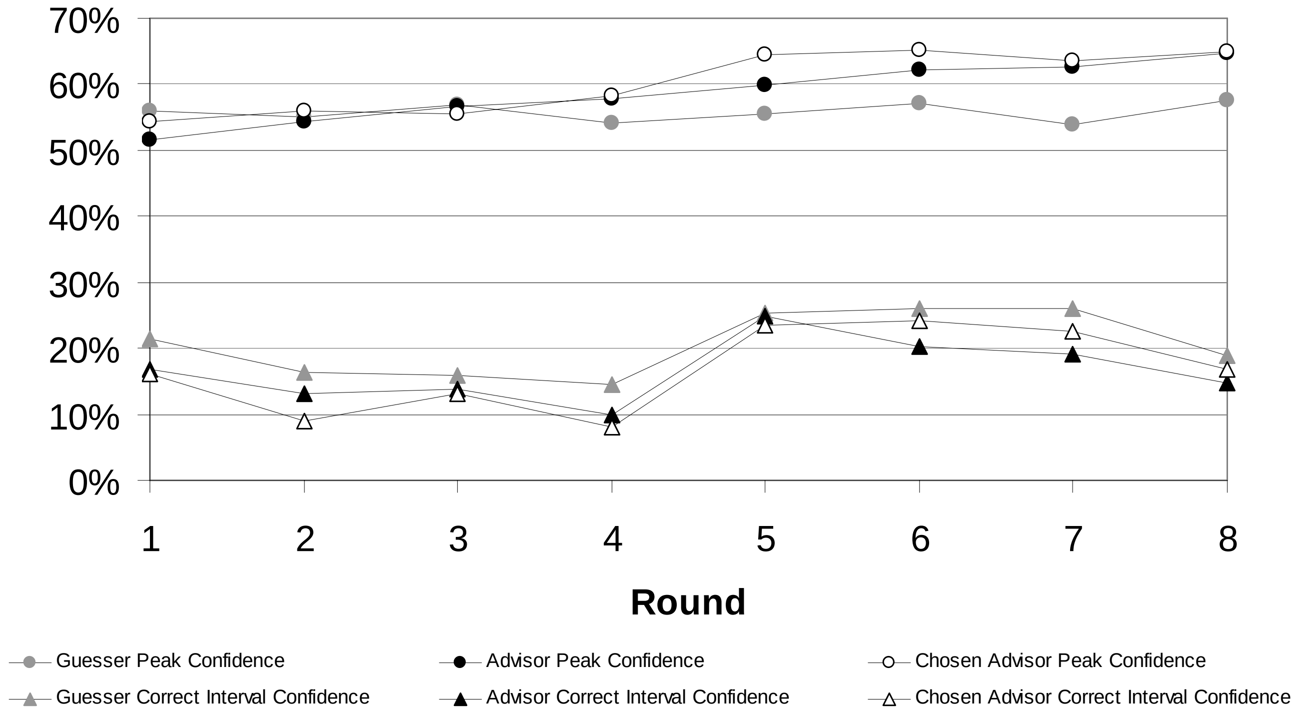

FIGURE 2

Absolute Difference Between Weighted Average Estimates and Correct Weights in pounds

(Advisor and Guesser Accuracy in Study 1)

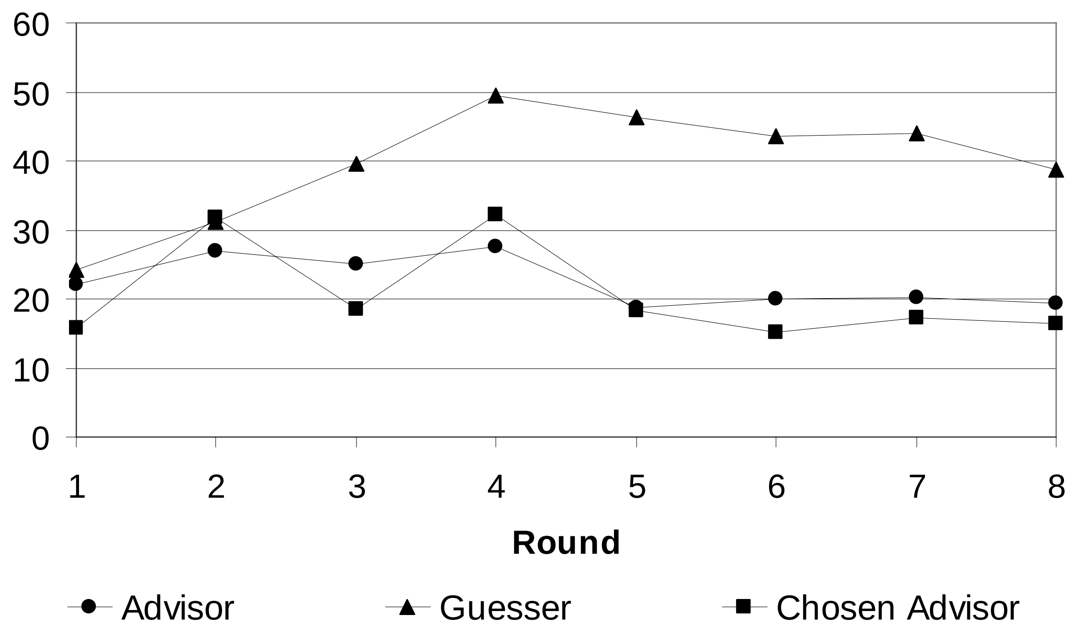

FIGURE 3

Peak and Correct Interval Confidence for Advisors and Guessers (Study 2)

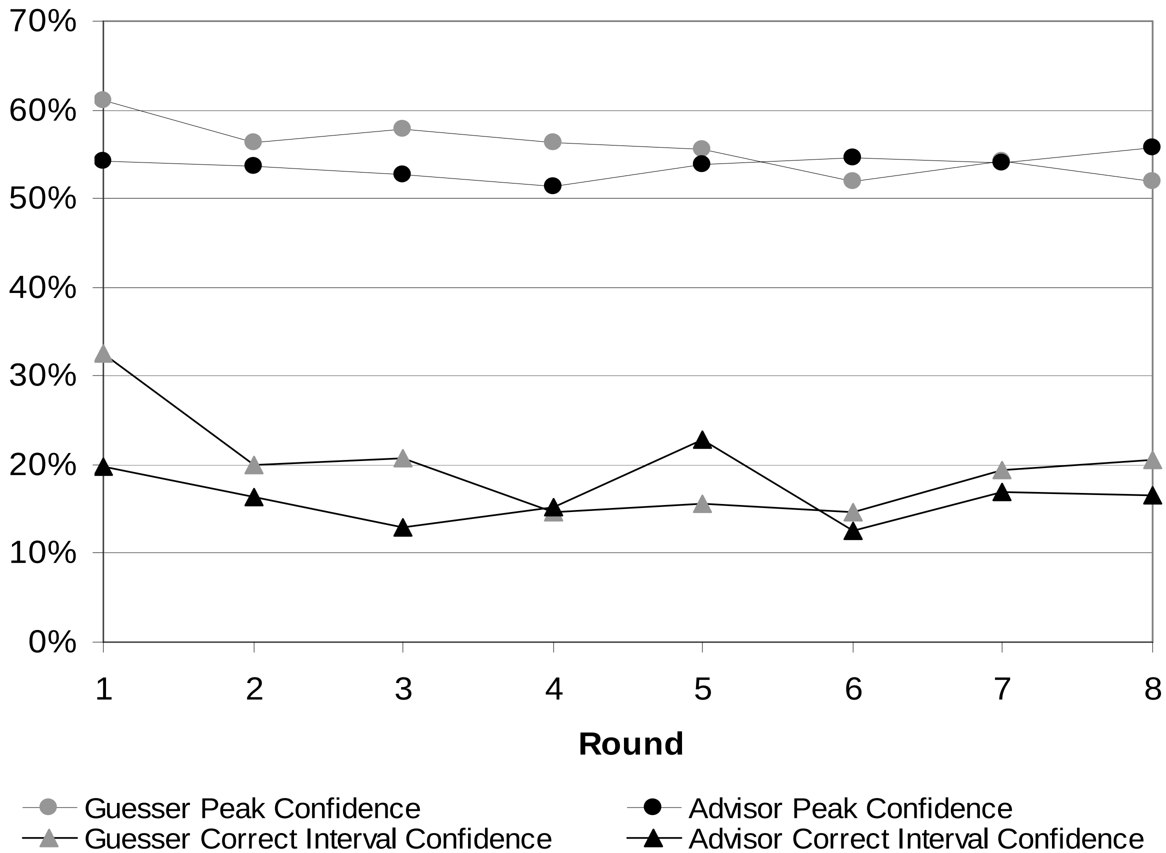

FIGURE 4

Absolute Difference Between Weighted Average Estimates and Correct Weights in pounds

(Advisor and Guesser Accuracy in Study 2)